MIMO Systems and Transmit Diversity

1 Introduction

So far we have investigated the use of antenna arrays in interference cancellation and for receive diversity. This final chapter takes a broad view of the use of antenna arrays in wireless communications. In particular, we will investigate the *capacity* of systems using multiple transmit and/or multiple receive antennas. This provides a fundamental limit on the data throughput in multiple-input multiple-output (MIMO) systems. We will also develop the use of transmit diversity, i.e., the use of multiple transmit antennas to achieve reliability (just as earlier we used multiple receive antennas to achieve reliability via receive diversity).

The basis for receive diversity is that each element in the receive array receives an *independent* copy of the same signal. The probability that all signals are in deep fade simultaneously is then significantly reduced. In modelling a wireless communication system one can imagine that this capability would be very useful on transmit as well. This is especially true because, at least in the near term, the growth in wireless communications will be asymmetric internet traffic. A lot more data would be flowing from the base station to the mobile device that is, say, asking for a webpage, but is receiving all the multimedia in that webpage. Due to space considerations, it is more likely that the base station antenna comprises multiple elements while the mobile device has only one or two.

In addition to providing diversity, intuitively having multiple transmit/receive antennas should allow us to transmit data faster, i.e., increase data throughput. The information theoretic analysis in this chapter will formalize this notion. We will also introduce a multiplexing scheme, transmitting multiple data streams to a single user with multiple transmit and receive antennas.

This chapter is organized as follows. Section 2 then presents a theoretical analysis of the capacity of MIMO systems. The following two sections, Sections 3 develops transmit diversity techniques for MIMO systems based on space-time coding. Section 4 then addresses the issue of maximizing data throughput while also providing reliability. We will also consider transmitting multiple data streams to a single user. This chapter ends in Section 5 with stating the fundamental tradeoff between data throughput (also called multiplexing) and diversity (reliability).

2 MIMO Capacity Analysis

Before investigating MIMO capacity, let us take a brief look at the capacity of single-input single-output (SISO) fading channels. We start with the original definition of capacity. This set of

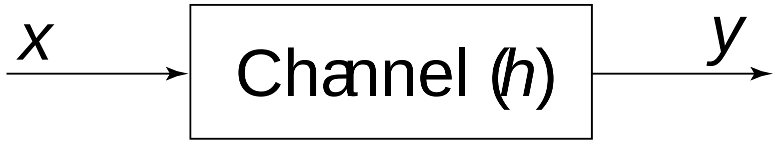


Figure 1: A single-input-single-output channel

notes assumes the reader knows the basics of information theory. See [1] for a detailed background. Consider the input-output system in Fig. 1. The capacity of the channel is defined as the maximum possible mutual information between the input (x) and output (y). The maximization is over the probability distribution of the input $f_X(x)$, i.e.

$$C = \max_{f_X(x)} [I(X;Y)] = \max_{f_X(x)} [h(Y) - h(Y/X)], \tag{1}$$

where $h(Y)$ is the entropy of the output Y.

For a SISO additive white gaussian noise (AWGN) channel, $y = x + n$, with $n \sim \mathcal{CN}(0, \sigma^2)$ and with limited input energy ($E\left\{|x|^2\right\} \leq E_s$), one can show that the capacity achieving distribution is Gaussian, i.e., $x \sim \mathcal{CN}(0, E_s)$ and $y \sim \mathcal{CN}(0, E_s + \sigma^2)$. It is not difficult to show that if n is Gaussian and has variance σ^2, $h(N) = \log_2(\pi e \sigma^2)$. Therefore $h(Y) = \log_2(\pi e(E_s + \sigma^2))$. Also, $h(Y/X)$ is the residual entropy in Y given the channel input X, i.e., it is the entropy in the noise term N. Therefore, $h(Y/X) = \log_2(\pi e \sigma^2)$ and the channel capacity, in bits/s/Hz, is given by

$$C = [h(Y) - h(Y/X)] = \log_2\left(\frac{E_s + \sigma^2}{\sigma^2}\right) = \log_2(1 + \rho), \tag{2}$$

where $\rho = E_s/\sigma^2$ is the signal-to-noise ratio (SNR).

In the case of a fading SISO channel, the received signal at the k-th symbol instant is $y[k] = h[k]x[k] + n[k]$. To ensure a compatible measure of power, set $\mathrm{E}\left\{|h[k]|^2\right\} = 1$ and $\mathrm{E}\left\{|x[k]|^2\right\} \leq E_s$. At this point there are two possibilities, a fixed fading channel with a random but unchanging channel gain and a slow, but fluctuating channel. In the first case, the capacity is given by

$$C = \log_2\left(1 + |h|^2 \rho\right), \tag{3}$$

where $\rho = E_s/\sigma^2$. An interesting aspect of this equation is that in a random, but fixed channel, the theoretical capacity may be zero. This is because, theoretically, the channel gain could be as close to zero making *guaranteeing* a data rate impossible. What is possible in this case is determining what are the chances a required capacity is available. This requires defining a new probability of outage, P_{out}, as the probability that the channel capacity is below a threshold rate R_0.

$$P_{\text{out}} = P(C < R_0) = P\left(|h|^2 > \frac{2^{R_0} - 1}{\rho}\right), \tag{4}$$

$$= 1 - \exp\left\{-\frac{2^{R_0} - 1}{\rho}\right\}, \tag{5}$$

where the final equation is valid for Rayleigh fading. Note that in the high-SNR regime ($\rho \to \infty$),

$$P_{\text{out}} \propto \frac{1}{\rho}, \tag{6}$$

i.e., at high SNR, the outage probability falls off inversely with SNR.

In the case of a time varying channel, assuming sufficient interleaving that the channel is independent from one symbol to the next, the average capacity over K channel realizations is

$$C_K = \frac{1}{K} \sum_{k=1}^{K} \left\{ \log_2 \left(1 + |h_k|^2 \rho \right) \right\}. \tag{7}$$

Based on the law of large numbers, as $K \to \infty$ the term on the right converges to the average or expected value. Hence,

$$C = \mathrm{E}_h \left\{ \log_2 \left(1 + |h|^2 \rho \right) \right\}, \tag{8}$$

where the expectation operation is taken over the channel values h. Note that this expression is non-zero and therefore with a *fluctuating* channel it is possible to guarantee the existence of an error-free data rate.

2.1 MIMO Systems

We now consider MIMO systems with the goal of evaluating the capacity of a system using N transmit and M receive antennas. We begin with the case of N parallel channels - basically N SISO channels operating in parallel. However, we will assume that the transmitter knows the N channels and can therefore *allocate power* intelligently to maximize capacity.

2.1.1 Parallel Channels

The N parallel channels are AWGN with a noise level of σ^2. The received data ($\mathbf{y}$) from input data $\mathbf{x}$ over N channels is modelled as

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{9}$$

$$\mathrm{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma^2 \mathbf{I}_N. \tag{10}$$

The transmitter has an energy budget of E_s which must be allocated across the N channels. The capacity of this channel is

$$C = \max_{\{E_n\} \sum_{n=1}^{N} E_n \leq E_s, E_n \geq 0} \sum_{n=1}^{N} \log_2 \left(1 + \frac{E_n |h_n|^2}{\sigma_n^2} \right), \tag{11}$$

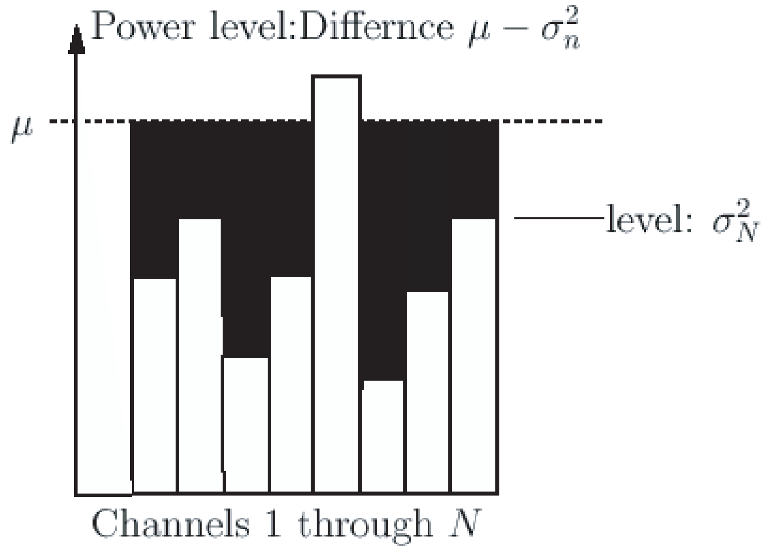


Figure 2: Illustrating Waterfilling.

where E_n is the energy allocated to the n^{th} channel. The equivalent Lagrange problem is[1]:

$$L(\{E_n\};\lambda) = \sum_{n=1}^{N} \log_2\left(1 + \frac{E_n |h_n|^2}{\sigma^2}\right) + \lambda\left(\sum_{n=1}^{N} E_n - E_s\right) \quad (12)$$

$$\Rightarrow \frac{\partial L}{\partial E_n} = \frac{|h_n|^2}{\sigma^2} \frac{\log_2(e)}{\left(1 + \frac{E_n|h_n|^2}{\sigma^2}\right)} + \lambda = 0, \quad (13)$$

$$\Rightarrow \forall n, \qquad \left(\frac{\sigma^2}{|h_n|^2} + E_n\right) = \mu \quad \text{(a constant)}. \quad (14)$$

Since $E_n \geq 0$,

$$E_n = \left(\mu - \frac{\sigma^2}{|h_n|^2}\right)^+, \quad (15)$$

where $(x)^+$ indicates only positive numbers are allowed, i.e. $(x)^+ = x$ if $x \geq 0$, else $(x)^+ = 0$. The constant μ is chosen to meet the total energy constraint. Equation (15) tells us how to allocate energy *given* knowledge of the channel attenuation through which the data must suffer.

Interestingly, the optimal power allocation scheme does *not* allocate all the power to the best channel. This is because the $\log_2(1 + \rho)$ expression for capacity implies a diminishing marginal returns on adding signal power (the capacity grows only as $\log_2$ at high SNR, but linearly at low-SNR). So providing some power to weaker channels can actually increase overall sum capacity.

[1]Note that the Lagrange problem being set up ignores the constraint that $E_n \geq 0$ for now and that this constraint is "added" later. A formal proof that this is OK will take us into a detour. The proof uses the fact that if we were to add this constraint (N of them), the associated Lagrange multiplier is either zero or the constraint is not met with equality.

This optimal scheme is known as *waterfilling.* An intuitive understanding of waterfilling (and why it is called so) may be obtained from Fig. 2, borrowed from Prof. Schlegel [2]. In the figure, σ_n^2 refers to the effective noise power at each time instant, $\sigma^2/|h_n|^2$. Waterfilling tells us that the optimal strategy is to 'pour energy' (allocate energy on each channel). In channels with lower noise power, more energy will be allocated. In channels with large noise power, the energy allocated is low . Some channels are so weak that the effective noise power becomes very large. Waterfilling tells us that transmitting any information on these channels is a waste of energy. If energy is allocated, the sum of the allocated energy and the effective noise power ($\sigma_n^2 = \sigma^2/|h_n^2|$) is a constant (the "water level", μ). Finally, if the channel were all equal, i.e. σ_n^2 were a constant, waterfilling leads to an equal energy distribution. Determining the water level, μ, is an iterative process.

The capacity on using the waterfilling approach is

$$C = \sum_{n=1}^{N} \log_2 \left(1 + \frac{E_n |h_n|^2}{\sigma^2} \right). \tag{16}$$

Aside: The result also leads to an interesting observation: if one could only focus on the times that the channel is in a "good" condition one could get enormous gains in capacity. Of course, this may not always be possible. However, thinking of a *multiuser* situation, if the channel to each user is changing with time, it is likely that at any time instant, one user has a good channel. By transmitting energy on that channel, overall capacity can be achieved in a *multiuser situation.* This is a new form of diversity called "opportunistic beamforming" [3].

Finally, if the channel is *not available* at the transmitter, clearly the best distribution scheme is to spread the energy evenly between all transmitters, i.e. $E_n = E_s/N$ and

$$C = \sum_{n=1}^{N} \log_2 \left(1 + \frac{E_s}{N\sigma^2} \right). \tag{17}$$

Note that since the log function increases significantly slower than the linear N term, the overall capacity is significantly larger than that for the SISO case.

2.1.2 Known MIMO Channels

We now turn to the more practical MIMO situation with N transmitters and M receivers with a full $M \times N$ channel matrix $\mathbf{H}$ in between. We will assume we know the channel matrix $\mathbf{H}$ at both the transmitter and receiver. Also, we will set $M \leq N$, however, the results here are easily extended for $M > N$. To ensure no artificial amplification in the channel, we shall set $\mathrm{E}\{|h_{mn}|^2\} = 1$. The data received at the M elements can be modelled as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{18}$$

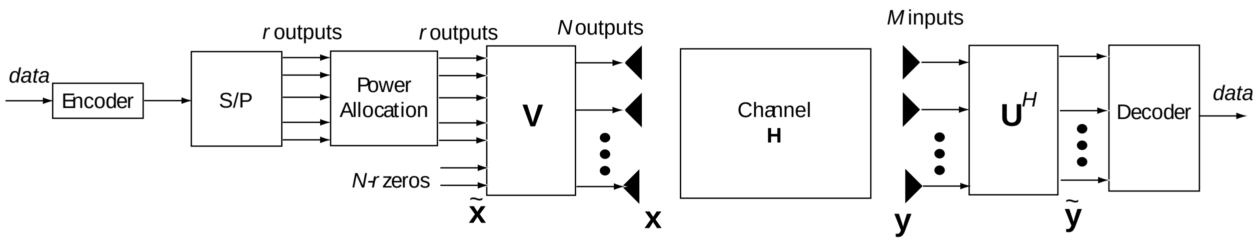


Figure 3: A communication system that achieves capacity.

where $\mathbf{H}$ is the full $M \times N$ channel matrix.

Based on the singular value decomposition[2] , one can decompose $\mathbf{H}$ as $\mathbf{H} = \mathbf{U\Sigma V}^H$, with $\mathbf{\Sigma} = [\text{diag}(d_1, d_2, \ldots, d_M) \,|\, \mathbf{0}_{M\times N-M}]$, where $d_m \geq 0$ are the M singular values of $\mathbf{H}$. Using Eqn. (18) and the fact that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_M$,

$$\begin{aligned} \mathbf{y} &= \mathbf{U\Sigma V}^H\mathbf{x} + \mathbf{n}, && (19)\\ \Rightarrow \mathbf{U}^H\mathbf{y} &= \mathbf{\Sigma V}^H\mathbf{x} + \mathbf{U}^H\mathbf{n}, && (20)\\ \Rightarrow \tilde{\mathbf{y}} &= \mathbf{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, && (21) \end{aligned}$$

where $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$ and $\tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$. *This transformed data in Eqn. (21) is equivalent to M parallel channels* with effective noise powers of $\sigma_m^2 = \sigma^2/d_m^2$. Note that if $\text{E}\{\mathbf{nn}^H\} = \sigma^2\mathbf{I}$, $\text{E}\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\} = \text{E}\{\mathbf{U}^H\mathbf{nn}^H\mathbf{U}\} = \sigma^2\mathbf{U}^H\mathbf{IU} = \sigma^2\mathbf{I}$. Furthermore, since $\mathbf{V}^H\mathbf{V} = \mathbf{I}_N$, the energy constraint remains the same, i.e., $\sum_{n=1}^N \tilde{E}_n = E_s$. Since the last $(N-M)$ columns of $\mathbf{\Sigma}$ are all zero, the last $(N-M)$ entries in $\tilde{\mathbf{x}}$ are irrelevant. In fact, if the rank of $\mathbf{H}$ is r, the system is equivalent to r parallel channels only. Note that $r \leq \min(N, M)$.

In the rotated (tilde) space MIMO communications is exactly the same as r parallel channels. The optimal power allocation is, therefore, the same waterfilling scheme as with the N parallel channels in Section 2.1.1. However, now the energy is spread over the eigen-channels, as opposed to physical channels. Figure 3 illustrates the communication system being considered. The data to be transmitted in encoded (if the encoder achieves capacity in an AWGN channel the overall scheme achieves channel capacity) and sent onto a serial-to-parallel converter with r outputs, where r is the rank of the channel matrix. *The waterfilling scheme is used to determine the powers of each element in these r outputs.* The r outputs are augmented with $(N-r)$ zeros to form the data vector $\tilde{\mathbf{x}}$. Multiplying with the right singular vector matrix $\mathbf{V}$ leads to the data vector $\mathbf{x}$ to be transmitted over the N elements. This transmission suffers channel $\mathbf{H}$. At the receiver the

[2]Any $M \times N$ matrix $\mathbf{A}$ can be decomposed as $\mathbf{A} = \mathbf{U\Sigma V}^H$. The columns of $\mathbf{U}$ are the M eigenvectors of $\mathbf{HH}^H$ and the columns of $\mathbf{V}$ are the N eigenvectors of $\mathbf{H}^H\mathbf{H}$. The $M \times N$ matrix $\mathbf{\Sigma}$ is a diagonal matrix of singular values. If $M \leq N$, $\mathbf{\Sigma} = [\,\text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_M) \,|\, \mathbf{0}_{M\times N-M}]$ where σ_m^2 are the M eigenvalues of $\mathbf{HH}^H$. Note that this is for an arbitrary rectangular matrix $\mathbf{A}$ and these singular values should not be confused with the noise power. Since $\mathbf{HH}^H$ and $\mathbf{H}^H\mathbf{H}$ are positive semi-definite matrices, $\mathbf{UU}^H = \mathbf{U}^H\mathbf{U} = \mathbf{I}_M$, $\mathbf{VV}^H = \mathbf{V}^H\mathbf{V} = \mathbf{I}_N$ and $\sigma_m \geq 0$. The matrix $\mathbf{U}$ ($\mathbf{V}$) is the matrix of left (right) singular vectors.

length-M data vector $\mathbf{y}$ is multiplied by the left singular vectors ($\mathbf{U}^H$) resulting in the transformed vector $\tilde{\mathbf{y}}$. This transformed vector is used for decoding the original data symbols.

The optimal energy distribution $\tilde{E}_m$ on the m-th channel and overall capacity are given by

$$\tilde{E}_m = \left(\mu - \frac{\sigma^2}{d_m^2}\right)^+, \tag{22}$$

$$C = \sum_{m=1}^{R} \log_2\left(1 + \frac{\tilde{E}_m d_m^2}{\sigma^2}\right). \tag{23}$$

To illustrate the workings of this capacity formula, let us consider four examples:

Case 1: 1 transmitter and M receivers, $\mathbf{H} = [h_1, h_2, \ldots, h_M]^T$, rank($\mathbf{H}$) = 1.

Since rank($\mathbf{H}$) = 1, only one singular value is non-zero and all the energy is allocated to this eigen-channel. This singular value and the resulting capacity are given by

$$d_1 = \sqrt{|h_1|^2 + |h_2|^2 + \ldots |h_M|^2}, \tag{24}$$

$$C = \log_2\left(1 + \frac{E}{\sigma^2}\sum_{m=1}^{M} |h_m|^2\right). \tag{25}$$

Case 2: N transmitters and 1 receiver, $\mathbf{H} = [h_1, h_2, \ldots, h_N]$, rank($\mathbf{H}$) = 1.

Since rank($\mathbf{H}$) = 1, only one singular value is non-zero and all the energy is allocated to this eigen-channel. This singular value and the resulting capacity are given by

$$d_1 = \sqrt{|h_1|^2 + |h_2|^2 + \ldots |h_N|^2}, \tag{26}$$

$$C = \log_2\left(1 + \frac{E}{\sigma^2}\sum_{n=1}^{N} |h_n|^2\right), \tag{27}$$

Note that this result is valid *only if the channel is known at the transmitter.*

Case 3: N transmitters and M receivers with perfect line of sight (LOS), without multipath.

Let d_t be the distance between the transmit elements and d_r the distance between the receive elements. The transmitter transmits in direction ϕ_t with respect to its baseline while the receiver receives from angle ϕ_r with respect to its baseline. In this case,

$$h_{mn} = \exp(jkd_r(m-1)\cos\phi_r)\exp(jkd_t(n-1)\cos\phi_t). \tag{28}$$

Note that even though the channel matrix $\mathbf{H}$ is $M \times N$, it is still rank-1 and $d_1 = \sqrt{NM}$. The capacity is given by

$$C = \log_2\left(1 + NM\frac{E_s}{\sigma^2}\right), \tag{29}$$

i.e., in line-of-sight conditions, the arrays at the transmitter and receiver only provide a power gain of NM.

Case 4: $N = M$ and the channel has full rank with equal singular values.

Since the square of the singular values of $\mathbf{H}$ are the eigenvalues of $\mathbf{H}\mathbf{H}^H$,

$$\sum_{m=1}^{M} d_m^2 = \mathrm{trace}\left(\mathbf{H}\mathbf{H}^H\right) = \sum_{n=1}^{N}\sum_{m=1}^{M} |h_{mn}|^2.$$

Since, on average, the each channel has unit power and we assume equal singular values, $d_m^2 = NM/M = N$, $\forall m$. Since all singular values are equal the energy allocation is clearly uniform ($E_m = E_s/N$) and

$$C = \sum_{m=1}^{M} \log_2\left(1 + \frac{E_s d_m^2}{N\sigma^2}\right) = \sum_{m=1}^{M} \log_2\left(1 + \frac{E_s}{\sigma^2}\right) = N\log_2\left(1 + \frac{E_s}{\sigma^2}\right). \tag{30}$$

Note the significant difference in the capacities described in Eqns. (29) and (30). Under perfect LOS conditions, the transmit and receive array only provide *power gain* and the capacity increases as the log of the number of elements. However, when the channel is set up such that each eigen-channel is *independent* and has equal power, the capacity gains *are linear*. The independent channels allow us to transmit *independent* data streams (N in the final example above), thereby increasing capacity.

In summary, in this section we have shown that a system with N transmitters and M receivers can be reduced to a problem of r parallel AWGN channels, where r is the rank of the channel matrix. To achieve the greatest gains in capacity, the channels from two different transmitters to the receivers must be independent and have equal power. The maximum possible gain in the channel capacity (over the SISO case) is the minimum of the number of transmitters and receivers, i.e., $\min(N, M)$. We will address this final constraint on the linear growth in capacity again in Section 2.1.4.

2.1.3 Channel Unknown at Transmitter

The analysis in Section 2.1.2 assumes both the transmitter and receiver know the channel matrix $\mathbf{H}$. However, in the more practical case that the channel is not known at the transmitter, but is known at the receiver, the approach is not valid. In this case, channel capacity must be determined as the maximum possible mutual information between input $\mathbf{X}$ and output $\mathbf{Y}$.

The capacity is given by

$$C = \max_{f_{\mathbf{X}}(\mathbf{x})} I(X;Y) = \max_{f_{\mathbf{X}}(\mathbf{x})} \left[H(\mathbf{Y}) - H(\mathbf{Y}/\mathbf{X})\right], \tag{31}$$

where $H(\mathbf{X})$ is the entropy in $\mathbf{X}$ with probability density function $f_{\mathbf{X}}(\mathbf{x})$ and is not to be confused with the channel matrix $\mathbf{H}$. Assuming channel matrix $\mathbf{H}$ is *known at the receiver*, the entropy in $\mathbf{Y}$, given the input data $\mathbf{X}$, is clearly only due to the noise $\mathbf{N}$. Assuming the noise to be complex, white and Gaussian with variance σ^2,

$$H(\mathbf{Y}/\mathbf{X}) = H(\mathbf{N}) = M \log_2(\pi e \sigma^2) = \log_2(\pi e \sigma^2)^M, \tag{32}$$

Given the channel, the entropy is $\mathbf{Y}$ is determined by the distribution of $\mathbf{X}$. We invoke the fact that the input distribution required to achieve capacity is *Gaussian*, i.e., $\mathbf{X}$ must be Gaussian distributed with $\mathbf{X} \sim N(0, \mathbf{\Sigma}_x)$ where $\mathbf{\Sigma}_x$ is the covariance matrix of $\mathbf{X}$ and whose diagonal entries are such that they meet the criterion of limited transmit energy.

From Eqn. (18), given $\mathbf{H}$, $\mathbf{Y}$ is also Gaussian with $\mathbf{Y} \sim N(0, \mathbf{\Sigma}_y)$ where $\mathbf{\Sigma}_y = \sigma^2 \mathbf{I}_M + \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H$ and $\mathbf{I}_M$ is the $M \times M$ identity matrix. Using the entropy result for the Gaussian pdf [1],

$$\begin{aligned} H(\mathbf{Y}) &= \log_2 \left[(\pi e)^M \det \mathbf{\Sigma}_y \right], & (33) \\ \Rightarrow C = \max_{f_{\mathbf{X}}(\mathbf{x})} I(\mathbf{X};\mathbf{Y}) &= \log_2 \left[(\pi e)^M \det \left(\sigma^2 \mathbf{I}_M + \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H \right) \right] - \log_2 \left(\pi e \sigma^2 \right)^M, & (34) \\ &= \log_2 \det \left(\mathbf{I}_M + \frac{1}{\sigma^2} \mathbf{H}\mathbf{\Sigma}_x\mathbf{H}^H \right), & (35) \end{aligned}$$

Based on an eigendecomposition of the covariance matrix of the input data, $\mathbf{\Sigma}_x$, one can show that the optimal covariance matrix is $\mathbf{\Sigma}_x = (E_s/N)\mathbf{I}_N$ [1, 2, 4] which corresponds to independent data streams and equal power distribution over all available channels. The capacity is therefore,

$$C = \log_2 \det \left(\mathbf{I}_M + \frac{E_s}{N\sigma^2} \mathbf{H}\mathbf{H}^H \right). \tag{36}$$

Note that, as with SISO channels, for a fixed MIMO channel unknown at the transmitter, the true capacity is zero since we cannot guarantee any minimum channel quality.

2.1.4 Fading MIMO Channels

So far we have focused on fixed channels. In the most practical situation, the channels vary as a function of time. In this case, the channel can change from one time instant to the next. Assuming sufficient interleaving to make the channel independent from one symbol instant to the next, the average capacity over a block of K data symbols is given by

$$\begin{aligned} C &= \frac{1}{K} \sum_{k=1}^{K} \max_{f_{\mathbf{X}}(\mathbf{x})} I(\mathbf{X}[\mathbf{k}];\mathbf{Y}[\mathbf{k}]) \\ &= \frac{1}{K} \sum_{k=1}^{K} \log_2 \det \left(\mathbf{I}_M + \frac{E_s}{N\sigma^2} \mathbf{H}[\mathbf{k}]\mathbf{H}[\mathbf{k}]^H \right). & (37) \end{aligned}$$

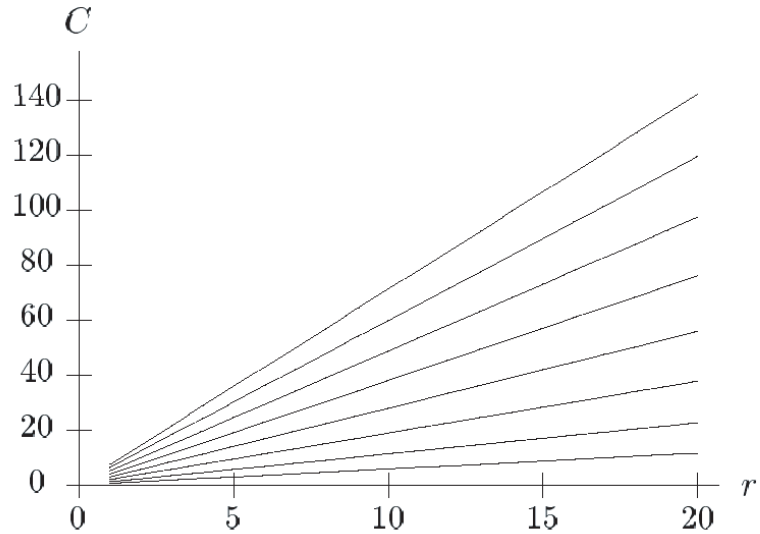


Figure 4: MIMO capacity in fading channels [5].

Based on the law of large numbers as $K \to \infty$ this approaches the expectation value of the right hand side in Eqn. (36) [4]

$$C = \mathrm{E}\left\{\log_2 \det\left(\mathbf{I}_M + \frac{E_s}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right)\right\}. \tag{38}$$

If $\{d_m^2,\ m = 1, 2, \ldots, M\}$ are the M eigenvalues of $\mathbf{H}\mathbf{H}^H$, the eigenvalues of $\left(\mathbf{I}_M + E_s/(N\sigma^2)\mathbf{H}\mathbf{H}^H\right)$ are $1 + E_s/(N\sigma^2)d_m^2$. The capacity in Eqn. (38) is then

$$C = \mathrm{E}\left\{\sum_{m=1}^{M}\log_2\left(1 + \frac{E_s}{N\sigma^2}d_m^2\right)\right\}, \tag{39}$$

where the expectation is taken over the M eigenvalues.

This result in Eqns. (38) and (39) is valid for any type of fading. Specializing this to the case of completely independent Rayleigh fading from each transmit element to each receive element, each individual entry in $\mathbf{H}$ is an independent complex Gaussian random variable. In this case, the matrix $\mathbf{H}\mathbf{H}^H$ is Wishart distributed [6]. In addition, the pdf of its eigenvalues are known [5, 7].

$$f(d_1^2, \ldots, d_M^2) = \frac{1}{MK_{M,N}} e^{-\sum_{m=1}^{M} d_m^2} \prod_{m=1}^{M} (d_m^2)^{N-M} \prod_{m<n} \left(d_m^2 - d_n^2\right)^2, \tag{40}$$

where $K_{M,N}$ is a normalizing factor and

$$C = \sum_{m=1}^{M} \mathrm{E}_{\{d_m^2\}}\left\{\log_2\left(1 + \frac{E_s d_m^2}{N\sigma^2}\right)\right\}, \tag{41}$$

$$\Rightarrow C = M\left[\mathrm{E}_{\{d_1^2\}}\log_2\left(1 + \frac{E_s d_1^2}{N\sigma^2}\right)\right] \tag{42}$$

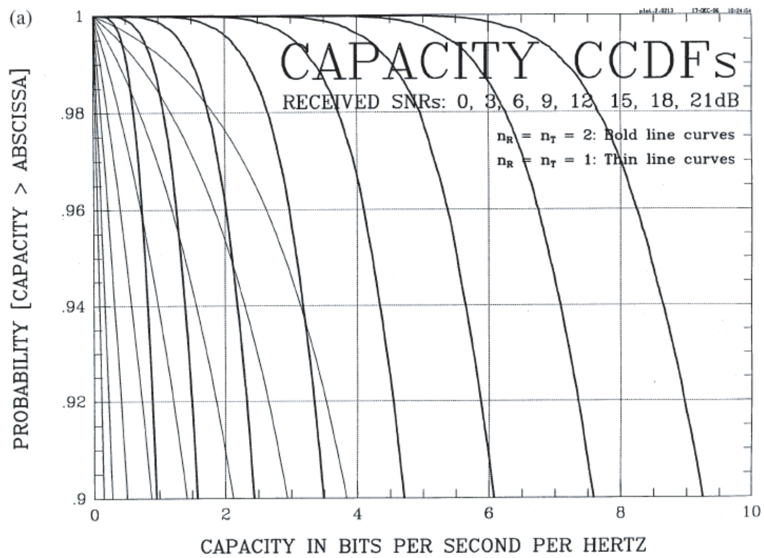


Figure 5: Outage probability in fading channels [8].

where the final expectation is taken over the pdf of an individual eigenvalue, found by marginalizing the multivariate pdf in Eqn. (40). The resulting capacity has been obtained by Telatar in [5] and shown in Fig. 4. The figure plots the capacity (in b/s/Hz) versus M or N. The eight plots are for different SNRs between 0dB and 35dB in steps of 5dB. Note the linear relationship between the capacity and the number of transmit and receive channels.

There are two important results included in here, one positive, one cautionary. First, just as when the channel is known at the transmitter, if the channel is Rayleigh and independent, it is possible to have *linear* increases in capacity in fading channels as well. The second (cautionary) result is that the increase is proportional to the minimum of the number of transmitters and receivers, i.e., $\min(N, M)$. This has important implications in a cellular network - it is reasonable to assume multiple elements at the base station. But, it is unlikely one could have more than one or two elements in a handheld device. In this case, multiple antennas at one end will only provide power gains, but not the parallel channels that provide large capacity gains.

The result in Eqn. (42) and Fig. 4 present the true capacity of a MIMO channel with independent Rayleigh fading, i.e., it is theoretically possible to have error-free transmission with rate below this capacity. Another figure of merit is outage probability such as derived in Eqn. (5) for SISO channels. In [8], Foschini and Gans evaluate the outage probability under Rayleigh fading. One of their results is shown in Fig. 5. Note the huge improvement in outage probability (here they plot the cumulative distribution, which is $(1 - P_{\text{out}})$ by moving from a SISO channel to $N = M = 2$. With a SNR of 21dB, the capacity of a SISO channel is larger than approximately 2.5b/s/Hz 96% of the time, while for $N = M = 2$ the capacity is larger than approximately 8.5b/s/Hz 96% of the time.

3 Transmit Diversity

So far, we have developed the capacity of MIMO systems in the case of the channel being known at the transmitter and receiver (leading to a waterfilling solution) and in the more practical case of the channel known at the receiver only (the results of [5,8]). This answers the question, "How fast can data be transmitted?", i.e., what is the theoretical maximum data rate that can be achieved in a MIMO system. We now investigate a different goal, using the multiple antennas to achieve reliability. We have already addressed this issue when the receiver has multiple receive antennas (receive diversity). Here we focus on *transmit diversity*. In a departure from the previous discussions, this will involve coding across the *space and time dimensions*. We begin with two remarkably simple schemes to achieve diversity on transmit, one inefficient, one efficient.

3.1 Space-Time Coding: Motivation

If using multiple receive antenna elements, we have shown that the optimal receiver is the maximal ratio combiner (MRC) which matches the receive weights to the channel. If the transmitted signal is $As_0u(t)$ within a symbol period of T_s, where s_0 is the symbol transmitted and $u(t)$ is the symbol waveform with unit energy, the received signal at the N elements (after the filter matched to the symbol waveform) and the combined signal are given by

$$\begin{aligned} \mathbf{x} &= \mathbf{g}s_0 + \mathbf{n} = \sqrt{E_s}\mathbf{h}s_0 + \mathbf{n}, & (43)\\ y &= \mathbf{h}^H\mathbf{x} = \sqrt{E_s}\left[\sum_{n=0}^{N-1}|h_n|^2\right]s_0 + \text{noise}, & (44) \end{aligned}$$

where E_s is the energy in the signal per symbol and the average energy in the fading term h_n is unity, i.e. $\mathrm{E}\{|h_n|^2\} = 1$. The MRC therefore results in the signal being multiplied by the sum of the powers in the channels. In transmit diversity, the array with N elements is at the transmitter. We will claim we have optimal (maximal ratio) transmission if we achieve a similar received signal. In the chapter on receive diversity we considered a system with a single transmit and N receive antennas. Here we will consider a system with N transmit and a single receive antenna. The receiver is assumed to know the channel.

An easy diversity scheme is to repeat the transmission of the same data symbol (s_0) over N symbol periods, one element at a time. At the receiver, the data from the N received signals from N symbol periods is written as a length-N vector $\mathbf{x}$ which is given by

$$\mathbf{x} = \sqrt{E_s}\mathbf{h}s_0 + \mathbf{n}, \qquad (45)$$

which has the same form as that for the receive diversity case. Note a crucial different though. This vector is the received signal at a *single element over N symbol intervals*. The vector $\mathbf{h}$ is the length-N vector of the channels from the N transmitting elements to the single receive element. The data vector $\mathbf{x}$ is processed as

$$y = \mathbf{h}^H \mathbf{x} = \sqrt{E_s} \left[\sum_{n=0}^{N-1} |h_n|^2 \right] s_0 + \text{noise}. \tag{46}$$

If the N channels are independent, this transmit scheme achieves diversity of order N.

This transmit diversity scheme is clearly very inefficient. Within any symbol period, only a single element is used. A single symbol is sent over N periods, i.e., one would need a bandwidth expansion of N to achieve the same data rate. On the other hand, this scheme shows that transmit diversity *is* possible. It also illustrates another important point - one cannot achieve transmit diversity by focusing on a single symbol period only. The scheme must also involve the time dimension - this is the basis for *space-time* coding - coding, by definition, introduces redundancy - to achieve transmit diversity one must introduce redundancy in both the space and time dimensions.

3.2 Diversity without wasting bandwidth

In [9] Alamouti presents a remarkably simple scheme to achieve transmit diversity, for an array of two elements, without any loss of bandwidth. The scheme transmits two symbols over two time periods (note again the time dimension is used). In the simplest case, the receiver has only a single element, though extensions are possible to receivers of multiple elements as well.

Denote two symbols to be s_0 and s_1. In the first symbol interval, transmit s_0 from the element #0 and s_1 from the element #1. In the next symbol interval, transmit $(-s_1^*)$ from element #0 and (s_0^*) from element #1 where the superscript * represents conjugation. The channel from the two elements to the receiver is assumed constant over both intervals $(2T_s)$. The two transmit antennas have a total energy budget of E_s, each symbol is transmitted with half the energy. Overall, the received signal over the two symbol intervals (y_0 and y_1) can be written as

$$y_0 = \sqrt{\frac{E_s}{2}} \left[h_0 s_0 + h_1 s_1 + n_0 \right], \tag{47}$$

$$y_1 = \sqrt{\frac{E_s}{2}} \left[-h_0 s_1^* + h_1 s_0^* + n_1 \right], \tag{48}$$

$$\Rightarrow \begin{bmatrix} y_0 \\ y_1^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_0 & h_1 \\ h_1^* & -h_0^* \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} + \begin{bmatrix} n_0 \\ n_1^* \end{bmatrix} \Rightarrow \mathbf{y} = \mathbf{H} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} + \mathbf{n} \tag{49}$$

Note that the second entry in the vector $\mathbf{y}$ is the conjugate of the data received on the second

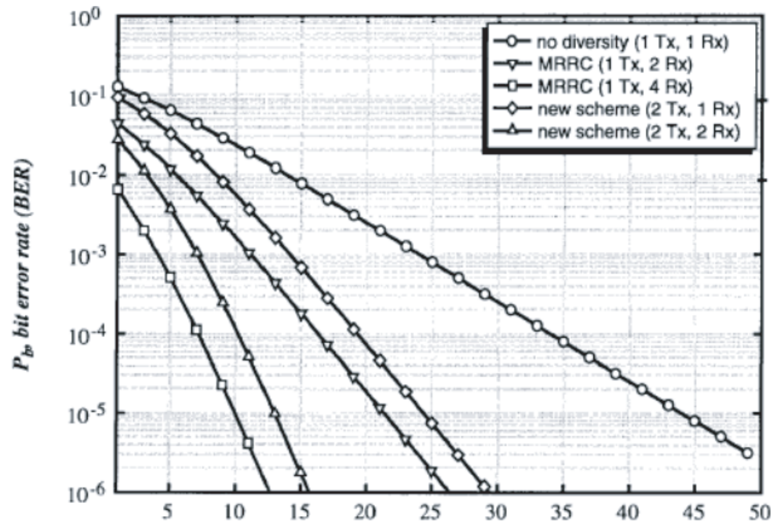


Figure 6: Performance of Alamouti's transmit diversity scheme

symbol interval.

$$\Rightarrow \mathbf{r} = \begin{bmatrix} r_0 \\ r_1 \end{bmatrix} = \mathbf{H}^H\mathbf{y} = \sqrt{\frac{E_s}{2}}\mathbf{H}^H\mathbf{H}\begin{bmatrix} s_0 \\ s_1 \end{bmatrix} + \mathbf{H}^H\mathbf{n}, \tag{50}$$

$$\Rightarrow r_0 = \sqrt{\frac{E_s}{2}}\left[|h_0|^2 + |h_1|^2\right]s_0 + h_0^* n_0 + h_1 n_1^*, \tag{51}$$

$$r_1 = \sqrt{\frac{E_s}{2}}\left[|h_0|^2 + |h_1|^2\right]s_1 - h_0^* n_1^* + h_1^* n_0. \tag{52}$$

Note that the equations for r_0 and r_1 include the squares of the two channel magnitudes, i.e., the received signal incorporates order-2 diversity. In addition, two symbols are transmitted over two symbol intervals and no bandwidth is wasted. This is the beauty of Alamouti's scheme. With a remarkably simple arrangement of transmitted and received data coupled with purely linear processing, order-2 diversity is achieved without any loss in bandwidth. The only 'penalty' is the halving in transmitted energy per symbol. We expect, therefore, the performance of Alamouti's scheme to be 3dB worse than the corresponding 2-element receive diversity case.

Figure 6 is taken from Alamouti's original paper [9]. The line furthest to the right corresponds to the SISO case of a single transmitter and receiver. The other two groups compare transmit and receive diversity. The middle curves compare two-element transmit and receive diversity. The 3dB loss for transmit diversity is clear. However, note Alamouti's transmit diversity scheme does achieve the order-2 diversity. This is similar to the curves on the left which compare four element receive diversity with the case of 2-transmit and 2-receive elements. Again, sharing the total power between the two transmitters causes a loss of 3dB. We will see soon that Alamouti's scheme, basically found serendipitously, is unique in its simplicity and efficiency.

3.3 Good Codes

In Section 3.2 we developed an excellent scheme to achieve diversity in the specific case of two transmit antennas. The scheme required careful arrangement of data over both space and time, introducing redundancy in the transmission process. This is reminiscent of error controlling coding wherein controlled redundancy is introduced to achieve reliable transmissions. Alamouti's scheme, therefore, is a good *space-time code.*

Left open, so far, are efficient schemes for transmit diversity in the general case. However, an immediate question arises "what are good codes?". This leads to an important question, "what makes a code good?". In [10], Tarokh, et al. answer this question in terms of a bound on the probability of error. The notes in this section are taken in a large part from their presentation on space-time coding (STC) in [10]. We will focus on independent Rayleigh fading.

Consider a MIMO system with N transmit and M receive antennas. The space-time code spans L symbols. The $M \times N$ channel matrix $\mathbf{H} = [h_{mn}]$ is assumed constant over these L symbols. The symbols themselves are normalized to have unit energy and each entry in the channel matrix satisfies $\mathrm{E}\{h_{mn}\} = 1$. The transmitter transmits the coded sequence $\underline{c} = \left\{c_1^0,\, c_1^1,\, \ldots,\, c_1^{(N-1)},\, c_2^0,\, c_2^1,\, \ldots,\, c_2^{(N-1)},\, c_L^0,\, c_L^1,\, \ldots,\, c_L^{(N-1)}\right\}$ over L time instants from the N elements. At time instant l, the symbol c_l^n is transmitted from element $\#n$. At each time instant l, the received data is given by

$$\begin{aligned} \mathbf{y}(l) &= \mathbf{H}\mathbf{x}(l) + \mathbf{n}, && (53) \\ \Rightarrow \mathbf{y}_m(l) &= \sqrt{E_s}\sum_{n=0}^{N-1} h_{mn}c_l^n + \mathbf{n}_m, && (54) \end{aligned}$$

where E_s is the energy received per symbol. A maximum-likelihood (ML) decoder uses this data to decode to a sequence $\underline{\tilde{c}} = \left\{\tilde{c}_1^0,\, \tilde{c}_1^1,\, \ldots,\, \tilde{c}_1^{(N-1)},\, \tilde{c}_2^0,\, \tilde{c}_2^1,\, \ldots,\, \tilde{c}_2^{(N-1)},\, \tilde{c}_L^0,\, \tilde{c}_L^1,\, \ldots,\, \tilde{c}_L^{(N-1)}\right\}$, which may not be the same as the transmitted sequence $\underline{c}$. However, note that since the transmitted sequence is $\underline{c}$ and the ML decoder decides on $\underline{\tilde{c}}$, both $\underline{c}$ and $\underline{\tilde{c}}$ *are valid codewords.*

Given the channel at the receiver, the probability of error, i.e., the probability that $\underline{c}$ is transmitted and $\underline{\tilde{c}}$ is decoded is bounded by

$$P(\underline{c} \to \underline{\tilde{c}}) \;\leq\; \exp\left[-d^2\left(\underline{c},\, \underline{\tilde{c}}\right)\frac{E_s}{4\sigma^2}\right], \tag{55}$$

where $d\left(\underline{c},\, \underline{\tilde{c}}\right)$ is the Euclidean distance between $\underline{c}$ and $\underline{\tilde{c}}$ weighted by the known channel, $\mathbf{H}$, at the receiver.

$$
\begin{aligned}
d^2\left(\underline{c}, \tilde{\underline{c}}\right) &= \sum_{m=0}^{M-1}\sum_{l=1}^{L}\left|\sum_{n=0}^{N-1} h_{mn}\left(c_l^n - \tilde{c}_l^n\right)\right|^2, & (56)\\
&= \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1} h_{mn}h_{mn'}^{*}\sum_{l=1}^{L}\left(c_l^n - \tilde{c}_l^n\right)\left(c_l^{n'} - \tilde{c}_l^{n'}\right)^{*}, & (57)\\
&= \sum_{m=0}^{M-1} \mathbf{\Omega}_m \mathbf{E}\mathbf{E}^H \mathbf{\Omega}_m^H, & (58)
\end{aligned}
$$

where * represents the conjugate transpose and

$$
\mathbf{\Omega}_m = \left[h_{m0}\; h_{m1}\; \ldots\; h_{m(N-1)}\right], \qquad (59)
$$

$$
\mathbf{E} = \begin{bmatrix} c_1^0 - \tilde{c}_1^0 & c_2^0 - \tilde{c}_2^0 & \cdots & c_L^0 - \tilde{c}_L^0 \\ c_1^1 - \tilde{c}_1^1 & c_2^1 - \tilde{c}_2^1 & \cdots & c_L^1 - \tilde{c}_L^1 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{N-1} - \tilde{c}_1^{N-1} & c_2^{N-1} - \tilde{c}_2^{N-1} & \cdots & c_L^{N-1} - \tilde{c}_L^{N-1} \end{bmatrix}. \qquad (60)
$$

Note that keeping with the notation [10], $\mathbf{\Omega}_m$, the channel from the N transmitters to the m-th receiving element is a row vector. Also, $\mathbf{E}$ is the $N \times L$ error matrix of the differences between the two codewords $\underline{c}$ and $\tilde{\underline{c}}$ over the N transmit elements and L time instants.

Since $\mathbf{E}\mathbf{E}^H$ is a positive semi-definite matrix, its eigenvectors are orthogonal to each other and one $\mathbf{E}\mathbf{E}^H = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, where $\mathbf{Q}$ is a unitary matrix. All eigenvalues in the diagonal of $\mathbf{\Lambda}$ satisfy $\lambda_n \geq 0$. Defining a new vector $\boldsymbol{\beta}_m = \mathbf{\Omega}_m\mathbf{Q}$, we have

$$
\begin{aligned}
\mathbf{\Omega}_m \mathbf{E}\mathbf{E}^H \mathbf{\Omega}_m^H &= \mathbf{\Omega}_m \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \mathbf{\Omega}_m^H = \boldsymbol{\beta}_m \mathbf{\Lambda} \boldsymbol{\beta}_m, & (61)\\
&= \sum_{n=0}^{N-1} \lambda_n \left|\beta_{mn}\right|^2, & (62)\\
\Rightarrow d^2(\underline{c}, \tilde{\underline{c}}) &= \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \lambda_n \left|\beta_{mn}\right|^2. & (63)
\end{aligned}
$$

Now, since the channel is assumed to be Rayleigh and any two channels are assumed independent, with unit average power, $\mathbf{\Omega}$ is zero-mean complex Gaussian and $\mathrm{E}\{\mathbf{\Omega}_m^H \mathbf{\Omega}_m\} = \mathrm{I}_N$. Since $\boldsymbol{\beta}_m$ is a linear combination of $\mathbf{\Omega}_m$, it too is zero-mean complex Gaussian and $\mathrm{E}\{\boldsymbol{\beta}_m^H \boldsymbol{\beta}_m\} = \mathbf{Q}^H \mathrm{E}\left\{\mathbf{\Omega}_m^H \mathbf{\Omega}_m\right\}\mathbf{Q} = \mathrm{I}_N$, i.e. $\boldsymbol{\beta}_m$ is also Rayleigh distributed and $|\beta_{mn}|^2$ is exponentially distributed with unit mean.

We know that if X is exponentially distributed with unit mean, $\mathrm{E}\{e^{-\gamma x}\} = 1/(1+\gamma)$. Putting this fact together with Eqns. (55) and (63),

$$\begin{aligned}
\mathrm{E}\left\{P(\underline{c}\rightarrow\underline{\tilde{c}})\right\} &\leq \leq \mathrm{E}\left\{\exp\left[-d^2(\underline{c},\underline{\tilde{c}})\frac{E_s}{4\sigma^2}\right]\right\}, && (64)\\
&\leq \mathrm{E}\left\{\exp\left[-\frac{E_s}{4\sigma^2}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\lambda_n\left|\beta_{mn}\right|^2\right]\right\}, && (65)\\
&\leq \prod_{m=0}^{M-1}\prod_{n=1}^{N-1}\mathrm{E}\left\{\exp\left[-\frac{E_s}{4\sigma^2}\lambda_n\left|\beta_{mn}\right|^2\right]\right\}, && (66)\\
&\leq \left[\prod_{n=0}^{N-1}\frac{1}{\left(1+\frac{E_s\lambda_n}{4\sigma^2}\right)}\right]^M. && (67)
\end{aligned}$$

Note that if rank $\left[\mathbf{EE}^H\right]=R$, the product in Eqn. (67) is only upto R and

$$\begin{aligned}
\mathrm{E}\left\{P(\underline{c}\rightarrow\underline{\tilde{c}})\right\} &\leq \left[\prod_{n=0}^{R-1}\frac{1}{\left(1+\frac{E_s\lambda_n}{4\sigma^2}\right)}\right]^M, && (68)\\
&\leq \left(\prod_{n=0}^{R-1}\lambda_n\right)^{-M}\left(\frac{E_s}{4\sigma^2}\right)^{-RM}, && (69)\\
&\leq \left(\left[\prod_{n=0}^{R-1}\lambda_n\right]^{1/R}\right)^{-RM}\left(\frac{E_s}{4\sigma^2}\right)^{-RM}. && (70)
\end{aligned}$$

Such a code is a called a R-space-time code.

In the chapter on receive diversity we defined the diversity order to be the slope of the BER v/s SNR curve (on a log-log scale). From Eqn. (70) we see that the space-time code provides a diversity order of RM. Second, the product of the eigenvalues of the error matrix provides an additional coding gain. Since $\underline{c}$ and $\underline{\tilde{c}}$ are two arbitrary codewords, we now know what makes a code good:

1. A good code has highest possible rank in the error matrix between any two codewords ($L \geq N$ and $R = N$), i.e., the Hamming distance between any two codewords must be N. This will provide the greatest *coding gain*
2. A good code has the highest product of eigenvalues of the error matrix [3], i.e., this gain, purely due to the choice of code is an additional *coding gain.*

In [10], Tarokh et al. also prove two limitations on the space-time code.

[3]This is not exactly the determinant since this is only the product of non-zero eigenvalues. However, since any decent space-time code design would have full rank, in practice this is often referred to as the determinant criterion.

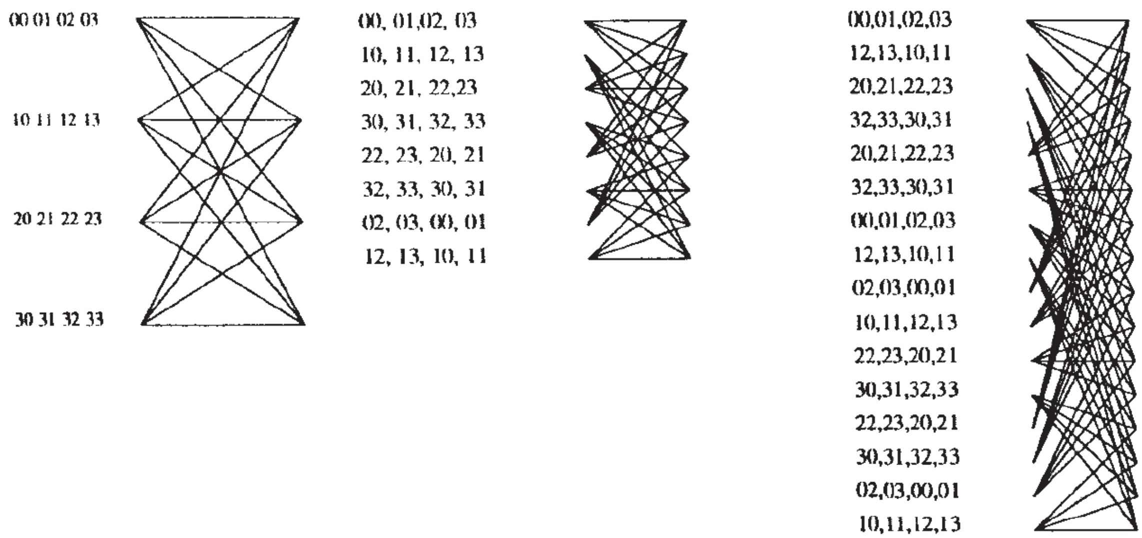


Figure 7: Space-Time Trellis Codes proposed by Tarokh et.al. for a 4-PSK constellation

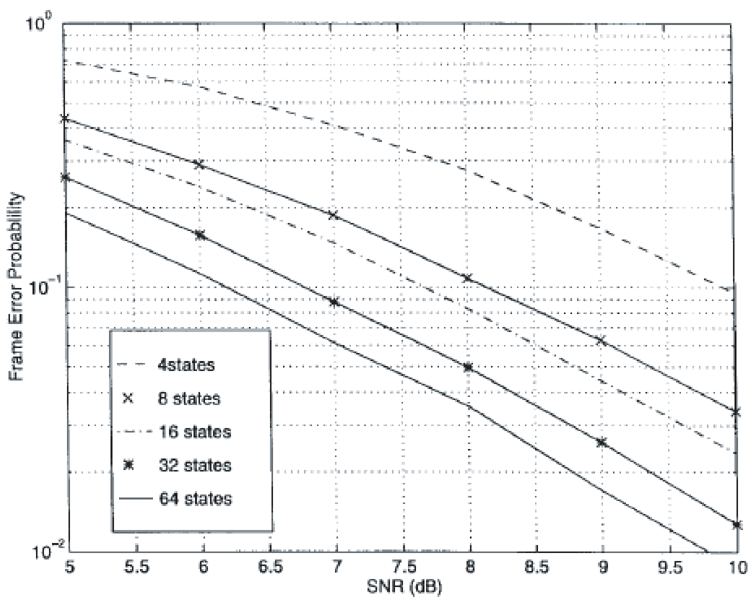


Figure 8: Results for the Space-Time Trellis Codes proposed in [10] for a 4-PSK constellation

Theorem 1: If the signal constellation has 2^b elements and $A_{2^{bL}}(N,R)$ is the number of codewords over L time instants with Hamming distance R, then the rate of the code (r), in bits/sec/Hz satisfies

$$r \leq \frac{\log_2 A_{2^{bL}}(N,R)}{L}. \tag{71}$$

Furthermore, $A_{2^{bL}}(N,R) < 2^{bL}$, i.e.

$$r \leq b \tag{72}$$

This tells us that to achieve a certain data throughput, one needs a certain level of complexity in the signal constellation. However, clearly that also makes for a more complicated code. This is especially true of the trellis codes suggested in [10].

The second result is specifically for trellis codes. The constraint length of a code is the number of steps before again reaching the all-zero codeword.

Theorem 2: A R-space-time code must have constraint length greater than or equal to $R-1$.

The second theorem tells us that to achieve a certain diversity order (R) one needs a suitably complicated trellis. Since ML decoding is exponential in the constraint length, this also means one needs a significantly complicated decoding mechanism. The decoding may therefore restrict the diversity achievable by a code. Note that the transmit diversity order cannot, in any case, greater than N, the number of transmit antennas.

3.4 Space Time Coding

Having described what makes a code good, Tarokh et al. also proposed some space-time trellis codes that meet the criteria developed in Section 3.3. Figure 7 provides the trellis diagrams for three codes presented in [10]. All three codes are for a 4-PSK constellation, achieving a rate of 2 b/s/Hz for a two-element array. The pairs of numbers on the left of the code represent the outputs of the encoder to be transmitted from element. For example, the first trellis code has four states (states 0-3). The data input is a serial train of 4-PSK symbols (data 0-3). If the code is in state 0 and the input is '1', a '0' is transmitted from antenna #0 and a '1' is transmitted from antenna #1. Similarly, if the code is in state 2 and the input is a '1', a '2' is transmitted from antenna #0 and a '1' is transmitted from antenna #1. The other two codes presented have 8 and 16 states. They all achieve a transmit diversity order of 2, but with increasing coding complexity, the *coding gain* is significantly increased.

Figure 8 plots the frame error rate versus SNR for several trellis codes with varying number of states. The system has two transmitters and two receivers. All codes achieve diversity order 4 (2 on transmit and 2 on receive). However, note the additional gain in error rate with increasing complexity of code. This is due to the coding gain, as described in Eqn. (67) in Section 3.3.

The work of [10] has served as a huge step forward in the understanding of transmit diversity and the area of space-time coding. The paper describes the essential properties to look for in a proposed space-time code and provides certain codes with these properties. However, the paper also points out some significant limitations of space-time coding. The two theorems in the earlier section in particular point out that there is no free lunch. Data transfer and coding gain arise at the expense of encoding complexity. There is another significant problem with the approach developed so far - the trellis code needs to be designed, effectively, by brute force. Furthermore, each code must be designed for every possible data constellation used, i.e., changing from a 4-PSK to a 8-PSK required complete redesign of the code. Finally, trellis decoding is inherently complex with exponential complexity in the number of states.

These problems, coupled with the extremely simple and flexible (in terms of signal constellation) encoding block coding scheme proposed by Alamouti, led Tarokh and others to investigate generalize the block coding concept.

3.5 Space-Time Block Codes

We saw in Section 3.2 a simple scheme for space-time coding that allows for order-2 diversity. This section illustrates another approach to space-time coding, one based on encoding blocks of data (a block of 2 data symbols in the Alamouti scheme). The symbols are arranged in a such a way that, at the receiver, ML decoding can be performed individually on each symbol independent of the other symbols. The data symbols, effectively, are 'orthogonal' to each other. In [11], Tarokh et.al. develop the theory for orthogonal space-time block coding (OSTBC) for an arbitrary number of elements, allowing for extremely simple decoding with almost no growth in complexity.

The development of OSTBC is based on the theory of orthogonal designs for real symbols unearthed by Tarokh and his co-authors in [11]. The authors extend these designs to include complex symbols. The discussion summarizes some of the key results that leads to OSTBC. Consider again a communication system with N transmitters and M receivers. The data is encoded over L time slots. At the l-th time instant, the transmitter transmits data vector $\mathbf{c}_l$. The received data is

$$\mathbf{y}_l = \mathbf{H}\mathbf{c}_l + \mathbf{n}_l. \tag{73}$$

Over the L time instants, the ML decoder finds the solution to

$$\underline{\tilde{c}} = \min_{\underline{c}} \sum_{l=1}^{L} \sum_{m=1}^{M} \left| y_l^m - \sum_{n=1}^{N} h_{mn} c_l^n \right|^2 . \tag{74}$$

In a $N \times N$ orthogonal design, N data symbols $\{c_n,\}$ are transmitted over N time instants using the N elements. The ML decoder is equivalent to

$$\underline{\tilde{c}} = \min_{\{c_n\}} \sum_{n=1}^{N} e_n, \tag{75}$$

where e_n is some measure of error. The key is that e_n depends on c_n only, i.e., *each symbol can be decoded individually* [4]. Note that an orthogonal code has full rank in its corresponding error matrix. Hence, the transmit diversity order, determined by the rank of the error matrix as in Eqn. (60), is N.

3.5.1 Real Designs

The theory of OSTBC starts with real designs, assuming real data. A real orthogonal design is a $N \times N$ matrix of "indeterminates" made of N variables x_n or $-x_n$. Given a block of N symbols, the indeterminates are replaced with the corresponding symbols. At time instant l, the l-th row is transmitted over the N antenna elements. For example, a 2×2 orthogonal design is

$$\mathcal{O}_2 = \begin{pmatrix} x_1 & x_2 \\ -x_2 & x_1 \end{pmatrix}. \tag{76}$$

Given a block of two symbols s_1 and s_2, at the first time instant, s_1 is transmitted from the first element and s_2 from the second element. In the second instant, $-s_2$ is transmitted from the first element and s_1 from the second element. Note that $\mathcal{O}_2\mathcal{O}_2^T = \left(x_1^2 + x_2^2\right)\mathbf{I}_2$.

A *linear processing orthogonal design* $\mathcal{E}$ allows each row of the design to be linear combinations of the N indeterminates. Given a row vector of indeterminates, $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, the l^{th} row of the matrix $\mathcal{E}$ is given by $\mathbf{x}\mathbf{A}_l$ for some matrix $\mathbf{A}_l$. The authors show that any linear design is equivalent to another linear design $\mathcal{L}$ where

$$\mathcal{L}\mathcal{L}^T = \left(\sum_{n=1}^{N} x_n^2\right)\mathbf{I}_N. \tag{77}$$

Constructing an orthogonal design therefore becomes equivalent to constructing matrices $\mathbf{A}_l$, $l = 1, \ldots, N$ that satisfy this requirement. The theory unearthed by Tarokh et al. is the Hurwitz-Radon family of K matrices $\mathbf{B}_l$, $l = 1, \ldots, K$ that satisfy

$$\mathbf{B}_l^T\mathbf{B}_l = \mathbf{I} \qquad \mathbf{B}_l^T = -\mathbf{B}_l \qquad \mathbf{B}_l\mathbf{B}_m = -\mathbf{B}_m\mathbf{B}_l \tag{78}$$

The l-th row of a linear design $\mathcal{L}$ is given by $\mathbf{x}\mathbf{B}_l$. Note that to construct a square orthogonal design, we need $K = N - 1$ (in addition, the first row of the design set to be the data vector $\mathbf{x}$ itself, i.e. $\mathbf{B}_1 = \mathbf{I}$).

[4] See [11] for the expression detailing e_n.

According to the theory of Hurwitz and Radon, if $N = 2^a b$ with b odd and $a = 4c + d$ with $c, d \geq 0$, the number of matrices in this family satisfies, $K < \rho(N) = 8c + 2^d$. Therefore, the only numbers that satisfy $K = N - 1$ are $N = 2$, 4, or 8. Therefore, only for a few specific cases is it possible to construct a $N \times N$ orthogonal design (a square design).

The authors then generalize linear designs to non-square matrices. A block of K data symbols (and possibly their negatives) are arranged in a linear $L \times N$ design. The l^{th} row of this matrix is transmitted in the l^{th} time slot over the N elements. The rate of this code is clearly $R = K/L$. A generalized linear design $\mathcal{G}$ satisfies

$$\mathcal{G}\mathcal{G}^T = \left(\sum_{k=1}^{K} x_k^2 \right) \mathbf{I} \tag{79}$$

Clearly a desirable property is to minimize L for a given rate R. This would minimize the block size in the encoding and decoding process. Denote as $A(R, N)$ the minimum L for which a linear design exists. The authors show that *for real symbols there exists a rate-1 linear design for any N.* For example,

$$\mathcal{G}_3 = \begin{pmatrix} x_1 & x_2 & x_3 \\ -x_2 & x_1 & -x_4 \\ -x_3 & x_4 & x_1 \\ -x_4 & -x_3 & x_2 \end{pmatrix}, \tag{80}$$

which transmits $K = 4$ symbols over $L = 4$ time instants using $N = 3$ antennas.

3.5.2 Complex Designs

So far, we have focused on real designs only, i.e., designs for real symbols. Clearly, in a communication system we are also interested in designs for complex symbols. A complex design uses as indeterminates symbol manipulations of the form $\pm x$, $\pm x^*$, $\pm jx$, $\pm jx^*$ where the superscript * represents the complex conjugate.

Some important results:

- Any complex design is equivalent to a real design with $x_n (= x_n^r + jx_n^i)$ replaced with $\begin{pmatrix} x_n^r & x_n^i \\ -x_n^i & x_n^r \end{pmatrix}$.
- Complex designs exist for $N = 2$ only.
- Any complex linear design is equivalent to a linear design $\mathcal{L}_c$ such that each entry in $\mathcal{L}_c$ is a linear combination of x_n or its conjugates.

- A $L \times N$ generalized linear design satisfied $\mathcal{G}_c \mathcal{G}_c^H = \left(\sum_{k=1}^{K} |x_k|^2 \right) \mathbf{I}$.
- *Full rate linear designs exist for N=2 only.*
- The maximum rate that can be guaranteed for an arbitrary N is $R = 1/2$. However, rate-3/4 "sporadic" codes exist for $N = 3$ and $N = 4$ and are provided in [11]. For example, the rate-3/4 code for $N = 3$ and $N = 4$ are

$$\mathcal{G}_3 = \begin{pmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} \\ \frac{x_3^*}{\sqrt{2}} & \frac{x_3^*}{\sqrt{2}} & \frac{-x_1 - x_1^* + x_2 - x_2^*}{2} \\ \frac{x_3^*}{\sqrt{2}} & -\frac{x_3^*}{\sqrt{2}} & \frac{x_2 + x_2^* + x_1 - x_1^*}{2} \end{pmatrix}, \tag{81}$$

and

$$\mathcal{G}_4 = \begin{pmatrix} x_1 & x_2 & \frac{x_3}{\sqrt{2}} & \frac{x_3}{\sqrt{2}} \\ -x_2^* & x_1^* & \frac{x_3}{\sqrt{2}} & -\frac{x_3}{\sqrt{2}} \\ \frac{x_3^*}{\sqrt{2}} & \frac{x_3^*}{\sqrt{2}} & \frac{-x_1 - x_1^* + x_2 - x_2^*}{2} & \frac{-x_2 - x_2^* + x_1 - x_1^*}{2} \\ \frac{x_3^2}{\sqrt{2}} & -\frac{x_3^*}{\sqrt{2}} & \frac{x_2 + x_2^* + x_1 - x_1^*}{2} & \frac{x_1 + x_1^* + x_2 - x_2^*}{2} \end{pmatrix}, \tag{82}$$

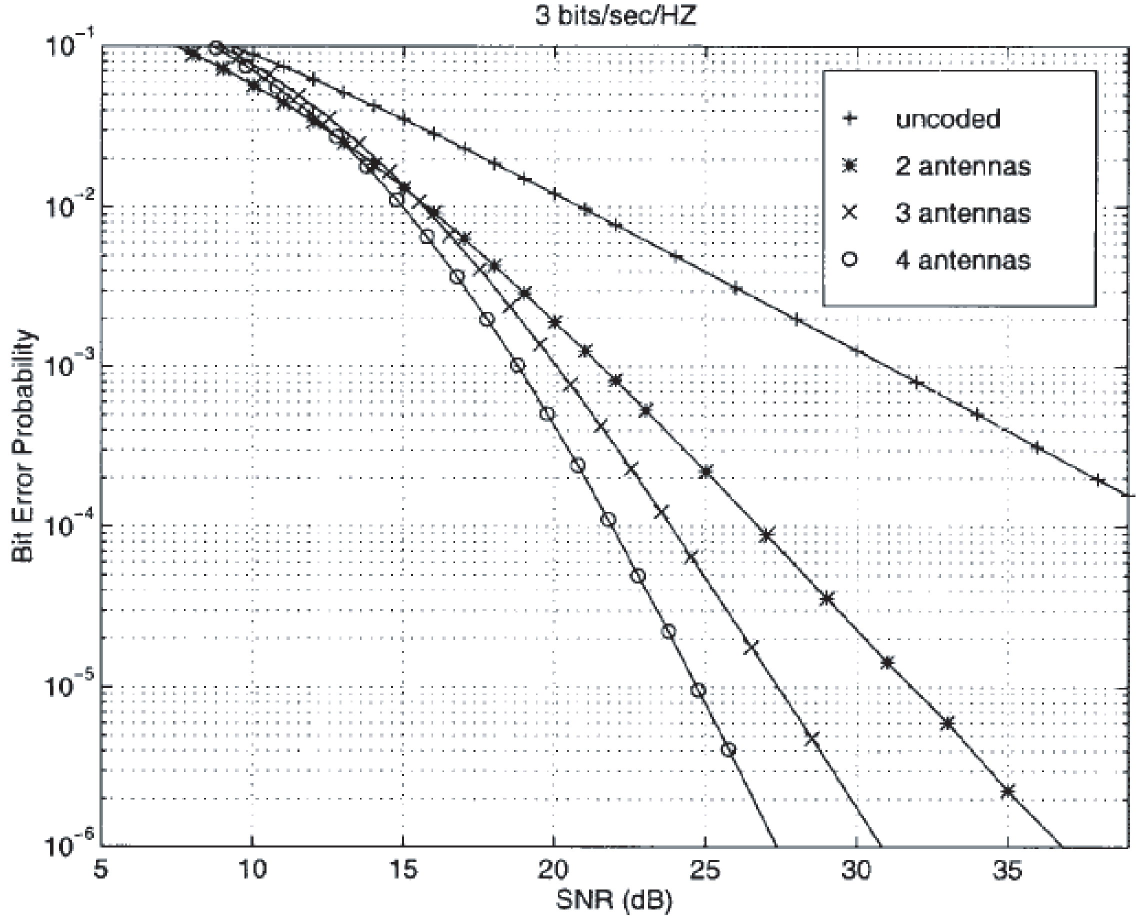
respectively.

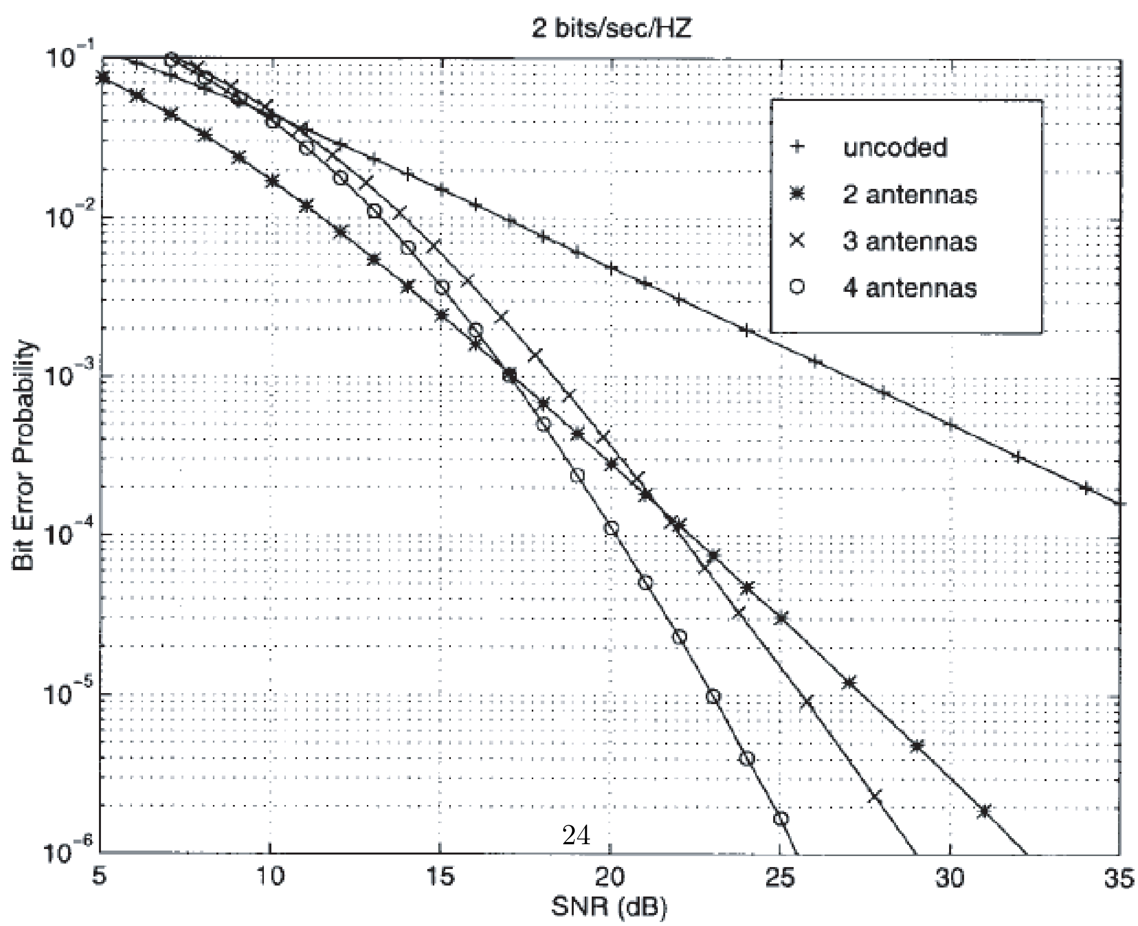
In 2001, an alternative $\mathcal{G}_4$ matrix was proposed in [12]:

$$\mathcal{G}_4 = \begin{pmatrix} x_1 & 0 & x_2 & -x_3 \\ 0 & x_1^* & x_3^* & x_2^2 \\ -x_2^* & -x_3 & x_1^* & 0 \\ x_3^* & -x_2 & 0 & x_1^* \end{pmatrix}, \tag{83}$$

which interestingly may "save" on power due the inclusion of some zeros in the transmission.

The performance of OSTBC codes are presented in a companion paper [13]. Figure 9 plots the BER versus SNR for two cases - rate of 3 b/s/Hz and 2b/s/Hz. For two antennas, as per the theorem in [10] and in Section 3.4, achieving this rate requires constellations of 8 and 4 elements


3 bits/sec/HZ
uncoded
2 antennas
3 antennas
4 antennas
Bit Error Probability
SNR (dB)


2 bits/sec/HZ
uncoded
2 antennas
3 antennas
4 antennas
Bit Error Probability
SNR (dB)

(8-PSK and 4-PSK here) respectively. However, for 3 and 4 antennas, achieving 3b/s/Hz requires a constellation of 16-elements (16-QAM here) since one can only achieve a rate-3/4 code. Similarly, for 2b/s/Hz, the transmitter uses 16-QAM with a rate-1/2 code.

Note that in the second figure, due to the greater number of constellation points to achieve the same rate, using only two elements does better than 3 or 4 elements even for up to a BER of 10^{-3}! This is because to satisfy the budget of a maximum total transmit energy of E_s, with three elements the signal from each element is transmitted with energy $E_s/3$. However, note that the three and four element schemes do achieve diversity orders of 3 and 4 respectively.

One interesting point listed above is that orthogonal complex rate-1 codes exist for $N = 2$ only. Alamouti's code, the orthogonal code for $N = 2$ is therefore unique in its simplicity and efficiency. This issue and the work in [11] has set off a lot of research in STBC that has higher spectral efficiency, efficiencies close to unity. In general, these codes trade-off spectral efficiency for complexity in decoding.

4 Tradeoffs Between Reliability and Throughput

So far we have analyzed the use of MIMO systems from two opposing points of view - how fast we can transmit data (capacity in Section 2) and reliability (transmit diversity in Section 3). Admittedly, unlike with transmit diversity, we have not developed any practical schemes that could achieve the promised data throughput (which is probably impossible, but we haven't even attempted to come close). In particular, as with capacity a good practical scheme should be able to achieve the linear gains (of order $\min(M, N)$) in data throughput. However, intuitively we understand that in transmitting data faster we will have to sacrifice reliability. This section discusses schemes to achieve greater throughput, before concluding a statement of the fundamental tradeoff between throughput and reliability.

In any case, for arrays with more than very few elements constructing a trellis code is difficult and ML decoding becomes prohibitively complex. We also saw that efficient and simple block codes exist only for $N \leq 4$.

4.1 Combined Beamforming and STC

In [14], Tarokh et al. present a scheme that represents a tradeoff between data throughput and diversity order. The scheme is based transmitting multiple data streams while limiting the coding complexity but yet exploiting all available spatial degrees of freedom. Consider a transmitting array of N elements divided into Q groups of sub-arrays, each with N_q elements, i.e., $\sum_{q=1}^{Q} N_q = N$. At the input, B bits are divided into Q blocks of B_q bits each. Using a low-complexity encoder (denoted as C_q, the B_q bits are encoded, yielding N_q symbols per time slot for transmission using N_q elements. The overall code can be represented as $C_1 \times C_2 \times \ldots \times C_Q$. Note that each code may itself be a trellis or block code. Let $\mathbf{c}_q$ denote the data transmitted using the q^{th} subarray.

The receiver decodes each block of data successively. Without loss of generality we start with decoding the first block. The data received at the M receiving elements is given by

$$\mathbf{y} = \mathbf{H}\mathbf{c} + \mathbf{n}, \tag{84}$$

$$= \left[\begin{array}{cccc|ccc} h_{11} & h_{12} & \cdots & h_{1N_1} & h_{1(N_1+1)} & \cdots & h_{1N} \\ h_{21} & h_{22} & \cdots & h_{2N_1} & h_{2(N_1+1)} & \cdots & h_{2N} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN_1} & h_{M(N_1+1)} & \cdots & h_{MN} \end{array}\right] \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_Q \end{bmatrix} + \mathbf{n}, \tag{85}$$

where the channel matrix $\mathbf{H}$ is partitioned to isolate the channel from the first subarray to the receiver. Only N_1 of the N transmitters are of interest. The data transmissions $\mathbf{c}_2$ to $\mathbf{c}_Q$ (through the channels from their respective transmitters) act as interference and degrade the decoding of the data in the first block, $\mathbf{c}_1$.

Denote as $\mathbf{\Omega}(C_1)$ the $M \times N_1$ channel matrix from the N_1 elements in the first subarray to the M receive elements. Similarly denote as $\mathbf{\Lambda}(C_1)$ the $M \times (N - N_1)$ matrix of the "interfering" channel[5], i.e.

$$\mathbf{\Omega}(C_1) = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1N_1} \\ h_{21} & h_{22} & \cdots & h_{2N_1} \\ \vdots & \vdots & \cdots & \vdots \\ h_{M1} & h_{M2} & \cdots & h_{MN_1} \end{bmatrix}, \qquad \mathbf{\Lambda}(C_1) = \begin{bmatrix} h_{1(N_1+1)} & h_{1(N_1+2)} & \cdots & h_{1N} \\ h_{2(N_1+1)} & h_{2(N_1+2)} & \cdots & h_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ h_{M(N_1+1)} & h_{M(N_1+2)} & \cdots & h_{MN} \end{bmatrix}. \tag{86}$$

Since $\mathbf{\Lambda}(C_1)$ has $N - N_1$ columns, if $M > N - N_1$, $\text{rank}(\mathbf{\Lambda}(C_1)) \leq N - N_1$ and there exist at least $M - (N - N_1) = (M - N + N_1)$ vectors that are orthogonal to all $(N - N_1)$ vectors in $\mathbf{\Lambda}(C_1)$. Denote as $\mathbf{\Theta}(C_1)$ as the $M \times (M - N + N_1)$ matrix whose columns are orthogonal to $\mathbf{\Lambda}(C_1)$. In addition, one can assume the columns of $\mathbf{\Theta}(C_1)$ are mutually orthonormal. Note that due to the constraint $M > N - N_1$, it is always possible to create matrix $\mathbf{\Theta}(C_1)$ given the channel matrix $\mathbf{H}$. Now consider a new set of "received" data

$$\tilde{\mathbf{y}}_1 = \mathbf{\Theta}(C_1)^H \mathbf{y}, \tag{87}$$

$$= \mathbf{\Theta}(C_1)^H \mathbf{\Omega}(C_1)\mathbf{c}_1 + \mathbf{\Theta}(C_1)^H \mathbf{\Lambda}(C_1) \begin{bmatrix} \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_Q \end{bmatrix} + \mathbf{\Theta}(C_1)^H \mathbf{n}, \tag{88}$$

$$= \mathbf{\Theta}(C_1)^H \mathbf{\Omega}(C_1)\mathbf{c}_1 + \mathbf{\Theta}(C_1)^H \mathbf{n} \tag{89}$$

[5]The notation is in a large part from Tarokh et.al. [14]

The noise term $\mathbf{\Theta}(C_1)^H\mathbf{n}$ is also zero-mean and satisfies $\mathrm{E}\left\{\mathbf{\Theta}(C_1)^H\mathbf{n}\mathbf{n}^H\mathbf{\Theta}(C_1)\right\} = \sigma^2\mathbf{\Theta}(C_1)^H\mathbf{\Theta}(C_1) = \sigma^2\mathbf{I}_{M-N+N_1}$, i.e., the noise term is still white. Note that all the interference has been eliminated.

This final equation is that of an equivalent space-time coded communication system with N_1 transmitters and $(M - N + N_1)$ receivers. On the data in $\mathbf{c}_1$ we therefore get a diversity order of $N_1(M - N + N_1)$, which is significantly lower than the potential diversity gain for a single data stream of MN. Note that the achieved diversity order makes sense since the data stream is transmitted using N_1 antennas and the M receivers must use $(N - N_1)$ degrees of freedom to suppress the $(N-N_1)$ transmissions, leaving $(M-N+N_1)$ degrees of freedom to enhance reliability (diversity) on the N_1 transmissions of interest.

Clearly one could repeat the process and achieve $N_q(M - N + N_q)$ diversity order. However, there is a more intelligent way of processing the same data. Assume the data in $\mathbf{c}_1$ is accurately decoded. At the receiver, we now know the transmitted data from the first subarray and the channel $\mathbf{\Omega}(C_1)$ and so when decoding $\mathbf{c}_2$, one can *eliminate* this source of interference from the data. Let

$$\mathbf{y}_2 = \mathbf{y} - \mathbf{\Omega}(C_1)\mathbf{c}_1 = \mathbf{\Lambda}(C_1)\begin{bmatrix}\mathbf{c}_2\\ \mathbf{c}_3\\ \vdots\\ \mathbf{c}_Q\end{bmatrix} + \mathbf{n} \tag{90}$$

This is equivalent to a communication system with M receivers and $N - N_1$ transmitters (or $N - N_1 - N_2$ interfering sources). To decode $\mathbf{c}_2$, only N_2 of these transmitters are of interest. Using the same scheme as described above, one can achieve $N_2\left(M - (N - N_1 + N_2)\right) = N_2(M - N + N_1 + N_2)$-order diversity. This is *greater* than the $N_2(M - N + N_2)$ order diversity attained using the "obvious" approach.

Repeating this process for each subarray, to decode $\mathbf{c}_q$ one can achieve diversity order $N_q\left(M - N + \sum_{p=1}^{q} N_p\right)$, i.e., each successive data block can be decoded with greater diversity order. This also leads to a variant on this scheme - since each block gets a different order of diversity, one can transmit these blocks with different power levels to achieve somewhat equal error rates. In [14] the authors suggest a power allocation in inverse proportion to diversity order.

IMP: The famous Bell Labs Layered Space-Time (BLAST) scheme [15] uses $N_q = 1$, with the maximum data throughput but minimum diversity. The scheme is impractical in $M \geq N$. In fact, it should be noted that the BLAST scheme was known before Tarokh's paper was published and probably inspired this work.

The scheme presented in [14] is flexible in that it allows for a tradeoff between throughput (multiplexing) and required reliability. However, it suffers from one significant drawback - the need for an adequate number of antennas $M > N - N_1$ such that a null space can be formed to cancel the interference. As mentioned earlier, in cellular communications, one could expect multiple antennas at the base station, however, a mobile device would not have many antennas. Thinking in terms of the downlink, expecting $M > N - N_1$ may not be realistic. Note that BLAST requires $M \geq N$.

An alternative analysis of the BLAST scheme (and one that clearly identifies the successive interference cancellation structure of the scheme) uses the QR decomposition. Consider the simple case of $M = N$. The transmission transmits N data streams in parallel ($N_q = 1$). Let the transmitted vector be $\mathbf{c} = [c_1 c_2 \dots c_N]$. The received signal is

$$\mathbf{y} = \mathbf{Hc} + \mathbf{n}. \tag{91}$$

Since $M = N$, $\mathbf{H}$ is square and we can write $\mathbf{H} = \mathbf{QR}$ where $\mathbf{Q}$ is unitary and $\mathbf{R}$ is upper triangular. Therefore,

$$\begin{aligned} \mathbf{y} &= \mathbf{QRc} + \mathbf{n}, \\ \tilde{\mathbf{y}} = \mathbf{Q}^H \mathbf{y} &= \mathbf{Rc} + \tilde{\mathbf{n}}, \end{aligned} \tag{92}$$

where $\tilde{\mathbf{n}} = \mathbf{Q}^H \mathbf{n}$ is the noise term with the same statistics as the original noise term $\mathbf{n}$ (since $\mathbf{Q}$ is unitary).

Since $\mathbf{R}$ is upper triangular, note that the M^{th} data stream effectively "sees" a SISO channel. Once this symbol has been decoded it can be cancelled before decoding the $(M-1)^{\text{th}}$ stream. Note that since the M^{th} data stream sees a SISO channel, the diversity order is 1.

4.2 Linear Dispersion Codes

In 2002 Hassibi and Hochwald [16] published their work on *linear dispersion* codes, a variation on block codes, to address the problem of flexibility in terms of system design, data rate coupled with diversity order. LD codes have some very useful properties:

- BLAST (as described above) and STBC are special cases of LD codes, though LD codes generally outperform both while being simple to encode.
- LD codes are extremely flexible and can be used for any values of M and N.
- They can be decoded using the successive interference cancellation technique described above (in Section 4.1) and others such as sphere decoding.
- LD codes are designed to maximize the mutual information between input and output (closely tied to capacity)

LD codes have one other property that the authors present as desirable, but may not be - the codes are designed with the specific values of M and N in mind, i.e., changing the number of receive antennas implies changing the code used. In practice the same transmitter may be communicating with many types of receivers - or receivers with different numbers of antennas. Using a different code for each receiver is impractical. The authors propose designing the code for a minimum number of receive antennas; however, that would in practice be $M = 1$ thereby significantly reducing the data throughput gains provided by LD codes.

Let us begin by investigating the information theoretic aspects of Alamouti's STBC for two transmit elements. Remember that Eqn. (36) tells us that a MIMO system with N transmit and M receive antennas has capacity,

$$C = \mathrm{E}\left\{\log_2 \det\left(\mathbf{I}_M + \frac{E_s}{N\sigma^2}\mathbf{H}\mathbf{H}^H\right)\right\}. \tag{93}$$

With $N = 2$ and $M = 1$, $\mathbf{H} = [h_0,\ h_1]$ and $C = \mathrm{E}\left\{\log_2\left(1 + \frac{E_s}{2\sigma^2}\left[|h_0|^2 + |h_1|^2\right]\right)\right\}$.

In using the Alamouti scheme, transmitting symbols s_0 and s_1 over two time slots, Eqn. (49) indicates that the received data can be written as

$$\Rightarrow \begin{bmatrix} y_0 \\ y_1^* \end{bmatrix} = \sqrt{\frac{E_s}{2}} \begin{bmatrix} h_0 & h_1 \\ h_1^* & -h_0^* \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} + \begin{bmatrix} n_0 \\ n_1^* \end{bmatrix}, \tag{94}$$

$$\Rightarrow \mathbf{y} = \mathcal{H} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} + \mathbf{n}, \tag{95}$$

where $\mathcal{H}\mathcal{H}^H = \left[|h_0|^2 + |h_1|^2\right]\mathbf{I}_2$. This final equation suggests that the mutual information between the input and the output for Alamouti's scheme is

$$C_{\text{Alamouti}} = \frac{1}{2}\mathrm{E}\left\{\log_2 \det\left(\mathrm{I}_2 + \frac{E_s}{2\sigma^2}\mathcal{H}\mathcal{H}^H\right)\right\}, \tag{96}$$

where we use C that seems to indicate capacity, but in reality is only the maximum mutual information between the input and output using Alamouti's scheme. The factor of half is because of the fact that we transmit the two symbols over two time slots. Using the fact that the determinant of a matrix is the product of its eigenvalues,

$$C_{\text{Alamouti}} = \mathrm{E}\left\{\log_2\left(1 + \frac{E_s}{2\sigma^2}\left[|h_0|^2 + |h_1|^2\right]\right)\right\}, \tag{97}$$

implying that if the encoder used to obtain symbols s_0 and s_1 is capacity achieving, Alamouti's code for $N = 2$ and $M = 1$ is also achieves capacity!

However, if one is using Alamouti's code for $N = 2$ and $M = 2$, the maximum mutual information between the input symbols and output can be shown to be

$$C_{\text{Alamouti}} = \mathrm{E}\left\{\log\left(1 + \frac{2E_s}{4\sigma^2}\left[|h_{00}|^2 + |h_{01}|^2 + |h_{10}|^2 + |h_{11}|^2\right]\right)\right\}, \tag{98}$$

where h_{mn} is the channel from antenna $\#n$ to antenna $\#m$. This expression is the capacity of $N = 4$, $M = 1$, not $N = 2, M = 2$. Due the linear gains in capacity with $N = 2, M = 2$, Alamouti's space-time code does not achieve capacity for $N = 2, M = 2$.

A similar analysis of the BLAST scheme as described in Section 4.1 indicates that the BLAST scheme achieves capacity since it transmits N data streams, but provides no diversity (the reliability of a SISO channel).

4.2.1 Linear Dispersion Code Design

Consider a linear code that transmits K data symbols over L time symbols, i.e., achieves a rate of $R = K/L$. The K complex symbols are $s_k = \alpha_k + j\beta_k$, $k = 1, \ldots, K$. The linear dispersion code is a $L \times N$ matrix $\mathcal{S}$ given by

$$\mathcal{S} = \sum_{k=1}^{K} \left(\alpha_k \mathbf{A}_k + j\beta_k \mathbf{B}_k \right), \tag{99}$$

where $\mathbf{A}_k$ and $\mathbf{B}_k$ are $L \times N$ matrices that define the code. For example, Alamouti's code uses $K = L = N = 2$ with

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ \mathbf{B}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \ \mathbf{B}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{100}$$

In time slot l the l^{th} row of $\mathcal{S}$ is transmitted over the N transmit antennas. In the l^{th} time slot, the received data is

$$\mathbf{x}(l) = \mathbf{H}\mathbf{s}(l) + \mathbf{n}_l, \tag{101}$$

where $\mathbf{s}(l)$ is transmitted at time slot l. Since $\mathbf{s}(l)$ is linearly dependent on the data symbols (represented in its real and imaginary parts α_k and β_k), if we were to stack the real and imaginary parts of $\mathbf{x}(l)$, $l = 1, \ldots, L$,

$$\mathbf{x} = \begin{bmatrix} x_{r1}(1) \\ x_{i1}(1) \\ x_{r2}(1) \\ x_{i2}(1) \\ \vdots \\ x_{rM}(1) \\ x_{iM}(1) \\ \hline x_{r1}(2) \\ x_{i1}(2) \\ \vdots \\ x_{rM}(2) \\ x_{iM}(2) \\ \hline x_{r1}(L) \\ x_{i1}(L) \\ x_{r2}(L) \\ x_{i2}(L) \\ \vdots \\ x_{rM}(L) \\ x_{iM}(L) \end{bmatrix}_{2ML \times 1} = \mathcal{H} \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \vdots \\ \alpha_K \\ \beta_K \end{bmatrix}_{2K \times 1} + \mathbf{n}, \tag{102}$$

where $x_{rm}(l)/x_{im}(l)$ is the real/imaginary part of the received data at the m^{th} element at the l^{th} time slot. The $2ML \times 2K$ effective channel matrix $\mathcal{H}$ is dependent on the code matrices $\mathbf{A}_k$ and $\mathbf{B}_k$, $k = 1, \ldots, K$.

The linear dispersion code maximizes the mutual information between the input data and the output, given by

$$\{\mathbf{A}_k, \mathbf{B}_k\} = \arg \max_{\mathbf{A}_k, \mathbf{B}_k, k=1,\ldots,K} \left[\frac{1}{2L} \log_2 \det \left(\mathbf{I}_{2ML} + \frac{E_s}{N\sigma^2} \mathcal{H}\mathcal{H}^H \right) \right], \tag{103}$$

where the factor of half is due to the fact that we are now using real data (we have separated out the real and imaginary parts). This optimization problem must be constrained to ensure that the transmit energy is bounded. The constraint is

$$\sum_{k=1}^{K} \operatorname{tr}\left(\mathbf{A}_k \mathbf{A}_k^H + \mathbf{B}_k \mathbf{B}_k^H \right) \leq 2NL. \tag{104}$$

The solution to this constrained optimization problem defines the code. The authors also show that

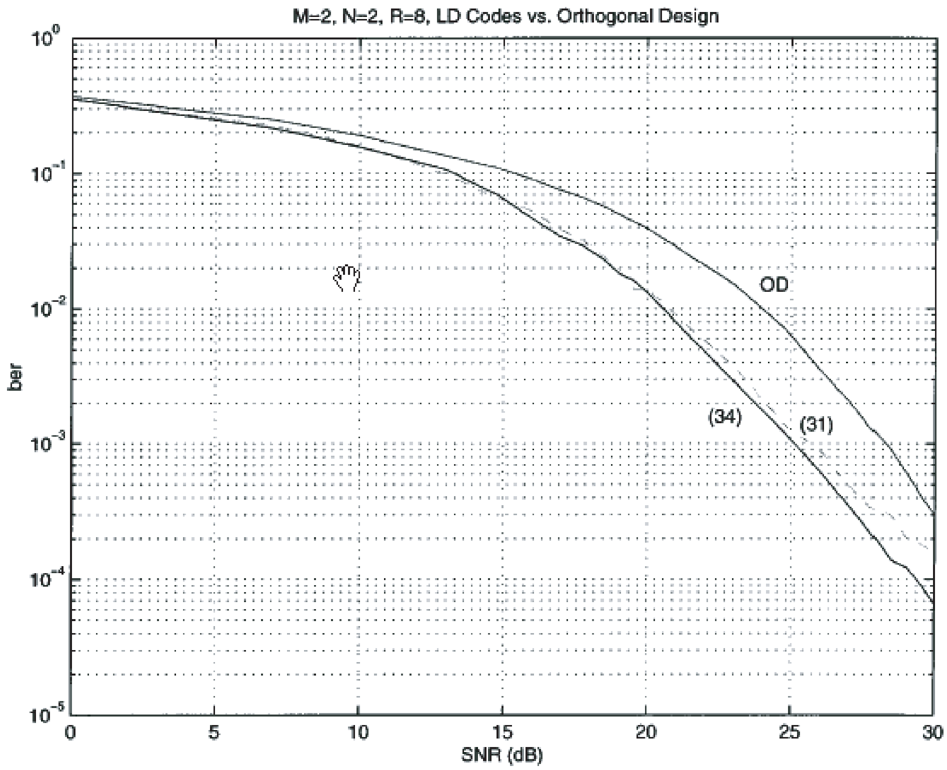


Figure 10: Comparing the performance of Alamouti and linear dispersion space-time codes.

the average pairwise error probability ($P_e(\text{avg.}) \leq C_{\max}^{-0.5}$), i.e., maximizing the mutual information also enables reliability.

Numerical Examples: The design methodology described above is tested against the Alamouti code for the case of $N = 2$, $M = 2$ and $R = 8$bits/channel use. The Alamouti code must therefore use a 256-point constellation (here 256-QAM). The dispersion code, designed for $K = 4$ and $L = 2$, uses a 16-QAM constellation. Figure 10 compares the performance of the two space-time codes. The line marked "OD" corresponds to the orthogonal Alamouti design whereas the lines marked (31) and (34) are two variations on the LD code design. Clearly both codes achieve order-2 diversity though both LD codes perform significantly better than Alamouti's code. Note that the there is a tradeoff in that ML decoding of the Alamouti code requires only symbol by symbol decoding where the LD code requires successive interference cancellation or sphere decoding.

The LD design methodology is also compared to the BLAST scheme for the case of $N = 2$, $M = 2$, $R = 4$ bits/channel use. The design uses $K = 4$. Figure 11 plots the results of using the two transmission schemes. The top two curves correspond to the BLAST scheme - the worse performance corresponds to the case (described above) with the successive cancellation and nulling which suffers from error propagation. The lower curve corresponds to ML decoding. The two

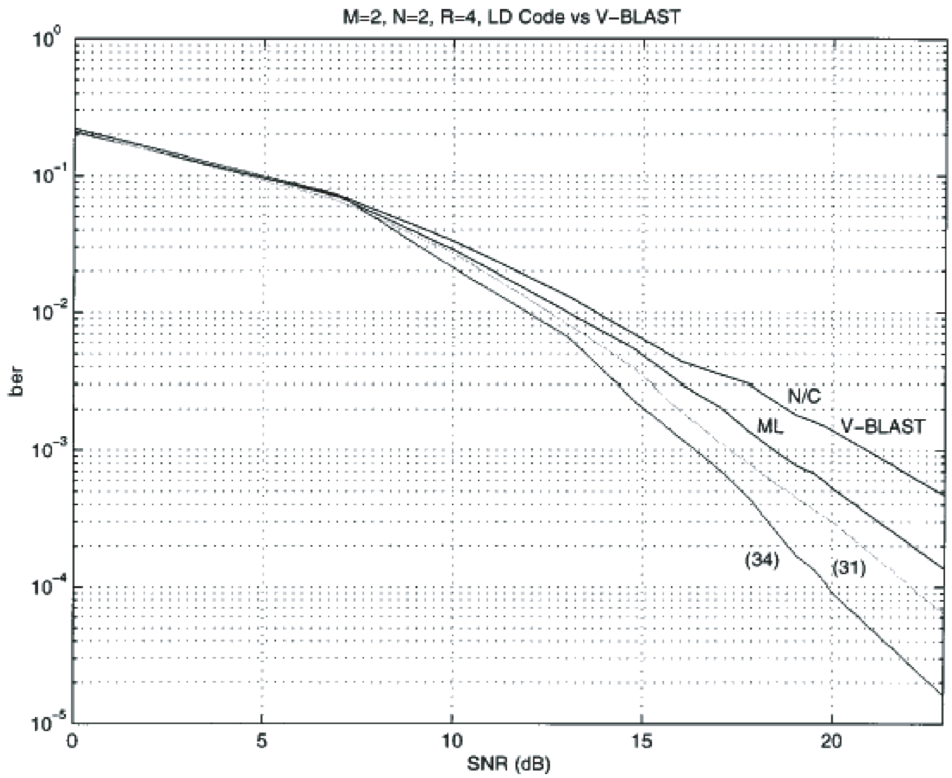


Figure 11: Comparing the performance of BLAST and linear dispersion space-time codes.

better curves correspond to the LD codes with sphere decoding[6]. Clearly, again, LD codes perform significantly better than the BLAST scheme. Note specifically that the LD codes achieve a greater diversity order.

In summary, linear dispersion codes, designed to maximize the mutual information between input and output ,with the specific values of M, N and rate in mind significantly outperform other well known codes. If required to deal with several types of receivers with varying numbers of receive antennas, the authors suggest to use a design for the minimum numbers of receive antennas. However, this, of course, results in some performance loss.

5 Diversity-Multiplexing Tradeoff

We end this chapter by stating the *fundamental tradeoff*, stated by Zheng and Tse, between diversity order and data throughput (also multiplexing) [17]. We have seen that given a diversity order of d the probability of error in Rayleigh channels is given by $P_e \propto \text{SNR}^{-d}$ at high SNR. Similarly, with a system using N transmit and M receive antennas, the capacity of the system raises (in the high SNR limit) as $C \propto \min(M, N) \log(\text{SNR})$.

[6] Outside the scope of this course. A sub-optimal efficient decoding technique.

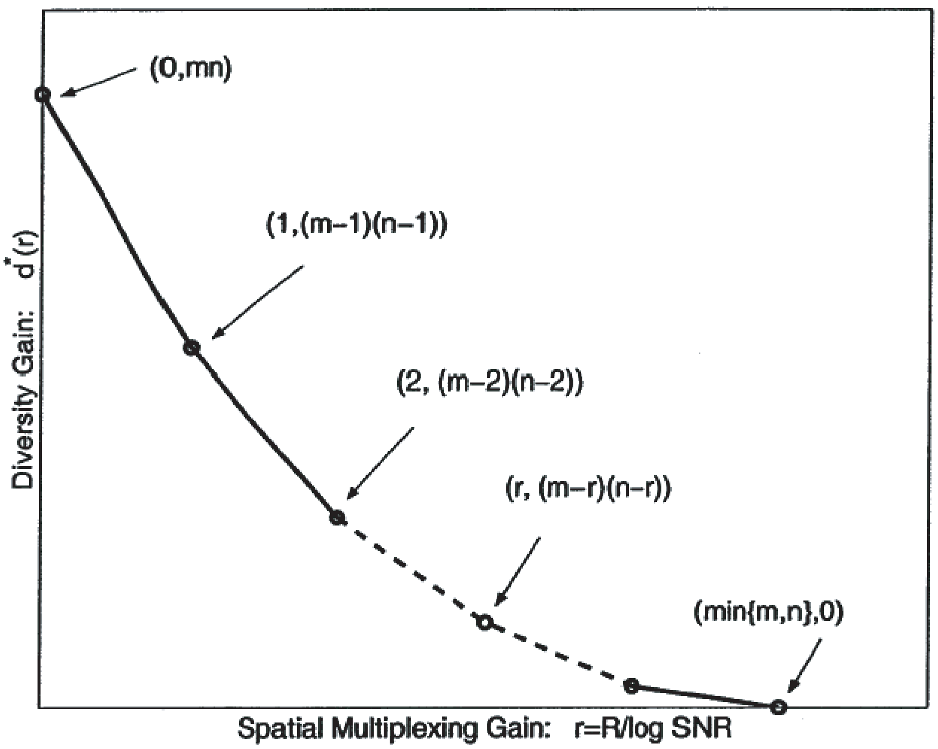


Figure 12: Illustrating the diversity-multiplexing tradeoff.

Note that the capacity expression indicates that to achieve capacity one must transmit data faster as the SNR increases. The authors define a scheme $\mathcal{C}(\mathrm{SNR})$ to achieve diversity gain d and *multiplexing gain* r if

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} = -d \tag{105}$$

$$\lim_{\mathrm{SNR}\to\infty} \frac{R(\mathrm{SNR})}{\log \mathrm{SNR}} = r \tag{106}$$

where $R(\mathrm{SNR})$ is the data throughput of the scheme. Note that the scheme assumes that as the SNR increases the data rate rises as well (possibly through adaptive modulation). Any scheme with a fixed data rate, however high, is said to have a *multiplexing gain of zero*. In this regard, all schemes we have developed so far have a multiplexing gain of zero. The thinking behind such a definition is that as the SNR increases one could either gain reliability (reduce P_e) or increase throughput or part of each. The definition determines how much of each we gain as SNR increases.

Let $d^\star(r)$ be the supremum of all possible schemes with multiplexing gain r. Clearly $d_{\max} = d^\star(0)$ and $r_{\max} = \sup(r : d^\star(r) > 0)$. The authors state the most interesting result:

Diversity Multiplexing Tradeoff: Consider the case of N transmit and M receive antennas. For a block length $L > M + N - 1$, the optimal tradeoff curve, $d^\star(r)$ is given by the piecewise-linear function connecting the points $(r, d^\star(r))$, $r = 0, 1, \ldots, min(M, N)$, where

$$d^\star(r) = (M - r)(N - r). \tag{107}$$

In particular, $d_{\max} = MN$ and $r_{\max} = \min(M, N)$.

Figure 12 illustrates the fundamental tradeoff between reliability (diversity order) and data rate (multiplexing). The curve is piecewise linear joining the points $(k, d^\star(k))$ with $d^\star(k)$ defined in Eqn. (107). The theorem states that it is not possible to use the available SNR in any manner better than this curve. Note that as expected at zero multiplexing gain, the diversity order that can be achieved is MN whereas if one where increasing the data throughput as $\min(M, N) \log(\mathrm{SNR})$, there is no diversity, i.e., the error rate *does not* fall with SNR.

Equation (107) suggests that at the integer points, when the multiplexing rate is r, the system communicates r parallel data streams. The transmitter and receive each use r degrees of freedom to eliminate the inter-stream interference leaving a diversity order of $(M - r)(N - r)$. Another interpretation uses the eigenvalues of the channel matrix [18]. A multiplexing rate of r says that the raw data rate of transmission is $R = r \log(SNR)$. An outage occurs when the mutual information falls below this rate. The mutual information is given by

$$C = \log_2 \det \left(\mathbf{I}_M + \frac{E_s}{N\sigma^2} \mathbf{H}\mathbf{H}^H \right) = \sum_{m=1}^{\min(N,M)} \log_2 \left(1 + \frac{E_s}{N\sigma^2} \lambda_m \right), \tag{108}$$

where $\lambda_m, m = 1, \ldots M$ are the eigenvalues of $\mathbf{H}\mathbf{H}^H$.

At high SNR, we can ignore the "1 + ". For an outage to occur we need

$$\sum_{m=1}^{\min(N,M)} \log_2 \left(\frac{SNR}{N} \lambda_m \right) < r \log_2(SNR). \tag{109}$$

The outage events are therefore controlled by the eigenvalues of $\mathbf{H}\mathbf{H}^H$. For an outage to happen, these eigenvalues have to be "bad" (small).

- If r is close to zero, an outage occurs if *all eigenvalues are poor*. This happens rarely (and yields diversity order MN).
- If $r = \min(N, M)$, we need *all eigenvalues to be large* to avoid an outage (note that the number of the 'large' eigenvalues effectively provides the linear term in front of the $\log_2(SNR)$.)
- Somewhere in between, to avoid an outage, we need r of the eigenvalues to be large enough, resulting in a diversity order of $(M - r)(N - r)$.

6 Summary

This chapter has covered a lot of ground. We started by developing the information theory of multiple input multiple output systems. The key was that in fading channels, the gains in capacity were linear in the number of *independent* channels between transmitter and receiver. The maximum gain is on order of $\min(N, M)$. So while all of this is very exciting, note that there are some significant costs that must be met to achieve these great gains. This set the theoretical limit on how fast we could transmit in a MIMO channel.

We then investigated transmit diversity - using the multiple inputs and outputs for reliable communications. In this regard, the work of Tarokh [10, 11, 14] proved crucial. We began with the extremely simple Alamouti code [9]. We showed that transmit diversity is possible, though requires use of the time dimension as well. We then asked the question - what makes a particular code good? In [10], the authors develop two criteria for good codes: the error matrix in Eqn. (60) must be full rank (diversity order) and the product of the eigenvalues must be maximized (to achieve coding gain). However, they also show that for large N, the coding must necessarily become prohibitively complex. In [11], the authors generalize the Alamouti space-time block coding scheme for an arbitrary number of elements. They show that it is not possible to have rate-1 orthogonal codes for $N > 2$. However, for $N = 3$ or $N = 4$, rate-3/4 orthogonal codes are possible. Finally, in [14], Tarokh et.al. generalize the BLAST concept to combine the benefits of space-time coding while yet using all available degrees of freedom.

We also investigated other schemes achieve a specific rate while attempting to maximize diversity. The linear dispersion codes of Hassibi and Hochwald allowed for a simple framework to design such codes. We wrapped up this chapter by stating a fundamental tradeoff between reliability (diversity order) and throughput (multiplexing gain).

7 Acknowledgements

My sincere thanks to Prof. David Tse [4] of U.C. Berkeley and Prof. Christian Schlegel [2] for access to their class notes covering similar material.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* John Wiley, 1991.

[2] C. Schlegel, “Statistical communication theory.” Lecture Notes for ELEN 7950-4, University of Utah, Spring 2002.

[3] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, June 2002.

[4] D. Tse and P. Viswanath, "Fundamentals of wireless communication." Lecture Notes for EE 290S, U.C. Berkeley, Fall 2002.

[5] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, November 1999.

[6] S. Haykin, *Adaptive Filter Theory.* Prentice Hall, 4th ed., 2002.

[7] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Annals of Mathematical Statistics,*, vol. 35, pp. 475–501, 1964.

[8] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[9] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Select Areas in Communications*, vol. 16, pp. 1451–1458, August 1998.

[10] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Spacetime codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[11] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Spacetime block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[12] G. Ganesan and P. Stoica, "Spacetime block codes: A maximum snr approach," *IEEE Transactions on Information Theory*, vol. 47, pp. 1650–1656, May 2001.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Spacetime block coding for wireless communications: performance results," *IEEE Journal on Select Areas in Communications*, vol. 17, pp. 451–460, March 1999.

[14] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and spacetime coding," *IEEE Transactions on Information Theory*, vol. 45, pp. 1121–1128, May 1999.

[15] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *Journal on Selected Areas in Communications*, vol. 17, p. 18411852, November 1999.

[16] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Transactions on Information Theory*, vol. 48, pp. 1804–1824, July 2002.

[17] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1073–1096, May 2003.

[18] H. Bolcskei, D. Gesbert, C. Papadias, and A. van der Veen, *Space-Time Wireless Systems.* Cambridge University Press, 2006.