Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading

BERTRAND M. HOCHWALD

Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974
hochwald@research.bell-labs.com

THOMAS L. MARZETTA

Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974
tlm@research.bell-labs.com

Abstract

Motivated by information-theoretic considerations, we propose a signalling scheme, *unitary space-time modulation*, for multiple-antenna communication links. This modulation is ideally suited for Rayleigh fast-fading environments, since it does not require the receiver to know or learn the propagation coefficients.

Unitary space-time modulation uses constellations of $T \times M$ space-time signals $\{\Phi_\ell,\ \ell = 1, \ldots, L\}$, where T represents the coherence interval during which the fading is approximately constant, and $M < T$ is the number of transmitter antennas. The columns of each Φ_ℓ are orthonormal. When the receiver does not know the propagation coefficients, which between pairs of transmitter and receiver antennas are modeled as statistically independent, this modulation performs very well either when the SNR is high or when $T \gg M$.

We design some multiple-antenna signal constellations and simulate their effectiveness as measured by bit error probability with maximum likelihood decoding. We demonstrate that two antennas have a 6 dB diversity gain over one antenna at 15 dB SNR.

Index Terms—Multi-element antenna arrays, wireless communications, channel coding, fading channels, transmitter and receiver diversity, space-time modulation

1 Introduction

Fading is traditionally regarded as a nuisance by the designers of wireless communications systems. Its effects are often mitigated by some combination of differential phase modulation, interleaving, or the transmission of pilot or training signals [1]. But, paradoxically, Rayleigh flat fading can be beneficial for a multiple-antenna communication link. It is shown in [6, 19] that, in a Rayleigh flat-fading environment, a link has a theoretical capacity that increases linearly with the smaller of the number of transmitter and receiver antennas, provided that the complex-valued propagation coefficients between all pairs of transmitter and receiver antennas are statistically independent and known to the receiver.

However, learning the fading coefficients becomes increasingly difficult as either the fading rate or number of transmitter antennas increases. In an effort to increase channel capacity or lower error probability, it is accepted practice to increase the number of transmitter antennas (thereby gaining "diversity" [9], [15]). But increasing the number of transmitter antennas increases the required training interval and reduces the available time in which data may be transmitted before the fading coefficients change. At vehicle speeds of 60 miles/hour, a mobile operating at 1.9 GHz has a fading coherence interval of about 3 ms, which for a symbol rate of 30 kHz corresponds to a fresh fade every 50–100 symbol periods. If several training symbols per transmitter antenna are needed, the coefficients for only a few antennas can be learned before a fresh fade occurs. Next-generation cellular systems in Europe will be expected to operate under very fast fading (trains moving at speeds up to 500 km/hr [20]) and hence it may be impractical to learn even the single coefficient between one transmitter and one receiver antenna.

Motivated by these considerations, we used Shannon theory in [8] to analyze multiple-antenna links without imposing any training schemes and with no assumed knowledge of the random fading coefficients. The complex fading coefficients between all pairs of transmitter and receiver antennas were modelled as independent with uniformly distributed phases and Rayleigh distributed magnitudes. The fading coefficients were piecewise constant over fixed time intervals, with channel coding performed over many such independent fading intervals. We showed that the channel capacity could not be increased by making the number of transmit antennas greater than the length of the fading interval, and found that the capacity-attaining signals had considerable structure. However, we did not explicitly address the problems of modulation and channel coding. In this paper, we use the structure derived in [8] to motivate a particular space-time modulation scheme.

The information-theoretic results in [8] suggest a signal constellation comprising complex-valued sig-

nals that are orthonormal with respect to time among the transmitter antennas. We call this signalling scheme *unitary space-time modulation*. When viewed as vector functions of time, the signals carry the message information entirely in their directions. In this paper, we explain in detail how to create, modulate, and demodulate unitary space-time modulation on a multiple antenna link operating in Rayleigh flat fading. Throughout most of the paper the propagation coefficients are assumed to be unknown to the receiver, but we also show how to use the modulation when the coefficients are known. When the receiver does not know the coefficients, no attempt to learn them is made. We concentrate on modulation and constellation design, and do not address coding issues that lower error probability by adding redundancy. We focus, instead, on raw or uncoded signal and bit error probabilities. When combined with appropriate channel coding, our proposed signal constellations can theoretically attain a high fraction of the channel capacity. Some multiple-antenna coding issues for receivers that know the channel appear in [18].

Section 2 presents the signal model and operating assumptions, and Section 3 reviews the information-theoretic foundations for unitary space-time modulation. In Section 4, we extend the information-theoretic justification by arguing that unitary space-time modulation is nearly optimal when the signal-to-noise ratio is high. In Section 5, we consider the use of unitary space-time modulation to transmit data across a multiple-antenna link, and discuss maximum likelihood demodulation and the properties a good constellation should have. In Section 6 some signal design issues are treated and simulations of a two-transmitter-antenna system are presented. We extend some of the piecewise-constant theory to continuous fading in Section 7.

The following notation is used throughout the paper: $\log x$ is the base-two logarithm of x, while $\ln x$ is base e. Given a sequence $b_1, b_2, \ldots$, of positive real numbers, we say that $a_n = O(b_n)$ as $n \to \infty$ if $|a_n|/b_n$ is bounded by some positive constant for sufficiently large n; we say that $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. Two complex vectors, a and b, are *orthogonal* if $a^\dagger b = 0$, where the superscript † denotes "conjugate transpose." The mean-zero, unit-variance, circularly-symmetric, complex Gaussian distribution is denoted $\mathcal{CN}(0, 1)$.

2 Multiple-Antenna Link: Signal Model

Consider a communication link comprising M transmitter antennas and N receiver antennas that operates in a Rayleigh flat-fading environment. Each receiver antenna responds to each transmitter antenna through a statistically independent fading coefficient that is constant for T symbol periods. The received signals are corrupted by additive noise that is statistically independent among the N receivers and the T symbol

periods. In complex baseband representation, during the T-symbol interval we transmit the signal $\{s_{tm},\ t = 1, \ldots, T,\ m = 1, \ldots, M\}$, and we receive the noisy signal $\{x_{tn},\ t = 1, \ldots, T,\ n = 1, \ldots, N\}$ related by the equation

$$x_{tn} = \sqrt{\rho/M} \sum_{m=1}^{M} h_{mn} s_{tm} + w_{tn}, \qquad t = 1, \ldots T, \ \ n = 1 \ldots N. \tag{1}$$

Here h_{mn} is the complex-valued fading coefficient between the mth transmitter antenna and the nth receiver antenna. The fading coefficients are constant for $t = 1, \ldots T$, and they are independent with respect to m and n and $\mathcal{CN}(0, 1)$ distributed, with density

$$p(h_{mn}) = \frac{1}{\pi} \exp\left\{-|h_{mn}|^2\right\}.$$

The transmitted signal has an average (over the M antennas) expected power equal to one,

$$\frac{1}{M} \sum_{m=1}^{M} \mathrm{E}\,|s_{tm}|^2 = 1, \qquad t = 1, \ldots, T. \tag{2}$$

The additive noise at time t and receiver antenna n is denoted w_{tn}, and is independent (with respect to both t and n), identically distributed $\mathcal{CN}(0, 1)$. The quantities in the signal model (1) are normalized so that ρ represents the expected signal-to-noise ratio (SNR) at each receiver antenna, independently of the number of transmitter antennas. We assume that the realizations of h_{mn}, $m = 1, \ldots, M$, $n = 1, \ldots, N$ are not known to the receiver or transmitter.

We use matrix notation for the transmitted signal S $(T \times M)$, and the received signal X $(T \times N)$. Conditioned on S, the received signal X has independent and identically distributed columns (across the N antennas); at a particular antenna, the T received symbols are zero-mean circularly-symmetric complex Gaussian, with $T \times T$ covariance matrix

$$\Lambda = I_T + (\rho/M) S S^\dagger, \tag{3}$$

where I_T is the $T \times T$ identity matrix. The received signal has conditional probability density,

$$p(X \mid S) = \frac{\exp\left(-\mathrm{tr}\left\{\Lambda^{-1} X X^\dagger\right\}\right)}{\pi^{TN} \det{}^N \Lambda}, \tag{4}$$

where "tr" denotes the trace function.

We assume, for now, that the fading coefficients change to new independent realizations every T symbol periods. This piecewise constant fading process mimics, in a tractable manner, the behavior of a continuously fading process. Furthermore, it is a very accurate representation of many TDMA, frequency hopping, or block-interleaved systems [13]. We consider continuous fading processes later. Each channel use (consisting of a block of T transmitted symbols) is independent of every other. Thus, data can be transmitted reliably at any rate less than the channel capacity, where the capacity is the least upper bound on the mutual information between X and S, or

$$C = \sup_{p(S)} I(X;S),$$

subject to the average power constraint (2), and where

$$\begin{aligned} I(X;S) &= \mathrm{E}\,\log\frac{p(X \mid S)}{p(X)} \\ &= \int dS\, p(S) \int dX\, p(X \mid S) \log\left\{\frac{p(X \mid S)}{\int d\acute{S}\, p(\acute{S}) p(X \mid \acute{S})}\right\}. \end{aligned} \tag{5}$$

The capacity C is measured in bits per block of T symbols. In general, one must code across multiple blocks to achieve capacity.

3 Summary of Known Capacity Results

The conditional density (4) has considerable symmetry arising from the statistical equivalence of each time-sample and of each transmitter antenna. The special properties of the conditional density, in combination with the concavity of the mutual information functional, lead to some general conclusions [8] that are summarized here.

3.1 Capacity limited by length of coherence interval; structure of capacity attaining signals

Theorem 1 *(Limit on number of transmitter antennas) For any coherence interval T and any fixed number of receiver antennas, the capacity obtained with $M > T$ transmitter antennas equals the capacity obtained with $M = T$ transmitter antennas.*

In what follows we assume that $M \leq T$.

Theorem 2 *(Structure of signal that achieves capacity) A capacity-achieving random signal matrix may be constructed as a product $S = \Phi V$, where Φ is an isotropically distributed $T \times M$ matrix whose columns are*

orthonormal, and V is an independent $M \times M$ real, nonnegative, diagonal matrix. Furthermore, we can choose the joint density of the diagonal elements of V to be unchanged by rearrangements of its arguments.

An *isotropically distributed* unit vector has a probability density that is unchanged when the vector is left-multiplied by any deterministic unitary matrix. Similarly, the isotropically distributed $T \times M$ matrix Φ obeys $\Phi^\dagger \Phi = I$, and has a density that is unchanged when it is left-multiplied by any $T \times T$ unitary matrix. In a natural way, Φ is the matrix counterpart of a complex scalar having unit magnitude and uniformly distributed phase. The joint probability density of Φ in terms of its M columns $\phi_1, \ldots, \phi_M$ is [8]

$$p(\Phi) = \left[\prod_{m=1}^{M} \frac{\Gamma(T+1-m)}{\pi^{T+1-m}} \right] \cdot \prod_{\substack{m_1, m_2 \\ m_1 \geq m_2}} \delta\left(\phi_{m_1}^\dagger \phi_{m_2} - \delta_{m_1 m_2} \right), \tag{6}$$

where $\delta(\cdot)$ is the Dirac delta function defined for complex arguments to be $\delta(\cdot) = \delta(\mathrm{Re}\,\{\cdot\}) \cdot \delta(\mathrm{Im}\,\{\cdot\})$, and $\delta_{m_1 m_2}$ is one when $m_1 = m_2$ and is zero otherwise. Substituting the structured S into (5) and performing some simplification yields

$$\begin{aligned} I(X;S) &= -TN \log e - N \cdot \sum_{m=1}^{M} \mathrm{E}\, \log\left(1 + \frac{\rho v_m^2}{M}\right) \\ &\quad - \int d\lambda \cdot p(\lambda) \cdot f(\lambda) \cdot \left[\log f(\lambda) - (\log e) \cdot \sum_{\ell=1}^{\min(N,T)} \lambda_\ell \right], \end{aligned} \tag{7}$$

where $v_1, \ldots, v_M$ are the nonnegative real diagonal entries of V,

$$\begin{aligned} f(\lambda) &\stackrel{\mathrm{def}}{=} \int dV \cdot \frac{p(V)}{\prod_{m=1}^{M} (1 + \frac{\rho}{M} v_m^2)^N} \\ &\quad \cdot \int d\Phi\, p(\Phi) \cdot \exp\left\{ \sum_{\ell=1}^{\min(N,T)} \sum_{m=1}^{M} \lambda_\ell \cdot \left(\frac{\rho v_m^2}{M + \rho v_m^2} \right) \cdot |\phi_{\ell m}|^2 \right\}, \end{aligned} \tag{8}$$

and

$$p(\lambda) \stackrel{\mathrm{def}}{=} \frac{e^{-\sum_{\ell=1}^{\min(N,T)} \lambda_\ell} \cdot \left(\prod_{\ell=1}^{\min(N,T)} \lambda_\ell \right)^{|T-N|} \cdot \prod_{i<j} (\lambda_i - \lambda_j)^2}{\prod_{\ell=1}^{\min(N,T)} \Gamma(T-\ell+1) \cdot \Gamma(N-\ell+1)}. \tag{9}$$

In the above, $p(V)$ denotes the joint density on $v_1, \ldots, v_M$, and $\lambda \stackrel{\mathrm{def}}{=} [\lambda_1, \ldots, \lambda_{\min(N,T)}]$. Computing the channel capacity requires maximizing $I(X;S)$ with respect to the joint probability density of the M nonnegative real diagonal elements of V. It is shown in [8] that we may choose $\mathrm{E}\, v_1^2 = \ldots = \mathrm{E}\, v_M^2 = T$.

The transmitted signal has the partitioned form $S = \left[\begin{array}{ccc} v_1\phi_1 & \cdots & v_M\phi_M \end{array} \right]$, where the M columns, representing the temporal signals fed into the M transmitter antennas, are mutually orthogonal. As we will argue, for either $T \gg M$, or for high SNR and $T > M$, setting $v_1 = \ldots = v_M = \sqrt{T}$, which we call *unitary space-time modulation*, achieves capacity.

3.2 Capacity bounds

An upper bound on capacity is obtained if we assume that the receiver is provided with a noise-free measurement of the propagation coefficients H. This *perfect-knowledge* upper bound is [6], [19]

$$C_u = T \cdot \mathrm{E} \log\det\left[I_N + \frac{\rho}{M} H^\dagger H\right] \tag{10}$$

per block of T symbols. When H is known to the receiver, the perfect-knowledge capacity bound is achieved with transmitted signal S whose elements are independent $\mathcal{CN}(0,1)$. For the special case $M = N = 1$ the perfect-knowledge capacity upper bound is $C_u = T(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho)$, where $\mathrm{E}_1(x) \stackrel{\mathrm{def}}{=} \int_x^\infty \frac{e^{-y}}{y}\, dy$ is the *exponential integral*.

A lower bound on capacity that we denote C_l is obtained by assigning unit probability mass to $v_1 = \cdots = v_M = \sqrt{T}$, substituting this mass function into (7), and integrating with respect to V. For the special case $M = N = 1$, the integration over Φ in (8) can be performed analytically to yield the capacity lower bound

$$\begin{aligned} C_l &= -T\log e - \log(1+\rho T) \\ &- \int_0^\infty \frac{(T-1)e^{-\lambda/(1+\rho T)}\gamma\left(T-1, \frac{\rho T\lambda}{1+\rho T}\right)}{\Gamma(T)(1+\rho T)\left[\frac{\rho T}{1+\rho T}\right]^{T-1}} \log\left[\frac{(T-1)e^{-\lambda/(1+\rho T)}\gamma\left(T-1, \frac{\rho T\lambda}{1+\rho T}\right)}{(1+\rho T)\left[\frac{\rho T\lambda}{1+\rho T}\right]^{T-1}}\right] d\lambda, \end{aligned} \tag{11}$$

where $\gamma(T,z) \stackrel{\mathrm{def}}{=} \int_0^z q^{T-1}e^{-q}dq$ is the *incomplete gamma* function. The next theorem, proven in [8], says that $C_l/T \to C/T \to C_u/T$, and the capacity-achieving distribution of v_1 is a unit mass at $\sqrt{T}$, as $T \to \infty$.

3.3 Asymptotic capacity and signal structure for $T \gg M$

Theorem 3 *(Capacity, asymptotically in T) Let $M = N = 1$. The capacity has the asymptotic expansion $(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho) - O\left(\sqrt{\frac{\log T}{T}}\right) = C_l/T \le C/T \le C_u/T = (\log e)e^{1/\rho}\mathrm{E}_1(1/\rho)$, as $T \to \infty$. This capacity is achieved as $T \to \infty$ by setting $v_1 = \sqrt{T}$ with probability 1.*

Heuristic considerations suggest strongly that Theorem 3 extends in a reasonable way to multiple transmitter and receiver antennas. Although H is unknown to the receiver, as T becomes large we could reserve a small portion of the coherence interval to send training data from which the receiver could estimate H, so C/T should approach C_u/T as $T \to \infty$ and this capacity would be attained by a transmitted signal S whose components are approximately independent $\mathcal{CN}(0,1)$. To demonstrate that $S = \sqrt{T}\Phi$, where $\Phi^\dagger\Phi = I$ and Φ is isotropically distributed, attains capacity, we note that as $T \to \infty$ the entries of S have distributions that approach independent $\mathcal{CN}(0,1)$ (see [8]). On the other hand, when $M = T$, setting $v_1 = \ldots = v_M = \sqrt{T}$ is not useful; in this case, $SS^\dagger = T \cdot \Phi\Phi^\dagger = T \cdot I_T$, so $p(X \mid S) = p(X)$ and no information is transmitted. In what follows we always assume that $M < T$.

4 Unitary Space-Time Modulation and High SNR

Unitary space-time modulation defined

The key results of the previous section say that: 1) There is no point in making the number of transmitter antennas greater than the duration of the coherence interval; 2) When the duration of the coherence interval is significantly greater than the number of transmitter antennas ($T \gg M$), setting $v_1 = \ldots = v_M = \sqrt{T}$ attains capacity. Taking our cue from these considerations, we define *unitary space-time modulation* to be the transmission of $S = \sqrt{T}\Phi$, where $\Phi^\dagger\Phi = I$. The previous section argues that unitary space-time modulation attains capacity for $T \gg M$. We now argue that unitary space-time modulation is optimal also for any fixed $T > M$, as $\rho \to \infty$. The following result, for the special case $M = N = 1$, shows that letting $v_1 = \sqrt{T}$ with probability one achieves capacity asymptotically as $\rho \to \infty$ for any fixed $T > 1$.

Theorem 4 *(Capacity, asymptotically in ρ) Let $M = N = 1$ and $T > 1$. The capacity has the asymptotic expansions*

$$C = \frac{T-1}{T}C_u + \log\left[\left(\frac{T}{e}\right)^{T-1}\frac{1}{\Gamma(T)}\right] + o(1) \tag{12}$$

$$= \log\left[\left(\frac{\rho T}{e^{\gamma+1}}\right)^{T-1}\frac{1}{\Gamma(T)}\right] + o(1) \tag{13}$$

as $\rho \to \infty$, where $\gamma = 0.5772\ldots$ is Euler's constant. This capacity is achieved as $\rho \to \infty$ by setting $v_1 = \sqrt{T}$ with probability 1.

Proof: See Appendix A.

Figure 1 displays, for $M = N = 1$ and $T = 2$, the exact capacity (obtained with the Blahut-Arimoto algorithm [2], [8]), the perfect-knowledge upper bound (10), the lower bound (11), and the expansion (12) as a function of ρ. Figure 2 is similar, except that $T = 5$, and we see that the lower bounds, asymptotic expansions, and capacities are essentially the same for all SNR's greater than 0 dB. Unlike the case in Theorem 3 where $T \to \infty$, when $\rho \to \infty$ we see that the capacity diverges away from the upper bound.

It is worth attempting to find an intuitive explanation for Theorem 4. The first term in (12) appears to be consistent with the strategy of sending a single known training symbol from which the receiver obtains a very accurate estimate for the fading coefficient, and then transmitting the remaining $T - 1$ symbols as if the fading coefficient were known to the receiver. The capacity thus obtained would correspond to approximately $T - 1$ perfect-knowledge channel uses, giving rise to the first term in (12); the remaining terms can be viewed as the penalty for estimating the fading coefficient imperfectly.

But this appealing argument does not explain why unitary space-time modulation $s = \sqrt{T}\Phi$, which has no explicit training, achieves capacity. Instead, let $s = v\Phi$, where v obeys $\mathrm{E}\, v^2 = T$ but is otherwise arbitrary, and consider the high-SNR received signal,

$$x \approx \sqrt{\rho} v h \Phi,$$

where x and Φ are T-dimensional vectors. The unit vector Φ, apart from its overall phase, can be determined very accurately from x, regardless of h. However, the scalar amplitude v cannot be determined so easily because it is multiplied by the unknown scalar h. Hence, when the SNR is high, transmitting information on Φ appears to be more profitable than transmitting on v. This suggests that we should simply set $v = \sqrt{T}$. Note that both this argument and Theorem 4 apply only if $T > 1$.

A similar intuitive argument suggests that Theorem 4 also holds for multiple transmitters and receivers; that is $v_1, \ldots, v_M \to \sqrt{T}$ as $\rho \to \infty$. For high SNR and $T > M$, the signal at the nth receiver antenna is

$$x_n \approx \sqrt{\rho/M} \sum_{m=1}^{M} v_m h_{mn} \phi_m, \tag{14}$$

where x_n and ϕ_m are T-dimensional vectors. Even for a very high SNR we cannot easily determine $v_1, \ldots, v_M$ because they are multiplied by the unknown fading coefficients $h_{1n}, \ldots, h_{Mn}$. However, the columns of Φ span an M-dimensional subspace of the T-dimensional complex vector space. In this vector space, the subspace is a hyperplane, and any two signals Φ_i and Φ_j that generate nonidentical subspaces

yield two distinct hyperplanes that intersect on some lower-dimensional hyperline. The probability of x_n falling on one of these intersections is zero. Hence, independently of h_{mn}, for high SNR we can perfectly distinguish Φ_i from Φ_j as long as their columns do not span the same subspace. (We demonstrate this effect in the next section by calculating the probability of mistaking one for the other.) Nevertheless, we do not have a proof that $v_1, \ldots, v_M \to \sqrt{T}$ as $\rho \to \infty$, for $M > 1$.

In short, when either $T \gg M$, or ρ is large with $T > M$, information-theoretic arguments say that the modulation of $v_1, \ldots, v_M$ is neither very interesting nor very useful. Rather one should use unitary space-time modulation, where $v_1 = \cdots = v_M = \sqrt{T}$ and where all message information is transmitted on the directions of the orthonormal columns of Φ. While information-theoretic arguments implicitly require the use of channel codes to attain capacity, we now consider the use of unitary space-time modulation in an uncoded form, and find design rules that help us generate good constellations of these signals.

5 ML Receiver for Unitary Space-Time Modulation

We now consider maximum likelihood (ML) reception of a constellation of L signals employing unitary space-time modulation,

$$S_\ell = \sqrt{T}\Phi_\ell, \qquad \ell = 1, \ldots, L,$$

where $\{\Phi_\ell,\ \ell = 1, \ldots, L\}$ are $T \times M$ complex matrices satisfying $\Phi_\ell^\dagger \Phi_\ell = I$. Ignore, for the moment, the problem of how to generate such a constellation. We derive the ML receiver and its performance when H is unknown and, for comparison, when H is known to the receiver (H is never known to the transmitter). It is customary to call the former receiver noncoherent and the latter receiver coherent.

5.1 Channel unknown to receiver

Maximum likelihood decoding becomes

$$\begin{aligned}
\Phi_{\mathrm{ml}} &= \arg \max_{\Phi_\ell \in \{\Phi_1, \ldots, \Phi_L\}} p(X \mid \Phi_\ell) \\
&= \arg \max_{\Phi_\ell \in \{\Phi_1, \ldots, \Phi_L\}} \frac{\exp\left(-\mathrm{tr}\left\{\left[I_T + (\rho T/M)\,\Phi_\ell \Phi_\ell^\dagger\right]^{-1} X X^\dagger\right\}\right)}{\pi^{TN} \det^N \left[I_T + (\rho T/M)\,\Phi_\ell \Phi_\ell^\dagger\right]} \\
&= \arg \max_{\Phi_\ell \in \{\Phi_1, \ldots, \Phi_L\}} \frac{\exp\left(-\mathrm{tr}\left\{\left[I_T - \frac{1}{1+M/\rho T}\Phi_\ell \Phi_\ell^\dagger\right] X X^\dagger\right\}\right)}{\pi^{TN} (1 + \rho T/M)^{MN}}
\end{aligned}$$

$$= \arg \max_{\Phi_\ell \in \{\Phi_1,\ldots,\Phi_L\}} \mathrm{tr}\,\{X^\dagger \Phi_\ell \Phi_\ell^\dagger X\}, \tag{15}$$

where the matrix formulas $\det(I + AB) = \det(I + BA)$ and $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ are used [17]. The ML receiver seeks to maximize the energy contained in the MN inner products that comprise $\Phi_\ell^\dagger X$.

Suppose now that $L = 2$, and Φ_1 and Φ_2 are transmitted with equal probability. The probability of decoding error is then

$$\begin{aligned} P_e &= \frac{1}{2}\mathrm{P}(\mathrm{tr}\,\{X^\dagger \Phi_2 \Phi_2^\dagger X\} > \mathrm{tr}\,\{X^\dagger \Phi_1 \Phi_1^\dagger X\} \mid \Phi_1 \text{ transmitted}) \\ &\quad + \frac{1}{2}\mathrm{P}(\mathrm{tr}\,\{X^\dagger \Phi_1 \Phi_1^\dagger X\} > \mathrm{tr}\,\{X^\dagger \Phi_2 \Phi_2^\dagger X\} \mid \Phi_2 \text{ transmitted}). \end{aligned} \tag{16}$$

As we show in the next theorem, the probability of error given that Φ_1 is transmitted is equal to the probability of error given that Φ_2 is transmitted, and P_e has a closed-form analytical expression that depends only on the singular values of the $M \times M$ matrix $\Phi_2^\dagger \Phi_1$.

Theorem 5 *(Two-signal error probability: H unknown) Suppose that two unitary space-time modulation signals Φ_1 and Φ_2 are transmitted with equal probability, and decoded with an ML receiver. Then the probability of error is*

$$P_e = \sum_j \mathrm{Res}_{\omega = i a_j} \left\{ -\frac{1}{\omega + i/2} \prod_{\substack{m=1 \\ d_m < 1}}^{M} \left[\frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)(\omega^2 + a_m^2)} \right]^N \right\}, \tag{17}$$

where $1 \geq d_1 \geq \ldots \geq d_M \geq 0$ are the singular values of the $M \times M$ matrix $\Phi_2^\dagger \Phi_1$, and

$$a_m \stackrel{\mathrm{def}}{=} \sqrt{\frac{1}{4} + \frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)}}.$$

Furthermore, P_e decreases as any d_m decreases, and has Chernoff upper bound

$$P_e \leq \frac{1}{2} \prod_{m=1}^{M} \left[\frac{1}{1 + \frac{(\rho T/M)^2 (1 - d_m^2)}{4(1 + \rho T/M)}} \right]^N. \tag{18}$$

Proof: See Appendix B.

For a single transmitter antenna ($M = 1$), d_1 is the magnitude of the inner product between Φ_1 and Φ_2.

For multiple transmitter antennas, $d_1, \ldots, d_M$ represent the similarity between the subspaces spanned by the columns of Φ_1 and Φ_2. The formula (17) is a closed-form expression that can be explicitly evaluated for any special case. See, for example, Appendix B, for the explicit evaluation when $d_1 = \ldots = d_M$. For given $d_1, \ldots, d_M$, the dependence of the probability of error on ρ and T is only through the product ρT.

Figure 3 displays the probability of error as a function of SNR for one transmitter and one receiver antenna ($M = N = 1$) and $T = 5$ for $d_1 = d = 0.0$, 0.4, and 0.8. Note that reducing d below 0.4 gains at most 1 dB in equivalent SNR. Figure 4 shows the probability of error as a function of d and SNR=0, 10, and 20 dB. Here we can see more clearly that reducing d below approximately 0.4 does not reduce the error by much. Figure 5 illustrates the probability of error for two transmitter antennas ($M = 2$), with $d_1 = d_2 = d$. Comparing this figure with Figure 3 reveals that for SNR's greater than 5 dB, two transmitter antennas can have significantly lower error probability than one with the same total transmitted power. This is seen more explicitly in Section 6. Figure 6 superimposes the $d = 0$ and $d = 0.8$ curves from Figures 3 and 5 for relatively low SNR. Observe that below approximately -2 dB, employing a second antenna with unitary space-time modulation actually increases the probability of error. This is not inconsistent with Theorems 3 and 4, which say that unitary space-time modulation is optimal for high SNR or large T. We conclude that when employing unitary space time modulation for given values of ρ, T, and N, there is an optimal number of transmitter antennas M that may be considerably smaller than T.

5.2 Channel known to receiver

We have justified unitary space time modulation $S = \sqrt{T}\Phi$ on information-theoretic grounds for receivers that do not know the channel, when either $T \gg M$ or ρ is large. Surprisingly, we can also justify this modulation when $T \gg M$ and when the receiver knows the channel. When the receiver knows the channel, capacity is achieved by an S matrix composed of independent $\mathcal{CN}(0,1)$ random variables. In Section 3 it is argued that $S = \sqrt{T}\Phi$ (with Φ isotropically distributed) approaches, in distribution, a matrix of independent $\mathcal{CN}(0,1)$ random variables as $T \to \infty$. Hence, for T sufficiently large, unitary space-time modulation is nearly optimal, even when the channel is known. Knowledge of H, however, mandates different criteria for designing a signal constellation.

When H is known to the receiver (although still random),

$$p(X \mid S, H) = \frac{1}{\pi^{TN}} \exp\left(-\mathrm{tr}\left\{(X - \sqrt{\rho/M}SH)(X - \sqrt{\rho/M}SH)^\dagger\right\}\right),$$

and maximum likelihood decoding is

$$\Phi_{\mathrm{ml}} = \arg \min_{\Phi_\ell \in \{\Phi_1,\ldots,\Phi_L\}} \mathrm{tr}\,\{(X - \sqrt{\rho T/M}\Phi_\ell H)(X - \sqrt{\rho T/M}\Phi_\ell H)^\dagger\}.$$

As shown in the next theorem, the two-signal probability of error depends on the singular values of the $T \times M$ difference $\Phi_2 - \Phi_1$.

Theorem 6 *(Two-signal error probability: H known) Suppose that two unitary space-time modulation signals Φ_1 and Φ_2 are transmitted with equal probability, and decoded with an ML receiver that knows H perfectly. Then the probability of error, averaged over H, is*

$$P_e = \sum_j \mathrm{Res}_{\omega=i\alpha_j} \left\{ -\frac{1}{\omega + i/2} \prod_{\substack{m=1 \\ \delta_m > 0}}^{M} \left[\frac{1}{(\rho T/M)\delta_m^2(\omega^2 + \alpha_m^2)} \right]^N \right\}, \tag{19}$$

where $2 \geq \delta_1 \geq \ldots \geq \delta_M \geq 0$ are the singular values of $\Phi_2 - \Phi_1$, and

$$\alpha_m \stackrel{\mathrm{def}}{=} \sqrt{\frac{1}{4} + \frac{1}{\rho T \delta_m^2/M}}.$$

Furthermore, P_e decreases as any δ_m increases, and has Chernoff upper bound

$$P_e \leq \frac{1}{2} \prod_{m=1}^{M} \left[\frac{1}{1 + \frac{\rho T}{4M}\delta_m^2} \right]^N. \tag{20}$$

Proof: See Appendix C.

We note that when H is known and S_1 and S_2 are arbitrary (i.e., do not necessarily have the unitary space-time structure) the derivation of exact probability of error in Appendix C still applies with minor changes. The probability of error and Chernoff bound for arbitrary S_1 and S_2 are still given by (19) and (20), but with $\delta_1, \ldots, \delta_M$ replaced by the singular values of $(S_2 - S_1)/\sqrt{T}$. See [18] for an alternative derivation of the Chernoff bound.

In general, there is no direct relationship between the known-H singular values $\delta_1, \ldots, \delta_M$, and the unknown-$H$ singular values $d_1, \ldots, d_M$. When $M = 1$, for example, we have $d_1 = |\Phi_2^\dagger \Phi_1|$ and $\delta_1 = \|\Phi_2 - \Phi_1\| = \sqrt{2 - 2\mathrm{Re}\,(\Phi_2^\dagger \Phi_1)}$, so for a given value of d_1, δ_1 can have the range of values $\sqrt{2(1-d_1)} \leq \delta_1 \leq \sqrt{2(1+d_1)}$.

For the special case $d_1 = \ldots = d_M = 0$ (the two signals are orthogonal), then $\delta_1 = \ldots = \delta_M = \sqrt{2}$,

and a direct comparison of (17) and (19) is meaningful. For high SNR, the Chernoff bounds for H unknown (18) and H known (20) are then

$$P_e \lesssim \frac{1}{2}\left(\frac{4}{\rho T}\right)^{MN} \text{ (unknown)}, \qquad P_e \lesssim \frac{1}{2}\left(\frac{2}{\rho T}\right)^{MN} \text{ (known)},$$

which suggests that the probability of error is a factor of approximately 2^{MN} lower when the receiver knows H than when it does not. Figure 7 shows the exact probability of error as a function of SNR when the two signals are orthogonal, for known and unknown H, and $M = N = 1$, and $T = 5$. For moderately high SNR's the knowledge of H yields a 3 dB gain, as expected.

We have seen that when H is known to the receiver, unitary space time modulation is a viable option for $T \gg M$. However, the maximum likelihood receivers for known H versus unknown H are considerably different, and so are the dependencies of probability of error on the signals. In the former we seek to maximize the singular values of $\Phi_2 - \Phi_1$, whereas in the latter we seek to minimize the singular values of $\Phi_2^\dagger \Phi_1$; these criteria are not compatible. Moreover, signal constellations for known H generally have to be larger than those for unknown H, reflecting the significantly higher channel capacity and lower error probability. When H is known, signals are distinguishable that would otherwise be indistinguishable if H were unknown, including antipodal pairs $\pm S$, as well as signals whose columns are permuted with respect to one another. The remainder of the paper considers only unknown H.

6 Design of Unitary Space-Time Modulation Constellations

We wish to design a constellation of L signals $\{S_\ell = \sqrt{T}\Phi_\ell,\ \ell = 1, \cdots, L\}$, where $\Phi_\ell^\dagger \Phi_\ell = I$. Since we assume no channel coding, the size of the constellation is $L = 2^{RT}$, where R is the data rate in bits per channel use. To minimize pairwise probability of error, one would like the singular values of the products $\Phi_{\ell_2}^\dagger \Phi_{\ell_1}$, $\ell_1 \neq \ell_2$ to be as small as possible. Unfortunately, we do not know of a way to minimize these singular values, nor can we visualize the properties of a good signal constellation. In constructing a constellation, we note that the pairwise probability of error is invariant to certain unitary transformations, including left-multiplication by a common $T \times T$ unitary matrix, $\Phi_\ell \to \Psi^\dagger \Phi_\ell,\ \ell = 1, \cdots, L$, and right-multiplication by arbitrary $M \times M$ unitary matrices, $\Phi_\ell \to \Phi_\ell \Theta_\ell,\ \ell = 1, \cdots, L$. Constellations that are related in this way are equally good.

6.1 Bound on d_1 for one transmitter antenna

With a single transmitter antenna ($M = 1$), the task is to find L unit vectors the magnitudes of whose inner products $(d_1)_{l_1,l_2},\ \ell_1 \neq \ell_2$ are as small as possible. As shown in the previous section, there is no direct relation between the magnitude of the inner product between two complex vectors and their Euclidean distance. There is a large body of literature on choosing collections of unit vectors that maximize their pairwise Euclidean distances (see [3] and the many references therein). However, the literature on choosing vectors that minimize their pairwise correlations appears to be smaller [10], [12], [22]. Moreover, the constellation design problem in T-dimensional complex space does not reduce to a design problem in $2T$-dimensional real space, because $d_1 = |\Phi_2^\dagger \Phi_1|$ does not equal the magnitude of the inner product between the real $2T$-dimensional vectors $[\mathrm{Re}\,(\Phi_1)^T\ \mathrm{Im}\,(\Phi_1)^T]$ and $[\mathrm{Re}\,(\Phi_2)^T\ \mathrm{Im}\,(\Phi_2)^T]$.

For given values of T and L, it is not known how small we can make $d_{\max} = \max_{l_1 \neq l_2} (d_1)_{l_1,l_2}$, the largest pairwise correlation between the signals. However, the following bound is available [10], [12]:

$$L \leq \frac{1 - d_{\max}^2}{k + 1 - (T+k) d_{\max}} \cdot \frac{T(T+1)\cdots(T+k)}{k!}, \tag{21}$$

where $k = 0, 1, \ldots$ is a free parameter. Solving this relation, for example, with $T = 5$ and $L = 32$ (which gives 32 signals in 5 time samples, or $R = 1$ bit/channel use), yields $d_{\max} \geq 0.46$. Hence, we would like to choose 32 complex 5-dimensional unit vectors, constituting our constellation, for which $d_{\max}$ is as close to 0.46 as possible. It is not known how tight the bound (21) is.

6.2 Algorithms for reducing $d_{\max}$

Starting with any constellation of unit vector signals for a single transmitter antenna $M = 1$, we describe a simple iterative algorithm for reducing $d_{\max}$:

1. Compute $d_{\max}$, the maximum of the magnitudes of all $L(L-1)/2$ distinct inner products, and choose a pair of vectors whose inner product is $d_{\max}$.

2. "Separate" the pair by moving each vector a small amount in opposite directions along the difference vector between the pair.

3. Renormalize the pair, if needed.

4. Repeat Steps 1–3 until $d_{\max}$ stops decreasing.

Using this technique with $T = 5$ and $L = 32$ (one bit per channel use) on a constellation of initially randomly generated unit vectors, we were able to achieve $d_{\max} = 0.515$. We see that we are not very far from the bound $d_{\max} \geq 0.46$. Figure 8 illustrates the correlations between the members of the constellation, $\Phi_1, \ldots, \Phi_{32}$.

This same algorithm may be generalized to multiple transmitter antennas $M > 1$ by identifying the pair of signals whose product yields the singular values that generate the worst (largest) Chernoff bound on error probability according to (18). "Separating" the signals can be aided by left-multiplying by unitary matrices, since this operation preserves the orthogonality of the columns in each signal. We omit the details. Figure 9 displays the bit error performance of constellations of unitary space-time modulated signals generated for $M = 1$ and $M = 2$ transmitter antennas, each with $R = 1$ bit/channel use and $T = 5$. We see that the bit error probability decreases approximately as $1/\rho^2$ for high SNR with two antennas, versus approximately as $1/\rho$ with one antenna. No attempt was made to assign the data bits to the unitary space-time signals optimally.

6.3 Adaptation to continuous fading

In certain TDMA, frequency hopping, or interleaving applications, the fading is approximately constant within a T-symbol block and is independent across blocks. However, in a mobile environment the fading may change gradually without piecewise jumps. If the fading process changes little within a symbol interval, one way to model the sampled received signal is to assign an autocorrelation function to the fading coefficients. One common autocorrelation function is Jakes', proposed in [9]. It is usually possible to select some value for T such that the fading is approximately constant over T symbols; in doing so, however, adjacent blocks of T symbols may be correlated as in Figure 10. Interleaving blocks of T symbols could remove this residual correlation. Instead, we describe a strategy that exploits the residual correlation between T-symbol blocks with a "seamless" modification to unitary space-time modulation.

Seamless unitary space-time modulation constrains all the entries in the first and the last rows of Φ_ℓ to have magnitude $1/\sqrt{T}$, i.e. $|[\Phi_\ell]_{1m}| = |[\Phi_\ell]_{Tm}| = 1/\sqrt{T}$, $m = 1, \cdots, M$. Suppose now that the signal Φ_j is to be transmitted immediately after the signal Φ_i. Recall that we can right-multiply Φ_j by any $M \times M$ unitary matrix without affecting its statistical properties at the receiver. Consequently, we can multiply Φ_j by the $M \times M$ diagonal unitary matrix Θ that makes the first row of $\Phi_j\Theta$ equal the last row of Φ_i, i.e. $[\Phi_j\Theta]_{1m} = [\Phi_i]_{Tm}$, $m = 1, \cdots, M$. Then, instead of transmitting all T rows of Φ_j, it is only necessary to transmit the last $T - 1$ rows of $\Phi_j\Theta$. Hence, each signal (except the very first) can be transmitted in

$T-1$ time samples rather than T, but the receiver can still exploit the T-symbol coherence interval to demodulate each signal; see [11] for single-antenna codes with this feature. It follows that the size of the signal constellation can be reduced from $L = 2^{RT}$ to $L = 2^{R(T-1)}$. For example, with $R = 1$ half the number of signals are needed.

It is worth noting that for $T = 2$ and $M = 1$ (fading approximately constant in blocks of two symbols, and one transmitter antenna), this form of seamless unitary space-time modulation is equivalent to conventional differential phase-shift modulation. To see this, suppose we wish to transmit one bit per channel use, $R = 1$. Then, using seamless unitary space-time modulation, we need only $L = 2^{R(T-1)} = 2$ signals in our constellation, each of which is a 2×1 vector whose first and last entries have magnitude $1/\sqrt{2}$. Since only two signals are required, making them orthogonal minimizes $d_1 = |\Phi_1^\dagger \Phi_2|$,

$$\Phi_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \qquad \Phi_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}.$$

Let binary message 0 be represented by Φ_1, and 1 by Φ_2. Suppose we want to transmit a binary 0 across the channel after having previously sent a 1 represented by Φ_2. Then we would multiply Φ_1 by -1 so that its first entry matched the last entry of the previously sent Φ_2. We then transmit only the second entry of the modified Φ_1, which is now $-1/\sqrt{2}$. Let $X_1, \ldots, X_3$ denote the three received symbols corresponding to the two transmitted data bits. The receiver then uses X_1 and X_2 to decode the first message bit, and X_2 and X_3 to decode the second. This modulation-demodulation process is exactly differential binary phase-shift keying (D-BPSK).

We now assume that the fading is correlated according to a Jakes model [9], with autocorrelation function $J_0(2\pi f_d t)$ where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind and f_d is the maximum nondimensional Doppler frequency in cycles/sample period. The fading processes shown in Figure 10 are generated according to this model. For $f_d = 0.01$ the first zero of the Bessel function is approximately $t = 38$. On the other hand, fading coefficients five time samples apart have correlation 0.976. Because of this high correlation, we may safely choose to design our constellation for any $T \leq 6$.

We now look at the performance of seamless unitary space-time modulation to transmit one bit per channel use ($R = 1$) across this continuously fading channel. Figure 11 shows the bit error rate for one ($M = 1$) and two ($M = 2$) transmitter antennas, and one receiver antenna. To generate this figure, signal constellations of size 2^{T-1} were designed for $T = 2, \ldots, 6$ according to the above principles. The receiver always decoded using maximum likelihood as if the fading were constant for T symbols. As explained

above, $M = 1$ and $T = 2$ corresponds exactly to D-BPSK, which is shown by the dashed line. With $M = 1$ and $T = 3, \ldots, 6$, the performance varies little with T, and is well approximated by the dashed line. On the other hand, with $M = 2$ (two transmitter antennas), the solid lines show that the performance varies greatly with T. As noted in Section 3, when $M = T$, unitary space time modulation is ineffective, and thus the error probability is 0.5 for $T = 2$. For $T = 3$, 4, and 5, the probability of error decreases monotonically very quickly as T increases. For $T = 5$ and two transmitter antennas, the probability of error is lower than for one transmitter antenna for all SNR's greater than 8 dB. Seamless unitary space-time modulation therefore realizes the diversity advantage of the second transmitter antenna for all reasonably high ρ. This behavior is consistent with our information-theoretic justification of unitary space-time modulation for high SNR in Section 4. The slightly worse performance at high SNR of $T = 6$, compared with $T = 5$, is possibly due to greater variation of the fading coefficients over six time samples than over five. Further experiments indicate that because the fading is so fast, increasing T beyond $T = 6$ degrades the performance even more.

7 Extensions of Theory to Continuous Fading

In the previous section, we successfully modified unitary space-time modulation to work over a fading channel with a Jakes' autocorrelation, even though the scheme was originally motivated by a piecewise constant fading model. In this section, we draw some theoretical conclusions about the optimal signals for fading channels, where, within each independent T-symbol block, the fading coefficients have an arbitrary time correlation. We refer to this time correlation as continuous fading. We obtain extensions of Theorems 1 (limiting the number of effective transmitter antennas) and Theorem 2 (structure of signal that achieves capacity).

Consider the model (1) where, within each block of T symbols, the fading coefficients now are independent, zero-mean, circularly-symmetric, stationary complex Gaussian random processes h_{tmn}. Thus, within a block of T symbols, the received signal is

$$x_{tn} = \sqrt{\rho/M} \sum_{m=1}^{M} h_{tmn} s_{tm} + w_{tn}, \qquad t = 1, \ldots T, \;\; n = 1 \ldots N. \tag{22}$$

The fading processes are independent from one T-symbol block to another, but within each block they are

correlated according to a known autocorrelation function $k(t)$

$$\mathrm{E}\{h_{t_1 m_1 n_1} h^*_{t_2 m_2 n_2}\} = \delta_{m_1 m_2}\delta_{n_1 n_2} k(t_1 - t_2), \tag{23}$$

where $k(0) = 1$. The formula for the conditional probability density (4) still applies but with the modified covariance matrix

$$\Lambda = I_T + (\rho/M)(SS^\dagger) \circ K, \tag{24}$$

where "$\circ$" denotes the Hadamard (i.e., element by element) matrix product, and K is the $T \times T$ Toeplitz covariance matrix, $[K]_{ij} = k(i-j)$. Note that in the former case of piecewise-constant fading, $[K]_{ij} = 1$.

It is realistic to assume that, within a block, the fading is a random process. Less realistic is the independence of the blocks, but this happens naturally if we assume that the blocklength T is long compared with the correlation time of the fading process. For then, the fading between different T-symbol blocks is independent, with the possible exception of a small number of samples near the boundaries of adjacent blocks. The block independence is more likely to be satisfied in TDMA systems such as IS-54/136, where a user does not have access to contiguous blocks.

Suppose that the fading autocorrelation function vanishes beyond some lag $\tau > 0$ that we call the *correlation time* of the fading, i.e., $k(t) = 0$ for $|t| = \tau, \tau + 1, \ldots$. The next theorem extends Theorem 1 to continuous fading.

Theorem 7 *(Limit on number of transmitter antennas in continuous fading) For any correlation time τ and any fixed number of receiver antennas, the capacity obtained with $M > \min(\tau, T)$ transmitter antennas can also be obtained with $M = \min(\tau, T)$ antennas.*

Proof: Suppose that $M > \min(\tau, T)$ and capacity is obtained for some joint probability density for the elements of the $T \times M$ matrix S. All but the $2\min(\tau, T) - 1$ central diagonal bands of the Toeplitz matrix K are zero; that is, $[K]_{ij} = 0,\ |i-j| \geq \min(\tau, T)$. The Hadamard product in (24) therefore causes the conditional probability density (4) to depend on only the $2\min(\tau, T) - 1$ central diagonal bands of $SS^\dagger$. A covariance-extension theorem in [5] states that one can always find a $T \times T$ Hermitian nonnegative-definite matrix Q whose rank is less than or equal to $\min(\tau, T)$, and whose $2\min(\tau, T) - 1$ central diagonal bands are proportional to the corresponding bands of $SS^\dagger$. Thus, we can find a Q satisfying

$$\frac{Q_{ij}}{\min(\tau, T)} = \frac{[SS^\dagger]_{ij}}{M}, \qquad \forall\, |i-j| < \min(\tau, T).$$

Since Q has rank at most $\min(\tau, T)$, it can be factored as $Q = S_1 S_1^\dagger$, where S_1 is a $T \times \min(\tau, T)$ matrix. Consequently, for any $T \times M$ matrix S, we can find a $T \times \min(\tau, T)$ matrix S_1 such that

$$\frac{(S_1 S_1^\dagger) \circ K}{\min(\tau, T)} = \frac{(SS^\dagger) \circ K}{M}. \tag{25}$$

This relation implicitly specifies a joint probability density for the elements of S_1 in terms of the joint probability density for the elements of S. We have the power constraint $\mathrm{E}\,\mathrm{tr}\,(S_1 S_1^\dagger)/\min(\tau, T) = \mathrm{E}\,\mathrm{tr}\,(SS^\dagger)/M = T$, which has been shown in [8] to achieve the same capacity as the stronger power constraint (2). Using $\min(\tau, T)$ transmitter antennas, we can therefore achieve the same capacity that can be achieved with M antennas. □

Few realistic autocorrelation functions vanish absolutely beyond some time lag. For the Jakes model considered in Section 6.3, the autocorrelation vanishes at $|t| = \tau \approx 38$. This limits the number of transmitter antennas to approximately 38.

We now determine some of the structure of the capacity-attaining signal in continuous fading. Because of Theorem 7, we assume that $M \leq \min(\tau, T)$. We define a random process $h_1, \ldots, h_T$ to be *cyclically stationary* if $p_{h_1,\ldots,h_T}(h_1, \ldots, h_T) = p_{h_1,\ldots,h_T}(h_{1+t \bmod T}, \ldots, h_{1+(T-1+t) \bmod T})$ for all t, where $p_{h_1,\ldots,h_T}(\cdot)$ is the joint density of $h_1, \ldots, h_T$. Intuitively, shifts in time of $h_1, \ldots, h_T$ "wrap around" without affecting their joint distribution, or, equivalently, the periodic extension of $h_1, \ldots, h_T$ is a stationary random process in the ordinary sense. The next theorem is the continuous-fading version of Theorem 2. Because the fading process is assumed to have less structure than in Theorem 2, the conclusions are weaker. However, the conclusion that the M transmitted signals should be time-orthogonal remains.

Theorem 8 *(Structure of signal that achieves capacity in continuous fading) The capacity attaining S can be chosen to have mutually orthogonal columns, and have joint density that is unchanged by rearrangements of its columns. Furthermore, the columns of S can be made jointly cyclically stationary if the fading is cyclically stationary.*

Proof: The singular value decomposition implies that the capacity-achieving signal S can always be factored into three terms $S = \Phi V \Psi^\dagger$, where Φ and Ψ are unitary matrices and V is real, nonnegative, and diagonal. Equations (4) and (24) imply that

$$p(X \mid \Phi V \Psi^\dagger) = p(X \mid \Phi V). \tag{26}$$

Dropping the third factor yields a new signal $S_1 = \Phi V$ that has the same mutual information as S, and

whose M columns are mutually orthogonal.

We now assume that the capacity-achieving S has mutually orthogonal columns. There are $M!$ ways of rearranging the columns of S, each corresponding to post-multiplying S by a $M \times M$ permutation matrix $P_{M\ell}$, $\ell = 1, \ldots, M!$. Each $SP_{M\ell}$ yields the same mutual information as S. Forming an equally-weighted mixture density for the transmitted signal involving all $M!$ arrangements of its columns yields a signal whose probability density is unchanged by rearranging its columns. The concavity of mutual information as a functional of the input density and Jensen's inequality together imply that the mutual information for this mixture is at least as great as that for S.

Let the fading be cyclically stationary. The transmitted signal may be cyclically shifted in time by pre-multiplying S by the $T \times T$ permutation matrix $P_{T\ell}$ satisfying

$$[P_{T\ell}S]_{tm} = s_{[1+(t-1-\ell) \bmod T]m}, \qquad t = 1, \ldots, T, \ m = 1, \ldots, M. \tag{27}$$

Forming an equally-weighted mixture density for the transmitted signal involving all T cyclic delays yields a density for the transmitted signal that is jointly cyclically stationary. In other words, the periodic extension in time of S is a multivariate (M-component) strict sense stationary random process. We now argue that the cyclic shift does not change the mutual information. Recall the model (22); we apply a cyclic shift in time of $+\ell$ to S, and $-\ell$ to X, to obtain

$$x_{[1+(t-1+\ell) \bmod T]n} = \sqrt{\rho/M} \sum_{m=1}^{M} h_{[1+(t-1+\ell) \bmod T]mn} s_{tm} + w_{[1+(t-1+\ell) \bmod T]n},$$
$$t = 1, \ldots T, \ \ n = 1 \ldots N.$$

The cyclic delay does not change the probability density of W because it is white, and it does not change the probability density of the fading because it is cyclically stationary. Consequently, the cyclic delay of the transmitted signal does not change the mutual information between it and the received signal, so Jensen's inequality implies that the mutual information for the mixture density is at least as great as that for the original signal. □

We make some final observations. First, in the above proof we assume that the fading is cyclically stationary. This is not restrictive since any wide-sense stationary fading process asymptotically becomes cyclically stationary as $T \to \infty$ [21]. Second, the role of the block length T is secondary to that of the coherence time τ. We impose the constraint that blocks of T symbols be independent because it allows

us to use the standard notions of mutual information and channel capacity per block-of-T-symbols. When $T \gg \tau$, the capacity per channel use becomes independent of T, and channel coding could be performed over the many independent fades that occur in a single T-block.

At present we are unable to say anything more about the general structure of the mutually orthogonal cyclically stationary signals that attain capacity. However, using what by now are familiar arguments, we can infer the structure for the limiting case $T \geq \tau \gg M$. One could send training symbols and estimate the fading coefficients and still have time to send data before the coefficients change. The capacity would approach the perfect knowledge capacity, the optimum signals would be approximately white Gaussian, so unitary space-time modulation would be approximately optimal.

8 Conclusions

Multiple element antenna arrays operating in Rayleigh flat fading can potentially sustain enormous data rates with moderate power in a narrow bandwidth. Our approach to this problem began with the premise that nobody knows the propagation coefficients and that the available transmission time should be spent sending message signals rather than training signals. Information-theoretic considerations then led us to unitary space-time modulation. Preliminary results indicate that this modulation can be highly effective, even though the receiver never explicitly learns the propagation coefficients.

We have derived performance criteria for unitary space-time modulation and indicated the properties that a signal constellation with low block probability of error should have. Our particular constellation designs were ad hoc, however, and the problem of how to design constellations systematically that have low probability of error and low demodulation complexity remains open. We have also not considered how to code across more than one block fading interval. Solutions to these problems are especially urgent for large T and high data rates.

Acknowledgments. The authors thank H. Landau, J. Mazo, J. Salz, R. Urbanke, and the late A. Wyner for helpful suggestions.

A Appendix: Asymptotic behavior of C as $\rho \to \infty$

For $M = N = 1$, we show that the mutual information generated by a given $p(v_1)$ can be no more than $o(1)$ larger than (11), the mutual information generated by $p(v_1) = \delta(v_1 - \sqrt{T})$, as $\rho \to \infty$.

We start by letting $p(v_1) = p_a\delta(v_1 - \sqrt{aT}) + p_b\delta(v_1 - \sqrt{bT})$ be composed of two masses, where $a = a(\rho)$ and $b = b(\rho)$ are positive functions of ρ that do not go to zero as $\rho \to \infty$, but are otherwise arbitrary. Since $\mathrm{E}\, v_1^2 = T$, it must hold that $ap_a + bp_b = 1$, and we assume that $p_a = p_a(\rho)$ and $p_b = p_b(\rho)$ are also functions of ρ. We allow a but not b to go to infinity as $\rho \to \infty$. (Allowing both would violate the power constraint.) It is then a simple matter to parallel the derivation of C_l in (11) to obtain the mutual information

$$I = (\log e) \cdot \left[-T - p_a \ln(1+\rho aT) - p_b \ln(1+\rho bT) - \ln\Gamma(T) - \int_0^\infty q(\lambda) \ln\Big(q(\lambda)/\lambda^{T-1} \Big) d\lambda \right], \quad \text{(A.1)}$$

where

$$q(\lambda) \stackrel{\text{def}}{=} p_a \frac{e^{-\lambda/(1+\rho aT)} \gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} + p_b \frac{e^{-\lambda/(1+\rho bT)} \gamma\Big(T-1, \frac{\rho bT\lambda}{1+\rho bT}\Big)}{\Gamma(T-1)(1+\rho bT)\left[\frac{\rho bT}{1+\rho bT}\right]^{T-1}}.$$

We look first at the first term in $\int_0^\infty q(\lambda) \ln\lambda \, d\lambda$, which is

$$\int_0^\infty d\lambda \, p_a \frac{e^{-\lambda/(1+\rho aT)} \gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} \ln\lambda \quad \text{(A.2)}$$

We break the integration into three disjoint ranges: $[0, 1/\rho^\varepsilon]$, $(1/\rho^\varepsilon, \rho^\varepsilon]$, and $(\rho^\varepsilon, \infty)$ for some arbitrary $0 < \varepsilon < 1$. When $\lambda \in [0, 1/\rho^\varepsilon]$, $\frac{\rho aT\lambda}{1+\rho aT} \to 0$ as $\rho \to \infty$, and the expansion

$$\gamma(T-1, z) = \frac{z^{T-1}}{T-1} + O(z^T), \qquad z \to 0 \quad \text{(A.3)}$$

and inequality $e^{-\lambda/(1+\rho aT)} \le 1$ therefore yield

$$\int_0^{1/\rho^\varepsilon} d\lambda e^{-\lambda/(1+\rho aT)} \gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big) \ln\lambda = O\Big(\frac{\ln\rho}{\rho^{T\varepsilon}}\Big). \quad \text{(A.4)}$$

Since $\gamma(T-1, \frac{\rho aT\lambda}{1+\rho aT}) \le \Gamma(T-1)$ for all λ,

$$\int_{1/\rho^\varepsilon}^{\rho^\varepsilon} d\lambda \, e^{-\lambda/(1+\rho aT)} \gamma(T-1, \frac{\rho aT\lambda}{1+\rho aT}) \ln\lambda \le \Gamma(T-1) \int_{1/\rho^\varepsilon}^{\rho^\varepsilon} d\lambda \, \ln\lambda = O\Big(\rho^\varepsilon \ln\rho\Big). \quad \text{(A.5)}$$

When $\lambda \in (\rho^\varepsilon, \infty)$, $\frac{\rho aT\lambda}{1+\rho aT} \to \infty$ as $\rho \to \infty$, and the expansion

$$\gamma(T-1,z) = \Gamma(T-1) - z^{T-2}e^{-z}(1+o(1)), \qquad z \to \infty \tag{A.6}$$

gives

$$\begin{aligned}
&\int_{\rho^\varepsilon}^{\infty} d\lambda e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)\ln\lambda \\
&= \int_{\rho^\varepsilon}^{\infty} d\lambda\, e^{-\lambda/(1+\rho aT)}(\ln\lambda)\Big[\Gamma(T-1) - \left(\frac{\rho aT\lambda}{1+\rho aT}\right)^{T-2} e^{-\frac{\rho aT\lambda}{1+\rho aT}}(1+o(1))\Big] \\
&= \int_{\rho^\varepsilon}^{\infty} d\lambda\, e^{-\lambda/(1+\rho aT)}(\ln\lambda)[\Gamma(T-1) - O(e^{-\rho^\varepsilon})] \\
&= (1+\rho aT)\Gamma(T-1)[1-O(e^{-\rho^\varepsilon})]\int_{\frac{\rho^\varepsilon}{1+\rho aT}}^{\infty} d\lambda e^{-\lambda}\ln[\lambda(1+\rho aT)] \\
&= (1+\rho aT)\Gamma(T-1)\left[[\ln(1+\rho aT) - \gamma + O\Big(\frac{\ln a\rho}{a\rho^{1-\varepsilon}}\Big)\right],
\end{aligned} \tag{A.7}$$

where $\gamma = 0.5772\ldots$ is Euler's constant. Joining (A.2), (A.4), (A.5), and (A.7), and repeating the calculations for the term involving p_b and b, we get

$$\int_0^\infty d\lambda\, q(\lambda)\ln\lambda^{T-1} = (T-1)[p_a\ln(1+\rho aT) + p_b\ln(1+\rho bT) - \gamma] + O\Big(\frac{\ln\rho}{\rho^{1-\varepsilon}}\Big), \tag{A.8}$$

where $0 < \varepsilon < 1$ is arbitrary.

We now look at $\int_0^\infty q(\lambda)\ln q(\lambda)\, d\lambda$. The first term is

$$\begin{aligned}
&\int_0^\infty d\lambda\, p_a \frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} \\
&\quad \cdot \ln\left[p_a \frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} + p_b \frac{e^{-\lambda/(1+\rho bT)}\gamma\Big(T-1, \frac{\rho bT\lambda}{1+\rho bT}\Big)}{\Gamma(T-1)(1+\rho bT)\left[\frac{\rho bT}{1+\rho bT}\right]^{T-1}}\right].
\end{aligned} \tag{A.9}$$

We break the integration into the same three disjoint ranges as before. For $\lambda \in [0, 1/\rho^\varepsilon]$, equation (A.3) yields

$$e^{-\lambda/(1+\rho aT)} = 1 - O\Big(\frac{\lambda}{1+\rho aT}\Big) \tag{A.10}$$

$$\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big) \;=\; \frac{\lambda^{T-1}}{T-1}(1+o(1)). \tag{A.11}$$

If $a = a(\rho)$ does not go to infinity, then neither term in the argument of the logarithm in (A.9) dominates the sum. If $a(\rho)$ goes to infinity, the second term dominates the sum and the logarithm in (A.9) behaves as $\ln(p_b\lambda^{T-1}/[(T-1)\Gamma(T-1)\rho bT]) = (T-1)\ln\lambda - \ln(\Gamma(T+1)\rho b)$ for large ρ. In either case we may then mimic the analysis of equation (A.2) to conclude that

$$\int_0^{1/\rho^\varepsilon} d\lambda\, q(\lambda)\ln q(\lambda) = O\Big(\frac{\ln\rho}{\rho^{1+\varepsilon}}\Big). \tag{A.12}$$

For $\lambda \in (1/\rho^\varepsilon, \rho^\varepsilon]$, the expansion (A.10) again applies, and equation (A.11) implies that

$$(1+o(1))\frac{1}{(T-1)\rho^{(T-1)\varepsilon}} \le \gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big) \le \Gamma(T-1).$$

Thus, the logarithm in (A.9) is $O(\ln\rho)$, and

$$\int_{1/\rho^\varepsilon}^{\rho^\varepsilon} d\lambda\, q(\lambda)\ln q(\lambda) = O\Big(\frac{\ln\rho}{\rho^{1-\varepsilon}}\Big). \tag{A.13}$$

Finally, for $\lambda \in (\rho^\varepsilon, \infty)$, we change the variable of integration to $\lambda' = \lambda/(\rho T)$, $\lambda' \in (1/(T\rho^{1-\varepsilon}), \infty)$. It follows from (A.6) that

$$\begin{aligned}
e^{-\lambda/(1+\rho aT)} &= e^{-\lambda'\rho T/(1+\rho aT)} = e^{-(1+O(1/\rho))\lambda'/a} \\
\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big) &= \gamma\Big(T-1,\frac{\rho^2aT^2\lambda'}{1+\rho aT}\Big) \\
&= \Gamma(T-1) - \Big(\frac{\rho^2aT^2\lambda'}{1+\rho aT}\Big)e^{-\frac{\rho^2aT^2\lambda'}{1+\rho aT}}(1+o(1)) \\
&= \Gamma(T-1) - O(e^{-\rho^\varepsilon}).
\end{aligned}$$

Hence,

$$q(\lambda) = q(\rho T\lambda') = \frac{1}{\rho T}\left[\frac{p_a}{a}e^{-(1+O(1/\rho))\lambda'/a} + \frac{p_b}{b}e^{-(1+O(1/\rho))\lambda'/b}\right]\Big(1+o(1)\Big),$$

and

$$\int_{\rho^\varepsilon}^{\infty} d\lambda\, q(\lambda)\ln q(\lambda) \;=\; \int_{1/(T\rho^{1-\varepsilon})}^{\infty} d\lambda'\, \rho T q(\rho T\lambda')\ln q(\rho T\lambda')$$

$$= -\ln \rho T + \int_0^\infty d\lambda \left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \ln\left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) + o(1). \quad \text{(A.14)}$$

Combining (A.12)–(A.14), we get

$$\int_0^\infty d\lambda q(\lambda) \ln q(\lambda) = -\ln(\rho T) + \int_0^\infty d\lambda \left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \ln\left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) + o(1), \quad \text{(A.15)}$$

and combining this equation with (A.8) gives

$$\begin{aligned} I &= (\log e) \cdot \Bigg[\ln(\rho T) + (T-2)[p_a \ln(1+\rho a T) + p_b \ln(1+\rho b T)] - T - \ln\Gamma(T) - (T-1)\gamma \\ &\quad - \int_0^\infty d\lambda \left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \ln\left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right)\Bigg] + o(1) \\ &= (\log e) \cdot \Bigg[(T-1)[\ln(\rho T) - \gamma] + (T-2)[p_a \ln a + p_b \ln b] - T - \ln\Gamma(T) \\ &\quad - \int_0^\infty d\lambda \left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \ln\left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right)\Bigg] + o(1). \end{aligned} \quad \text{(A.16)}$$

We have that $p_a + p_b = a p_a + b p_b = 1$. Therefore, by Jensen's theorem,

$$p_a \ln a + p_b \ln b \le \ln(a p_a + b p_b) = 0, \quad \text{(A.17)}$$

with equality if and only if $a = b = 1$. Furthermore, as shown in [4], for any density $p(\lambda)$ supported on $\lambda \ge 0$ and satisfying $\int_0^\infty d\lambda\, \lambda\, p(\lambda) = 1$,

$$-\int_0^\infty d\lambda\, p(\lambda) \ln p(\lambda) \le 1,$$

with equality if and only if $p(\lambda) = e^{-\lambda}$. Hence

$$-\int_0^\infty d\lambda \left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \ln\left(\frac{p_a}{a} e^{-\lambda/a} + \frac{p_b}{b} e^{-\lambda/b}\right) \le 1 \quad \text{(A.18)}$$

with equality if and only if $a = b = 1$. Thus, for $T > 1$ and as $\rho \to \infty$, I is maximized by choosing $a = b = 1$ with probability one, which collapses the two distinct masses into one at $v_1 = \sqrt{T}$. When

$a = b = 1$, then $I = C_l$, and from (A.16) we therefore have the expansion

$$C_l = (\log e)\Big[(T-1)[\ln(\rho T) - \gamma] - T - \ln\Gamma(T) + 1\Big] + o(1) \tag{A.19}$$

$$= \log\left[\left(\frac{\rho T}{e^{\gamma+1}}\right)^{T-1}\frac{1}{\Gamma(T)}\right] + o(1). \tag{A.20}$$

Since $C_u/T = (\log e)e^{1/\rho}E_1(1/\rho) = \log(\rho/e^{\gamma}) + O(1/\rho)$, we may also write

$$C = \frac{T-1}{T}C_u + \log\left[\left(\frac{T}{e}\right)^{T-1}\frac{1}{\Gamma(T)}\right] + o(1). \tag{A.21}$$

In our argument for showing that any density with two distinct masses asymptotically generates less mutual information than a single mass at $\sqrt{T}$, we have explicitly prevented one of the two masses from being located at $v_1 = 0$. We now show that a mass at $v_1 = 0$ must have probability that tends to zero as $\rho \to \infty$. As before, we assume that there are two masses, with one at $v_1 = \sqrt{aT}$ with probability p_a, but we place the other at $v_1 = 0$ with probability p_b. The mutual information I is then as in (A.1), but with $b = 0$, and

$$q(\lambda) = p_a\frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} + p_b\frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}.$$

We analyze I as in the previous manner, and begin with $\int_0^\infty q(\lambda)\ln\lambda^{T-1}d\lambda$. Its first term is the same as (A.2), yielding

$$\int_0^\infty d\lambda\, p_a\frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}}\ln\lambda^{T-1} = (T-1)p_a[\ln(\rho aT) - \gamma] + o(1) \tag{A.22}$$

(Compare (A.8).) The second term is

$$\int_0^\infty d\lambda\, p_b\frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}\ln\lambda^{T-1} = (T-1)p_b\frac{\Gamma'(T)}{\Gamma(T)}. \tag{A.23}$$

The integral $\int_0^\infty q(\lambda)\ln q(\lambda)\,d\lambda$ is now analyzed. The first term is

$$\int_0^\infty d\lambda\, p_a\frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}}\ln\left[p_a\frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1, \frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} + p_b\frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}\right]. \tag{A.24}$$

We note that the density

$$\frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}}$$

has a maximum value of approximately $1/(\rho aT)$, and is effectively supported for λ in an interval that increases linearly with ρ, beyond which it decays exponentially. On the other hand, the density

$$\frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}$$

has its maximum value $e^{-(T-1)}(T-1)^{T-1}/\Gamma(T)$ at $\lambda = T-1$, and decays exponentially as λ increases, independently of ρ. By breaking the integration in (A.24) into the three usual disjoint ranges (and omitting the tedious details), we conclude that (A.24) approaches

$$\int_0^\infty d\lambda\, p_a \frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} \ln\left[p_a \frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}}\right]$$
$$= p_a[-\ln(\rho aT) - 1 + \ln p_a] + o(1) \tag{A.25}$$

as $\rho \to \infty$.

The second term in $\int_0^\infty q(\lambda)\ln q(\lambda)\,d\lambda$ is

$$\int_0^\infty d\lambda\, p_b \frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)} \ln\left[p_a \frac{e^{-\lambda/(1+\rho aT)}\gamma\Big(T-1,\frac{\rho aT\lambda}{1+\rho aT}\Big)}{\Gamma(T-1)(1+\rho aT)\left[\frac{\rho aT}{1+\rho aT}\right]^{T-1}} + p_b \frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}\right]$$

and the same arguments show that this approaches (as $\rho \to \infty$)

$$\int_0^\infty d\lambda\, p_b \frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)} \ln\left[p_b \frac{e^{-\lambda}\lambda^{T-1}}{\Gamma(T)}\right] = p_b[-T + (T-1)\Gamma'(T)/\Gamma(T) - \ln\Gamma(T) + \ln p_b]. \tag{A.26}$$

Combining equation (A.1) with (A.22)–(A.26), we deduce that

$$I = (\log e)\,[p_a(T-1)[\ln(\rho aT) - \gamma] + p_a[-T - \ln\Gamma(T) + 1] - p_a\ln p_a - p_b\ln p_b] + o(1).$$

This expression is clearly maximized by letting $p_a(\rho) \to 1$ as $\rho \to \infty$. Hence, any mass at $v_1 = 0$ in the capacity-achieving distribution must have probability that decays to zero as $\rho \to \infty$.

We have been focusing on $p(v_1)$ with two distinct masses, and now outline how to generalize the

above arguments to show that any $p(v_1)$ asymptotically generates less mutual information than $p(v_1) = \delta(v_1 - \sqrt{T})$. First note that the expansion (A.16) can be immediately generalized to n masses $p(v_1) = \sum_{j=1}^{n} p_j \delta(v_1 - \sqrt{a_j T})$, to obtain

$$\begin{aligned} I &= (\log e) \cdot \Bigg[(T-1)[\log(\rho T) - \gamma] + (T-2) \sum_{j=1}^{n} p_j \ln a_j - T - \ln \Gamma(T) \\ &\quad - \int_0^\infty d\lambda \left(\sum_{j=1}^{n} \frac{p_j}{a_j} e^{-\lambda/a_j} \right) \ln \left(\sum_{j=1}^{n} \frac{p_j}{a_j} e^{-\lambda/a_j} \right) \Bigg] + o(1). \end{aligned} \tag{A.27}$$

Provided that $a_1, \ldots, a_n$ are taken from some finite positive interval, the asymptotic expansion (A.27) is uniform, and hence remains valid even if we let n become unbounded (say, as a function of ρ). As $\rho \to \infty$, the mutual information in (A.27) is therefore maximized by having $a_1, \ldots, a_n \to 1$, which reduces the multiple masses to a single mass at $v_1 = \sqrt{T}$. On a finite interval, we can uniformly approximate any continuous density with masses, and because I is concave in $p(v_1)$ (see [8]), we can approximate I arbitrarily closely as well (we are here overlooking many technical details about optimizing in an infinite-dimensional space). The preceding argument therefore tells us that we are asymptotically better off replacing the continuous density on this finite interval with a mass at $v_1 = \sqrt{T}$.

B Appendix: Two-Signal P_e for H Unknown—Exact and Chernoff Upper Bound

From (15), the probability of error when Φ_1 is transmitted is

$$P_{e|1} \stackrel{\text{def}}{=} \mathrm{P}\left(\frac{\rho T/M}{1 + \rho T/M} \mathrm{tr}\left\{ X^\dagger (\Phi_2 \Phi_2^\dagger - \Phi_1 \Phi_1^\dagger) X \right\} > 0 \mid \Phi_1 \right), \tag{B.1}$$

where the multiplicative factor $(\rho T/M)/(1 + \rho T/M)$ is included to simplify the algebraic manipulations in this section. Using the singular value decomposition,

$$\Phi_2^\dagger \Phi_1 = \Theta D \Upsilon^\dagger, \tag{B.2}$$

where Θ and Υ are $M \times M$ unitary matrices, and $D = \mathrm{diag}(d_1, \ldots, d_M)$ is diagonal, real, and nonnegative, we rewrite (B.1) as

$$\begin{aligned} P_{e|1} &= \mathrm{P}\left(\frac{\rho T/M}{1+\rho T/M} \mathrm{tr}\left\{ X^\dagger (\Phi_2 \Theta \Theta^\dagger \Phi_2^\dagger - \Phi_1 \Upsilon \Upsilon^\dagger \Phi_1^\dagger) X \right\} > 0 \mid \Phi_1 \right) \\ &= \mathrm{P}\left(\mathrm{tr}\left\{ Y_2^\dagger Y_2 - Y_1^\dagger Y_1 \right\} > 0 \mid \Phi_1 \right), \end{aligned} \tag{B.3}$$

where

$$\left[\begin{array}{c} Y_1 \\ Y_2 \end{array} \right] = Y = \left(\frac{\rho T/M}{1+\rho T/M} \right)^{1/2} \left[\begin{array}{c} \Upsilon^\dagger \Phi_1^\dagger \\ \Theta^\dagger \Phi_2^\dagger \end{array} \right] X.$$

The N columns of X are independent, identically distributed zero-mean complex Gaussian vectors with covariance matrix $I_T + (\rho T/M)\Phi_1 \Phi_1^\dagger$. Consequently the N columns of Y are also independent; any column y_n has $2M \times 2M$ covariance matrix

$$\mathrm{E}\left\{ y_n y_n^\dagger \mid \Phi_1 \right\} = \left[\begin{array}{cc} (\rho T/M) I_M & (\rho T/M) D \\ (\rho T/M) D & \frac{\rho T/M}{1+\rho T/M} [I_M + (\rho T/M) D^2] \end{array} \right]. \tag{B.4}$$

Note that this covariance matrix depends on Φ_1 and Φ_2 only through the singular value matrix D. We conclude that $P_{e|1}$ depends only on D. If we were to interchange Φ_1 and Φ_2, D would be unchanged, and thus $P_e = P_{e|1} = P_{e|2}$.

From (B.2), we have that $D = (\Phi_2 \Theta)^\dagger (\Phi_1 \Upsilon)$. The matrices $\Phi_2 \Theta$ and $\Phi_1 \Upsilon$ each comprise M orthogonal unit vectors, implying that every $d_1, \ldots, d_M$ is equal to an inner product between unit vectors. Consequently, $0 \leq d_m \leq 1$.

B.1 Exact P_e

We obtain a closed-form expression for the probability of error by inverting the characteristic function of $\mathrm{tr}\left\{ Y_2^\dagger Y_2 - Y_1^\dagger Y_1 \right\}$. The use of the characteristic function for this type of calculation is well-known [14, 16]. Equation (B.3) may be written as

$$P_e = \mathrm{P}\left(\sum_{n=1}^{N} \sum_{m=1}^{M} (|y_{m+M,n}|^2 - |y_{m,n}|^2) > 0 \mid \Phi_1 \right). \tag{B.5}$$

The structure in equation (B.4) implies that Y comprises MN statistically independent pairs of correlated random variables, $\{(y_{m,n}\ y_{m+M,n}),\ m = 1, \ldots, M,\ n = 1, \ldots, N\}$, with covariance

$$\mathrm{E}\left(\left[\begin{array}{c} y_{m,n} \\ y_{m+M,n} \end{array} \right] \left[\begin{array}{cc} y_{m,n}^* & y_{m+M,n}^* \end{array} \right] \Big| \Phi_1 \right) = \left[\begin{array}{cc} \rho T/M & (\rho T/M) d_m \\ (\rho T/M) d_m & \frac{\rho T/M}{1+\rho T/M}[1 + (\rho T/M) d_m^2] \end{array} \right],$$

where d_m is the mth diagonal entry of D, and this covariance does not depend on n. The characteristic function for each independent term in (B.5) is therefore

$$\begin{aligned} &\mathrm{E}\ \left(\exp\{-i\omega(|y_{m+M,n}|^2 - |y_{m,n}|^2)\} \mid \Phi_1 \right) \\ &\quad = \det{}^{-1} \left(I + i\omega \left[\begin{array}{cc} \rho T/M & (\rho T/M) d_m \\ (\rho T/M) d_m & \frac{\rho T/M}{1+\rho T/M}[1 + (\rho T/M) d_m^2] \end{array} \right] \left[\begin{array}{cc} -1 & 0 \\ 0 & 1 \end{array} \right] \right) \\ &\quad = \frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)} \left[\omega^2 - i\omega + \frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)} \right]^{-1} \\ &\quad = \frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)[(\omega - i/2)^2 + a_m^2]} \end{aligned} \tag{B.6}$$

when $d_m < 1$, where

$$a_m \stackrel{\mathrm{def}}{=} \sqrt{\frac{1}{4} + \frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)}}.$$

When $d_m = 1$, the characteristic function is identically one for all ω.

The region of convergence of the expectation in (B.6) is $\frac{1}{2} - a_m < \mathrm{Im}(\omega) < \frac{1}{2} + a_m$. The characteristic function of $\mathrm{tr}\,\{Y_2^\dagger Y_2 - Y_1^\dagger Y_1\}$ is the product of the M terms (B.6) raised to the Nth power. We invert the characteristic function, artfully choosing a particular integration contour within its region of convergence to obtain the error probability as

$$\begin{aligned} P_e &= \frac{1}{2\pi} \int_0^\infty dz \int_{-\infty+i/2}^{\infty+i/2} d\omega \exp(i\omega z) \cdot \prod_{m=1}^{M} \mathrm{E}\ \left(\exp\{-i\omega(|y_{m+M,n}|^2 - |y_{m,n}|^2)\} \mid \Phi_1 \right) \qquad \text{(B.7)} \\ &= -\frac{1}{2\pi i} \int_{-\infty+i/2}^{\infty+i/2} d\omega \frac{1}{\omega} \prod_{\substack{m=1 \\ d_m<1}}^{M} \left[\frac{1 + \rho T/M}{(\rho T/M)^2 (1 - d_m^2)[(\omega - i/2)^2 + a_m^2]} \right]^N. \end{aligned}$$

The exponential decay of the inner integral in (B.7) as $z \to \infty$ for $\mathrm{Im}\,(\omega) = 1/2$ justifies the above

interchange of integration. With a change of variables, we obtain

$$P_e = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega + i/2} \prod_{\substack{m=1 \\ d_m<1}}^{M} \left[\frac{1+\rho T/M}{(\rho T/M)^2(1-d_m^2)(\omega^2+a_m^2)}\right]^N, \tag{B.8}$$

where the integration is now along the real axis. We close the contour of integration from the positive real axis to the negative real axis with a semicircle sweeping the entire upper-half complex plane in a counterclockwise direction. Since each term in the product in the integrand decays for large $|\omega|$ as $1/|\omega|^{2N}$, the semicircle itself does not contribute to the total integral but encloses all the integrand's upper-half plane poles $ia_1, \ldots, ia_M$. The integral may therefore be evaluated by Cauchy's theorem;

$$P_e = \sum_j \mathrm{Res}_{\omega=ia_j} \left\{ -\frac{1}{\omega + i/2} \prod_{\substack{m=1 \\ d_m<1}}^{M} \left[\frac{1+\rho T/M}{(\rho T/M)^2(1-d_m^2)(\omega^2+a_m^2)}\right]^N \right\}. \tag{B.9}$$

This is equation (17).

We evaluate this expression for the special case of equal singular values $d_1 = \ldots = d_M = d < 1$. In this case, we must evaluate the residue of an order MN pole,

$$P_e = -\left[\frac{1+\rho T/M}{(\rho T/M)^2(1-d^2)}\right]^{MN} \frac{1}{\Gamma(MN)} \left[\frac{\mathrm{d}^{MN-1}}{\mathrm{d}\omega^{MN-1}} \frac{1}{(\omega+i/2)(\omega+ia)^{MN}}\right]_{\omega=ia}.$$

The following easily verified identity,

$$\frac{\mathrm{d}^k}{\mathrm{d}\omega^k} \frac{1}{(\omega+i/2)(\omega+ia)^{MN}} = (-1)^k \sum_{j=0}^{k} \frac{\Gamma(MN+k-j)\Gamma(k+1)}{\Gamma(MN)\Gamma(k+1-j)} \cdot (\omega+i/2)^{-(1+j)} (\omega+ia)^{-(MN+k-j)},$$

for $k = MN-1$ gives

$$P_e = \left[\frac{1+\rho T/M}{(2+\rho T/M)^2 - (\rho dT/M)^2}\right]^{MN} \cdot \sum_{j=0}^{MN-1} \frac{\Gamma(2MN-1-j)}{\Gamma(MN)\Gamma(MN-j)} \left[\frac{2\sqrt{(2+\rho T/M)^2-(\rho dT/M)^2}}{(\rho T/M)(1-d^2)^{1/2} + \sqrt{(2+\rho T/M)^2-(\rho dT/M)^2}}\right]^{1+j}.$$

B.2 Chernoff upper bound and monotonicity

We evaluate the Chernoff upper bound in an unconventional but concise way. Since P_e is real-valued, we take the real part of (B.8) to obtain

$$\begin{aligned} P_e &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{\substack{m=1 \\ d_m<1}}^{M} \left[\frac{1+\rho T/M}{(\rho T/M)^2(1-d_m^2)(\omega^2+a_m^2)}\right]^N, \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{m=1}^{M} \left[\frac{1}{1+\frac{(\rho T/M)^2(1-d_m^2)(\omega^2+1/4)}{1+\rho T/M}}\right]^N \qquad \text{(B.10)} \\ &\leq \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{m=1}^{M} \left[\frac{1}{1+\frac{(\rho T/M)^2(1-d_m^2)}{4(1+\rho T/M)}}\right]^N \\ &= \frac{1}{2}\prod_{m=1}^{M} \left[\frac{1}{1+\frac{(\rho T/M)^2(1-d_m^2)}{4(1+\rho T/M)}}\right]^N, \qquad \text{(B.11)} \end{aligned}$$

which is (18). It turns out that (B.11) is, in fact, exactly the Chernoff bound obtained by computing (see, e.g. [21])

$$P_e \leq \frac{1}{2}e^{\mu(\lambda)},$$

where

$$\mu(\lambda) = \ln \mathrm{E}\left\{\exp\left[\lambda\left(\ln p(X \mid \Phi_2) - \ln p(X \mid \Phi_1)\right)\right] \mid \Phi_1\right\},$$

and where $0 \leq \lambda \leq 1$ is a free parameter that is chosen to minimize $\mu(\lambda)$. To help see this, we note that $\mu(\lambda)$ is merely the logarithm of the previously computed characteristic function for $\omega = i\lambda$, and is minimized at $\lambda = 1/2$. The exact expression for P_e is derived in (B.8) by integrating the characteristic function along the line $\mathrm{Im}\,(\omega) = 1/2$. This process "tilts" the likelihood ratio by just the right amount needed to obtain (B.11) as the Chernoff bound.

Finally, to see that decreasing any d_m decreases the total error probability, observe that, for any ω, the integrand in (B.10) decreases as d_m decreases.

C Appendix: Two-Signal P_e for H Known—Exact and Chernoff Upper Bound

We follow the same strategy as in Appendix B, and therefore abbreviate the discussion. With H known, albeit random, the average error probability, given that Φ_1 is transmitted, is

$$\begin{aligned} P_{e|1} &= \mathrm{P}\left(-\mathrm{tr}\left\{\left[X-(\rho T/M)^{1/2}\Phi_2 H\right]\left[X-(\rho T/M)^{1/2}\Phi_2 H\right]^\dagger \right.\right. \\ &\quad \left.\left. -\left[X-(\rho T/M)^{1/2}\Phi_1 H\right]\left[X-(\rho T/M)^{1/2}\Phi_1 H\right]^\dagger\right\} > 0 \mid \Phi_1\right) \\ &= \mathrm{P}\left(\mathrm{tr}\left\{\left[(\rho T/M)^{1/2}(\Phi_2-\Phi_1)H+W\right]\left[(\rho T/M)^{1/2}(\Phi_2-\Phi_1)H+W\right]^\dagger + WW^\dagger\right\} > 0\right). \end{aligned}$$

We use the singular value decomposition, $\Phi_2 - \Phi_1 = \Xi\Delta\Omega^\dagger$, where Δ is diagonal, real, and nonnegative, and Ξ and Ω are unitary matrices. Because H and $\Omega^\dagger H$ have the same distribution, and W and ΞW have the same distribution, we have that

$$P_{e|1} = \mathrm{P}\left(\mathrm{tr}\left\{\left[(\rho T/M)^{1/2}\Delta H+W\right]\left[(\rho T/M)^{1/2}\Delta H+W\right]^\dagger + WW^\dagger\right\} > 0\right). \tag{C.1}$$

The probability of error only depends on the singular values, and hence $P_e = P_{e|1} = P_{e|2}$. The singular value decomposition implies that $\Delta = \Xi^\dagger\Phi_2\Omega - \Xi^\dagger\Phi_1\Omega$, or $\delta_m = [\Xi^\dagger\Phi_2\Omega]_{mm} - [\Xi^\dagger\Phi_1\Omega]_{mm}$, where the columns of each of the bracketed matrices are orthonormal unit vectors. Consequently $0 \le \delta_m \le 2$.

C.1 Exact P_e

As in Appendix B, we take the characteristic function of the log-likelihood ratio (the expectation being with respect to H as well as W), and we obtain the probability of error as the integral

$$\begin{aligned} P_e &= -\frac{1}{2\pi i}\int_{-\infty+i/2}^{\infty+i/2} d\omega \frac{1}{\omega}\prod_{m=1}^{M}\left[\frac{1}{1+(\omega^2-i\omega)(\rho T/M)\delta_m^2}\right]^N \\ &= -\frac{1}{2\pi i}\int_{-\infty+i/2}^{\infty+i/2} d\omega \frac{1}{\omega}\prod_{\substack{m=1\\ \delta_m>0}}^{M}\left[\frac{1}{(\rho T/M)\delta_m^2[(\omega-i/2)^2+\alpha_m^2]}\right]^N \\ &= -\frac{1}{2\pi i}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega+i/2}\prod_{\substack{m=1\\ \delta_m>0}}^{M}\left[\frac{1}{(\rho T/M)\delta_m^2(\omega^2+\alpha_m^2)}\right]^N \end{aligned} \tag{C.2}$$

$$= \sum_j \mathrm{Res}_{\omega=i\alpha_j} \left\{ -\frac{1}{\omega + i/2} \prod_{\substack{m=1 \\ \delta_m>0}}^{M} \left[\frac{1}{(\rho T/M)\delta_m^2(\omega^2+\alpha_m^2)} \right]^N \right\},$$

where

$$\alpha_m = \sqrt{\frac{1}{4} + \frac{1}{\rho T \delta_m^2/M}}.$$

This proves equation (19).

For the special case where $\delta_1 = \ldots = \delta_M = \delta$, we have the exact expression

$$P_e = \frac{1}{(4+\rho T\delta^2/M)^{MN}} \sum_{j=0}^{MN-1} \frac{\Gamma(2MN-1-j)}{\Gamma(MN)\Gamma(MN-j)} \left[\frac{2\sqrt{4+\rho T\delta^2/M}}{\sqrt{\rho T\delta^2/M}+\sqrt{4+\rho T\delta^2/M}} \right]^{1+j}.$$

C.2 Chernoff upper bound and monotonicity

As in Appendix B, the Chernoff bound is computed by applying an elementary inequality to the exact probability of error (C.2). The result is

$$\begin{aligned} P_e &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{\substack{m=1 \\ \delta_m>0}}^{M} \left[\frac{1}{(\rho T/M)\delta_m^2(\omega^2+\alpha_m^2)} \right]^N \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{m=1}^{M} \left[\frac{1}{1+(\rho T/M)\delta_m^2(\omega^2+1/4)} \right]^N \qquad \text{(C.3)} \\ &\le \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \frac{1}{\omega^2+1/4} \prod_{m=1}^{M} \left[\frac{1}{1+\frac{\rho T}{4M}\delta_m^2} \right]^N \\ &= \frac{1}{2} \prod_{m=1}^{M} \left[\frac{1}{1+\frac{\rho T}{4M}\delta_m^2} \right]^N, \end{aligned}$$

which is (20).

Finally, to see that increasing any δ_m decreases the total error probability, observe that, for any ω, the integrand in (C.3) decreases as δ_m increases.

References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading Channels: Information-theoretic and communications aspects," *IEEE Trans. Info. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[2] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Info. Theory*, vol. 18, pp.460-473, July 1972.

[3] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*, 2nd edition, Springer-Verlag, New York, 1993.

[4] T. Cover and J. Thomas, *Elements of Information Theory*, New York: John Wiley, 1991.

[5] R. L. Ellis and D. C. Lay, "Rank-preserving extensions of band matrices", *Linear and Multilinear Algebra*, vol. 26, pp. 147–179, 1990.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.

[8] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Info. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[9] W. C. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1993.

[10] G. A. Kabatyanskii and V. I. Levenshtein, "Bounds for packings on a sphere and in space," *Problems of Information Transmission*, vol. 14, no. 1, pp. 1–17, 1978.

[11] Y. Kofman, E. Zehavi, and S. Shamai (Shitz), "nd-Convolutional codes—Parts I and II" *IEEE Trans. Info. Theory*, vol. 43, pp. 558–589, Mar. 1997.

[12] J. E. Mazo, "Some theoretical observations on spread-spectrum communications," *Bell. Sys. Tech. J.*, vol. 58, pp. 2013–2023, Nov. 1979.

[13] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.

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

[15] N. Seshadri and J. H. Winters, "Two signalling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wire. Inf. Net.*, vol. 1, no. 1, pp. 49–59, 1994.

[16] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, New York: McGraw-Hill, 1966.

[17] T. Söderström and P. Stoica, *System Identification*, London: Prentice Hall, 1989.

[18] V. Tarokh, N. Seshadri and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Info. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[19] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Laboratories internal Technical Memorandum, 1995.

[20] FRAMES Multiple access proposal for the UMTS radio interface-SMG2, Workshop on UMTS Radio Interface Technologies, Dec. 1996.

[21] H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I. New York: John Wiley and Sons, 1968.

[22] L. R. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Info. Theory*, vol. 20, pp. 397–399, May 1974.

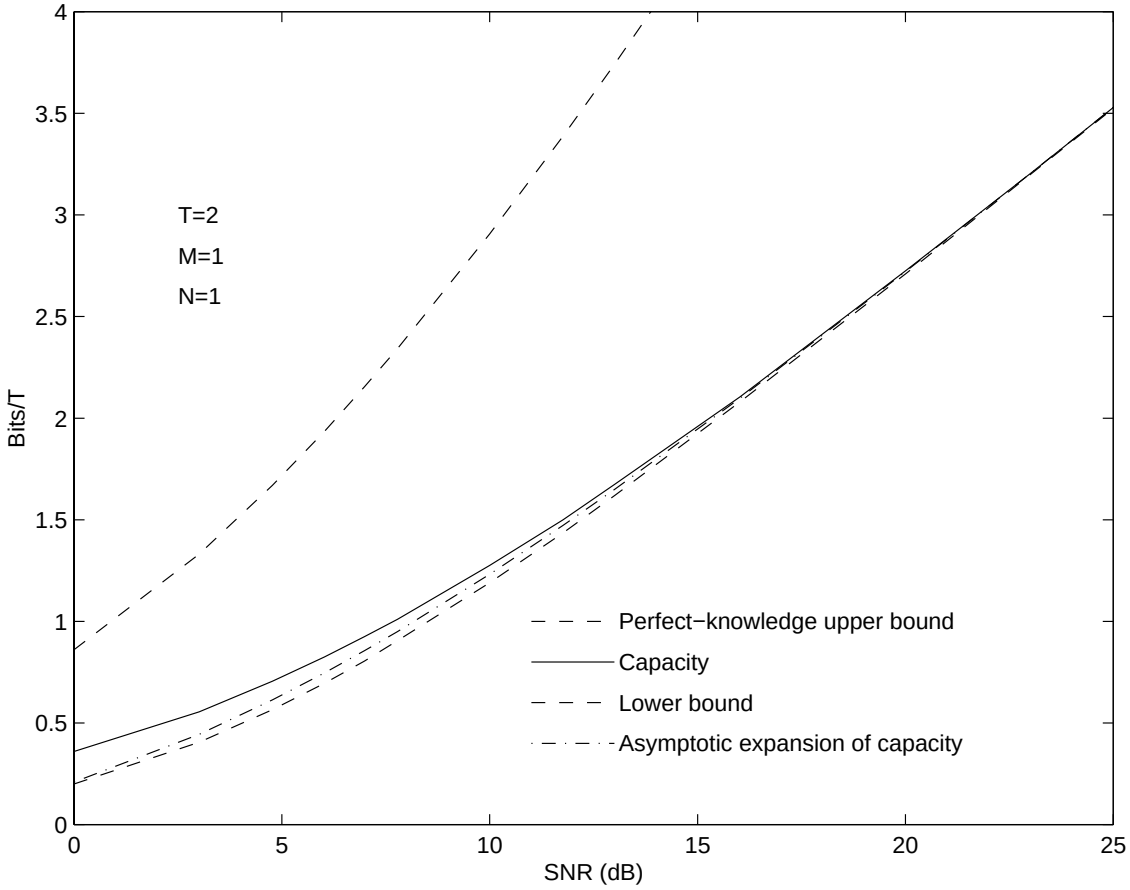


Figure 1: Normalized capacity, and upper and lower bounds, versus SNR ρ ($T = 2$, one transmitter and one receiver antenna). The lower bound and capacity meet as $\rho \to \infty$. However, unlike the case where $T \to \infty$, the capacity never meets the perfect-knowledge upper bound.

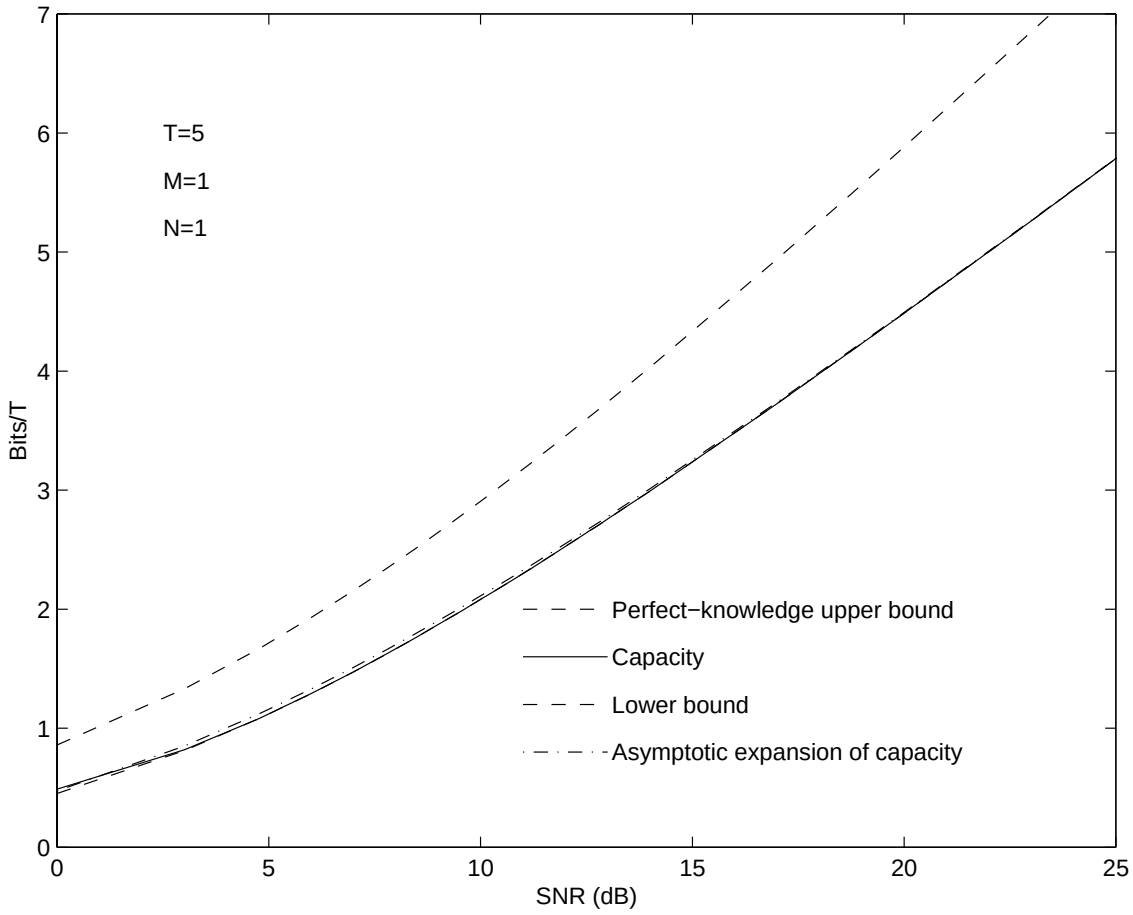


Figure 2: Normalized capacity, and upper and lower bounds, versus SNR ρ as in Figure 1, but with $T = 5$.

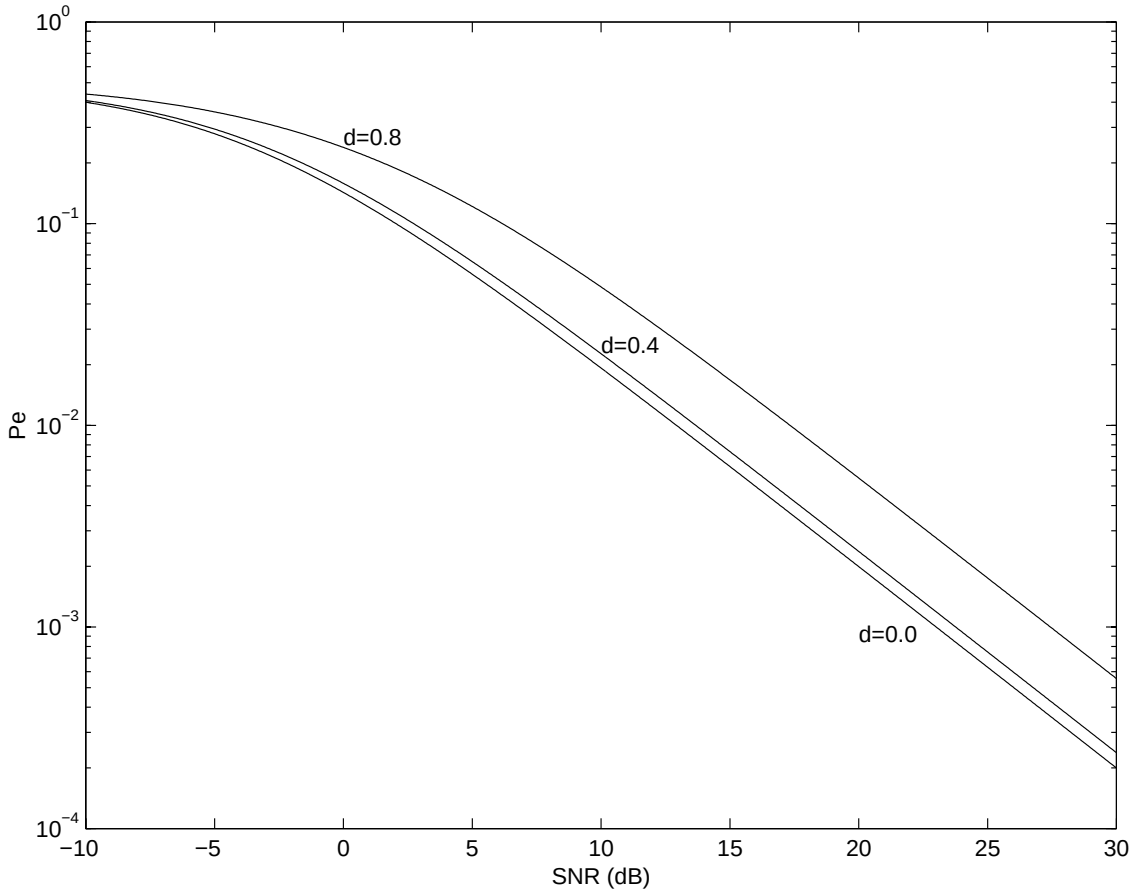


Figure 3: Two-signal probability of error vs. SNR for one transmitter and one receiver antenna ($M = N = 1$), $T = 5$, and $d = 0.0$, 0.4, 0.8.

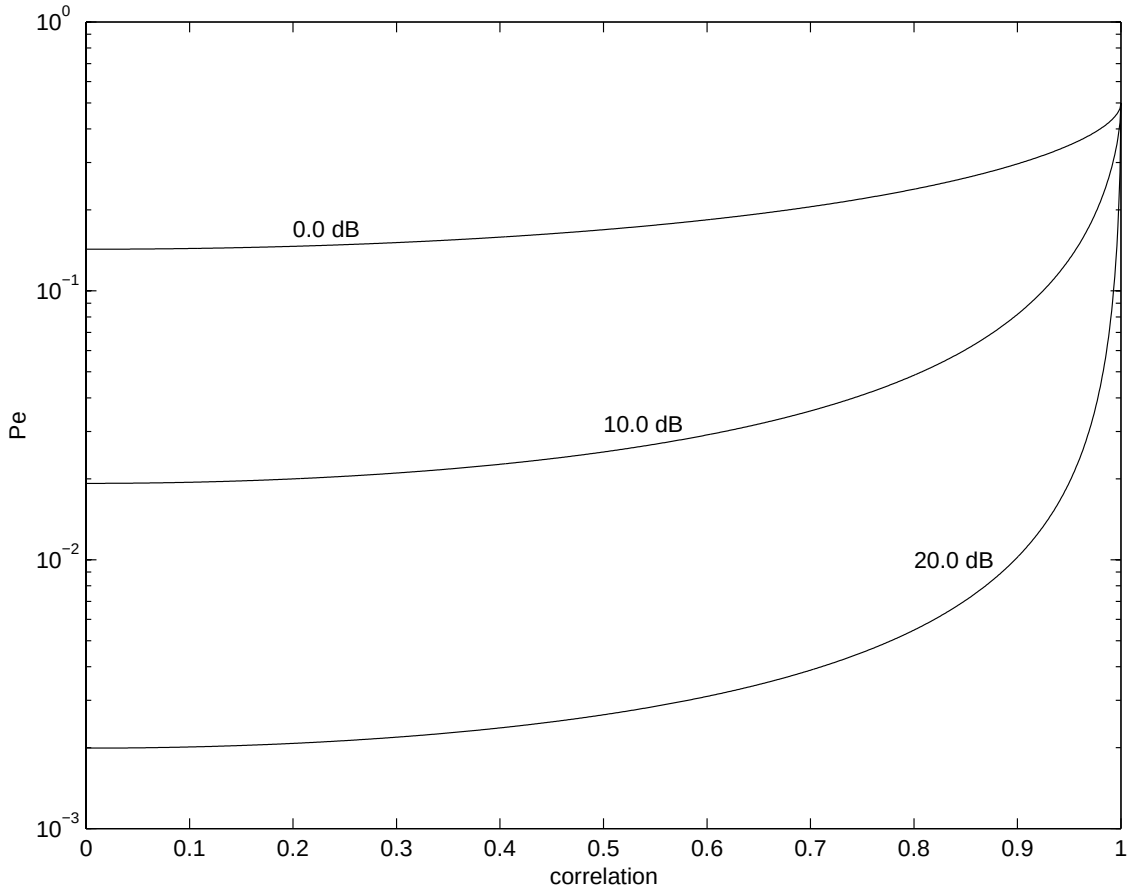


Figure 4: Two-signal probability of error vs. correlation d_1 for one transmitter and receiver antenna ($M = N = 1$), $T = 5$, and SNR=0, 10, 20 dB.

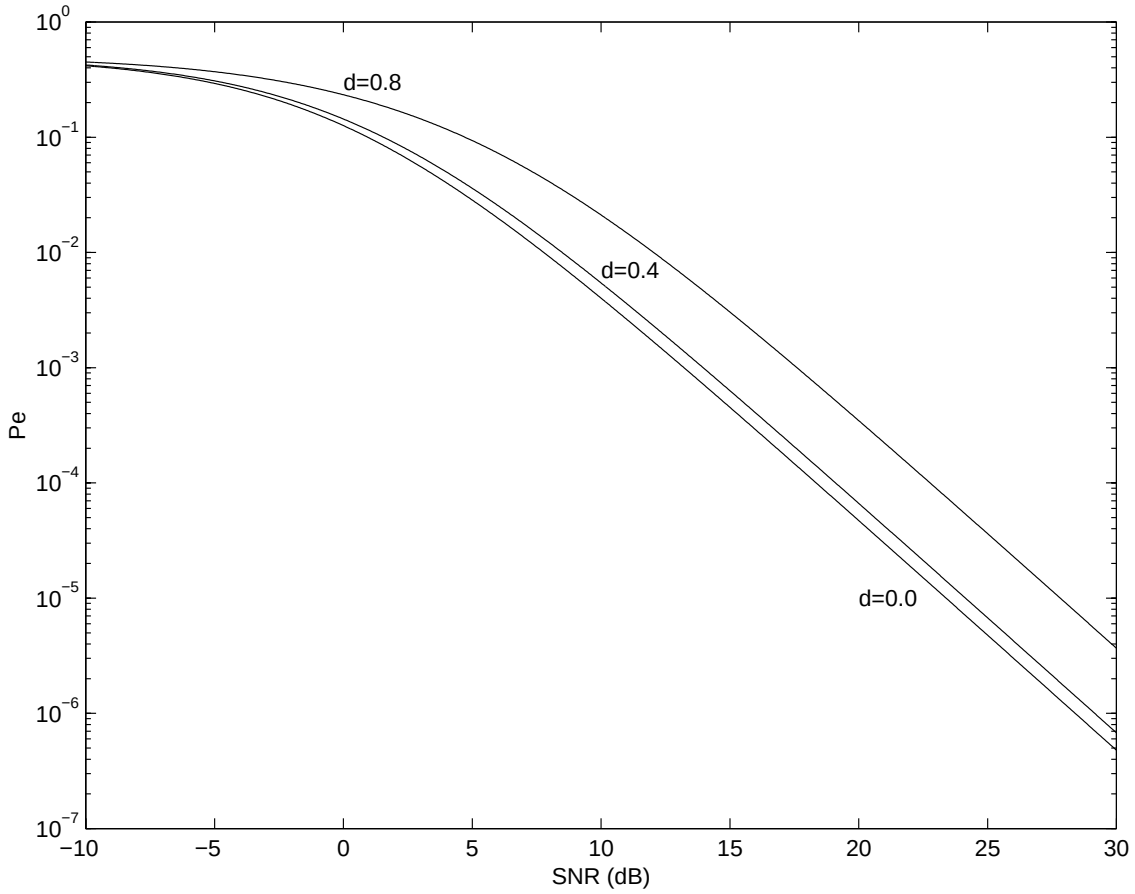


Figure 5: Two-signal probability of error vs. SNR for two transmitter antennas and one receiver antenna ($M = 2$, $N = 1$), $T = 5$, and $d_1 = d_2 = d = 0.0$, 0.4, 0.8.

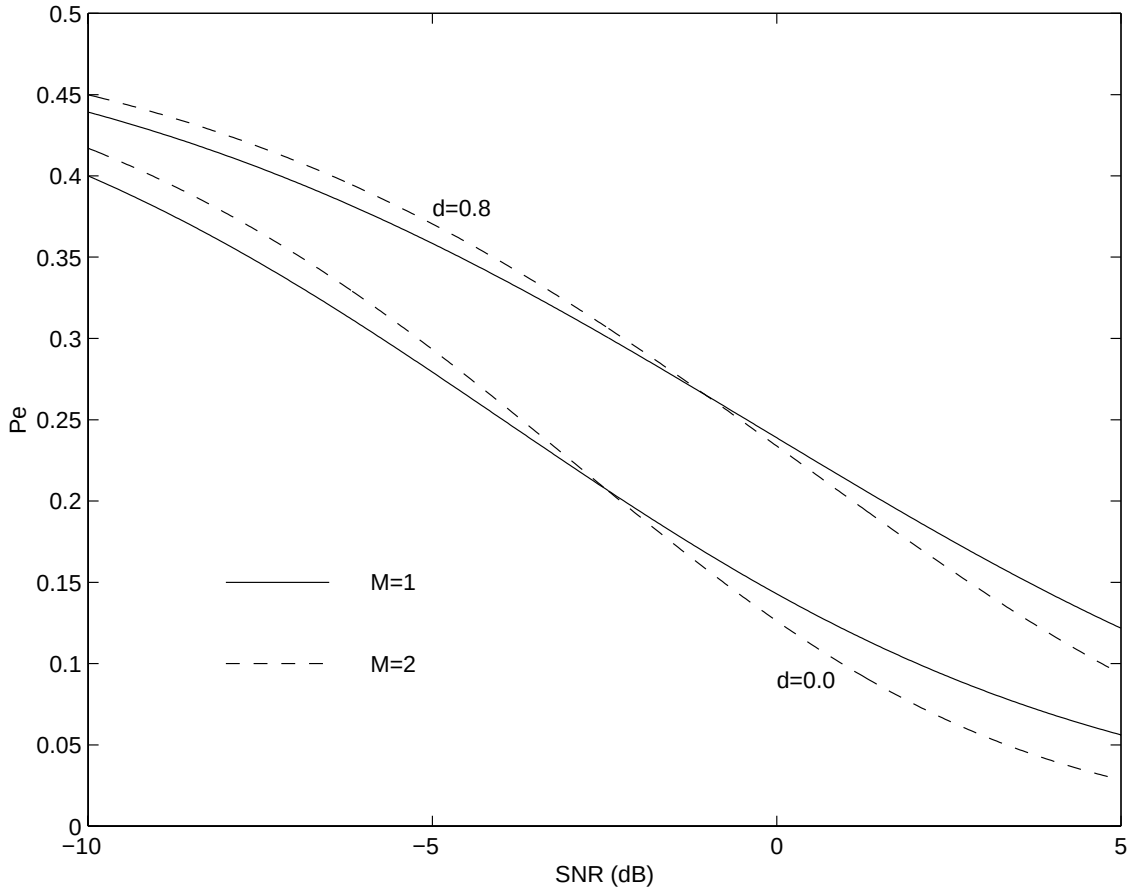


Figure 6: Two-signal probability of error vs. SNR for one ($M = 1$, solid curves), and two ($M = 2$, dashed curves) transmitter antennas, one receiver antenna ($N = 1$), $T = 5$, and $d = 0.0$, 0.8.

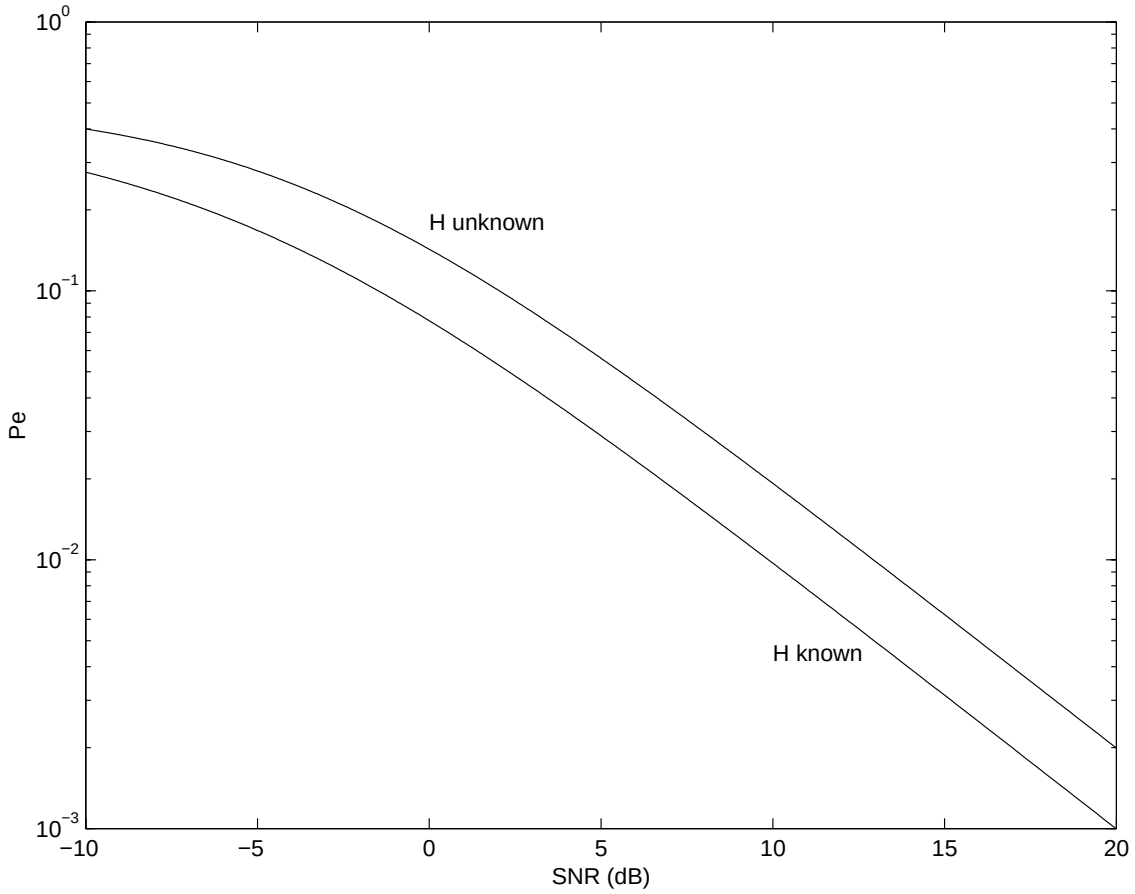


Figure 7: Two-signal probability of error vs. SNR for H unknown ($d = 0$) compared with H known ($\delta = 1.414\ldots$), and one transmitter and one receiver antenna ($M = N = 1$), and $T = 5$.

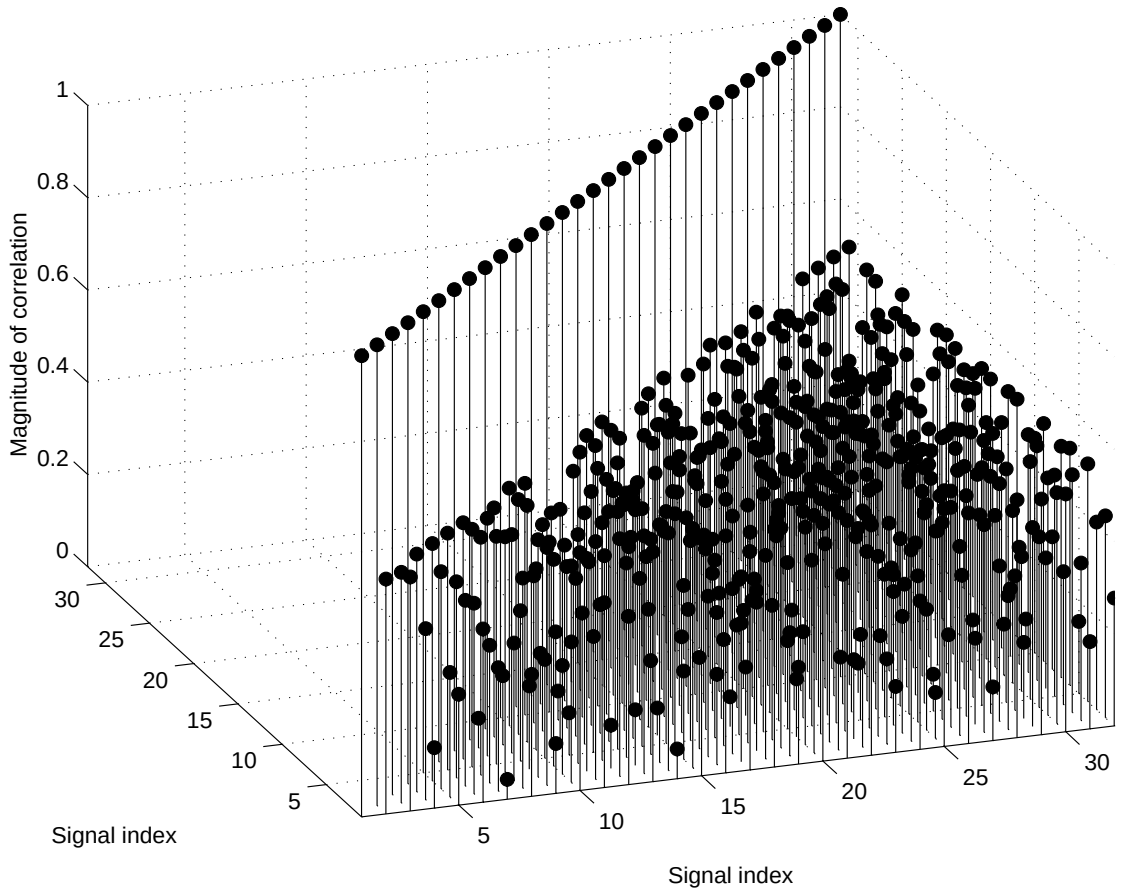


Figure 8: Magnitudes of correlations between $\Phi_1, \ldots, \Phi_{32}$ for $T = 5$. The diagonal entries with value 1.0 represent each signal correlated with itself.

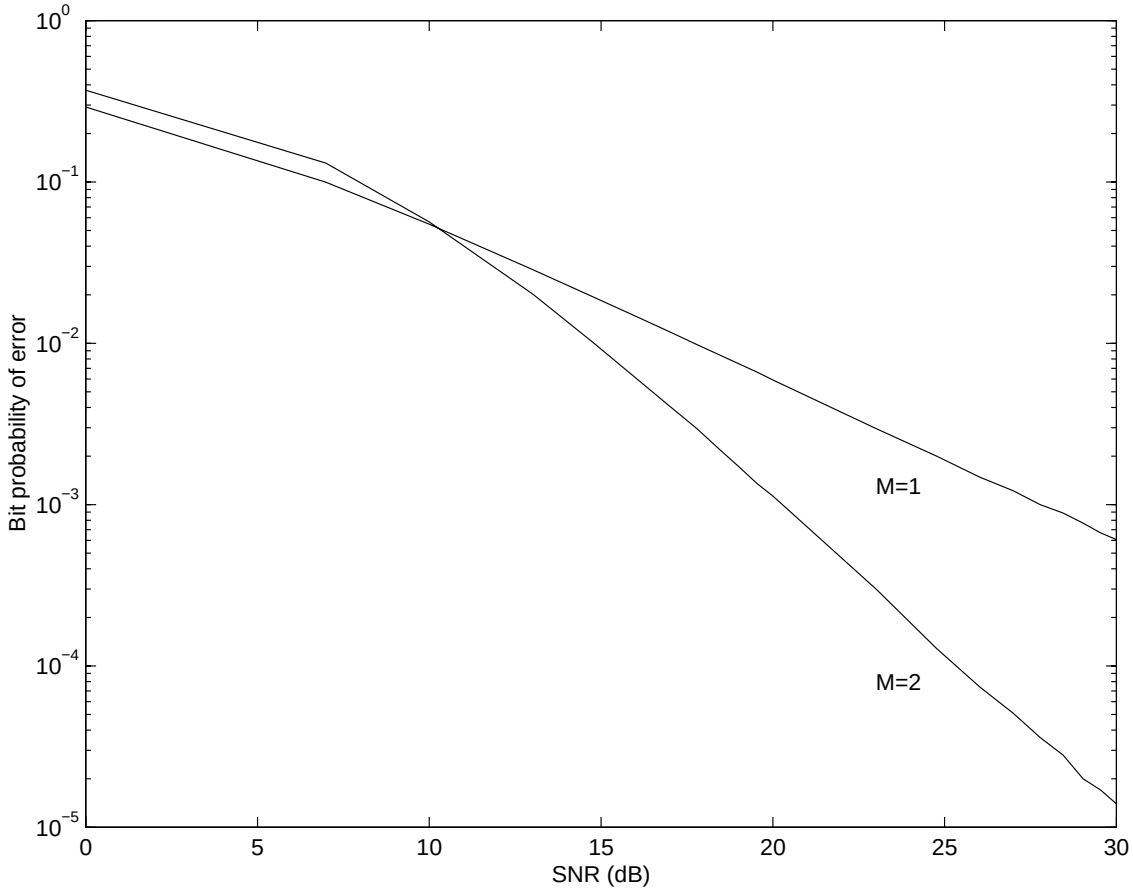


Figure 9: Performance of unitary space-time constellations for $M = 1$ versus $M = 2$ transmitter antennas for $T = 5$ as a function of SNR ρ, with $R = 1$ bit/channel use.

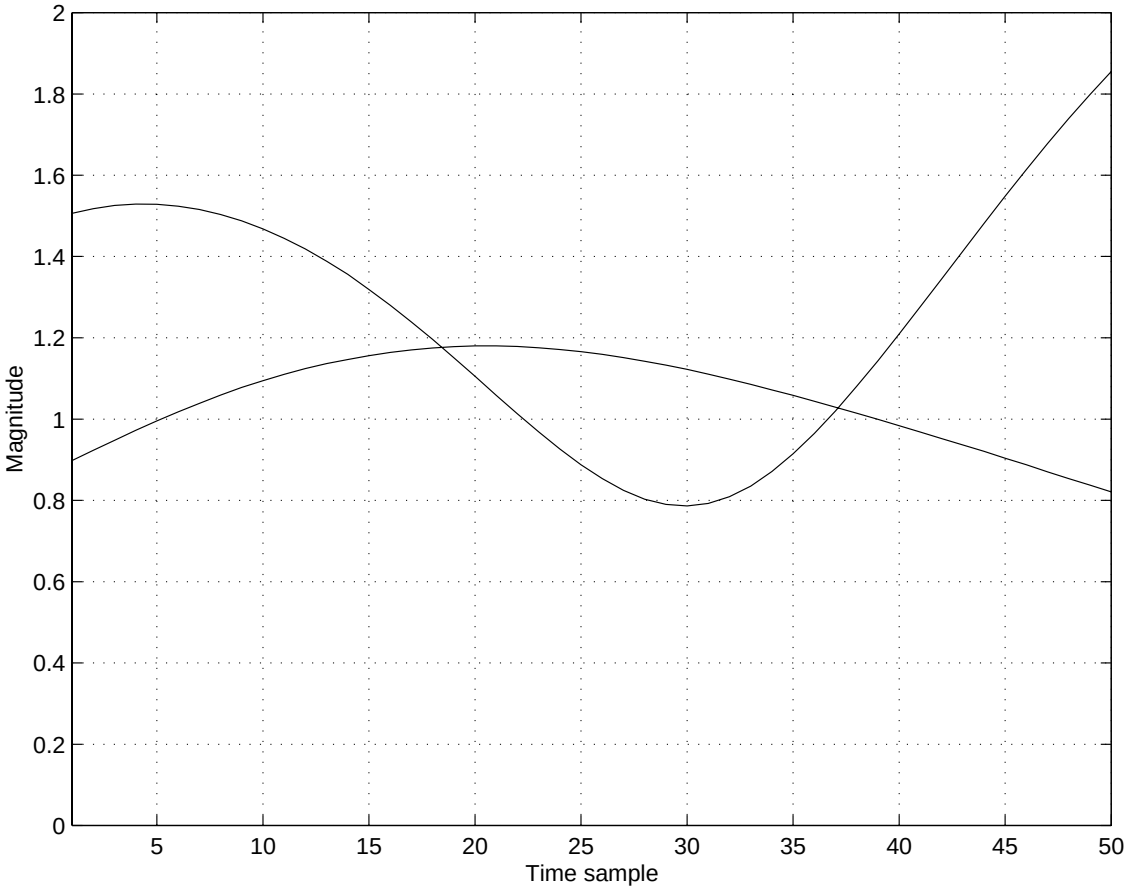


Figure 10: Magnitude of two typical independent realizations of a Jakes fading process with $f_d = 0.01$ cycles/sample.

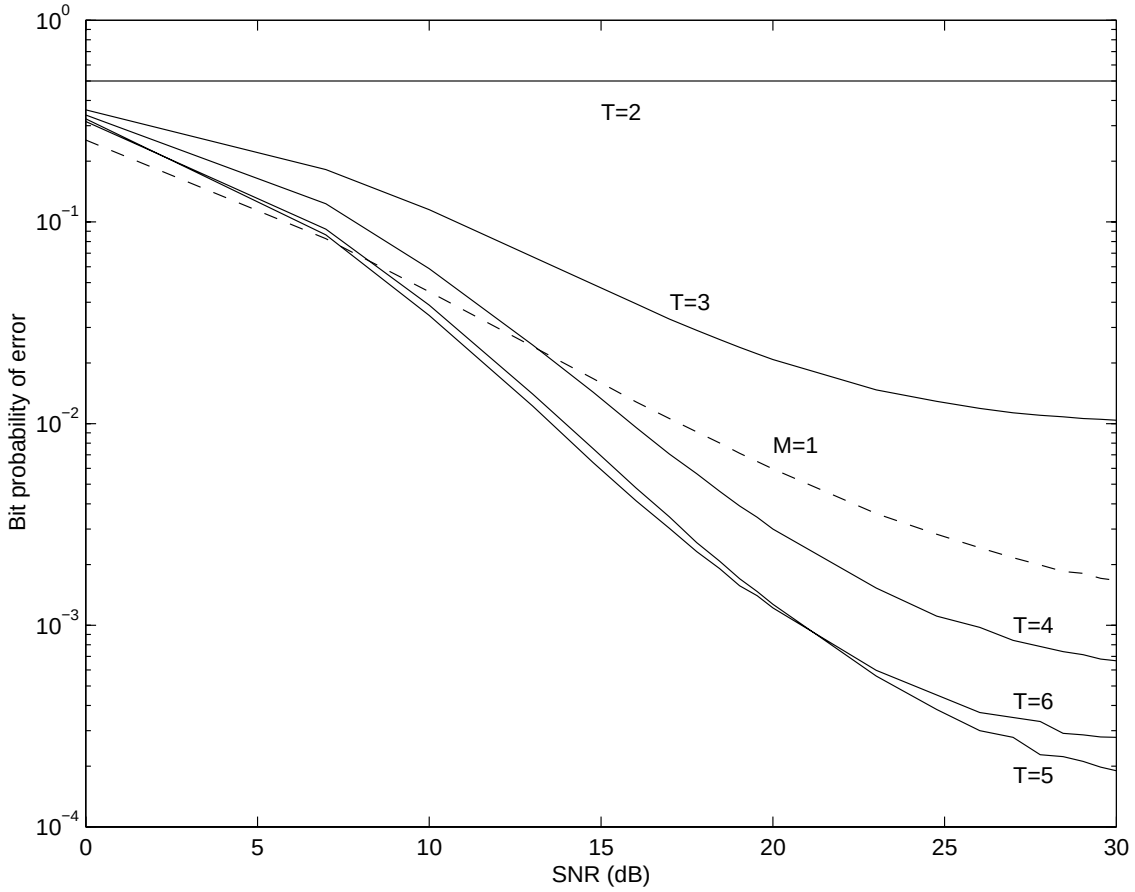


Figure 11: Unitary space-time modulation performance for one (dashed line) and two (solid lines) transmitter antennas sending $R = 1$ bit per channel use with constellations designed for $T = 2, \ldots, 6$. The fading is a Jakes process with $f_d = 0.01$ cycles/sample and there is one receiver antenna. The one-antenna probability of error varies little with T and is well-approximated by the D-BPSK dashed line. The two-antenna probabilities of error vary greatly with T. The best overall performance for high SNR occurs for $T = 5$.