Space-Time Autocoding

BERTRAND M. HOCHWALD   THOMAS L. MARZETTA   BABAK HASSIBI

Mathematical Sciences Research Center
Bell Laboratories
Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974
{hochwald, tlm, hassibi}@research.bell-labs.com

Abstract

Prior treatments of space-time communications in Rayleigh flat fading generally assume that channel coding covers either one fading interval—in which case there is a nonzero "outage capacity"—or multiple fading intervals—in which case there is a nonzero Shannon capacity. However, we establish conditions under which channel codes span only one fading interval and yet are arbitrarily reliable. In short, space-time signals are their own channel codes. We call this phenomenon *space-time autocoding*, and the accompanying capacity the *space-time autocapacity*.

Let an M-transmitter-antenna, N-receiver-antenna Rayleigh flat fading channel be characterized by an $M \times N$ matrix of independent propagation coefficients, distributed as zero-mean, unit-variance complex Gaussian random variables. This propagation matrix is unknown to the transmitter, it remains constant during a T-symbol coherence interval, and there is a fixed total transmit power. Let the coherence interval and number of transmitter antennas be related as $T = \beta M$ for some constant β. A $T \times M$ matrix-valued signal, associated with $R \cdot T$ bits of information for some rate R is transmitted during the T-symbol coherence interval. Then there is a positive space-time autocapacity C_a such that for all $R < C_a$, the block probability of error goes to zero as the pair $(T, M) \to \infty$ such that $T/M = \beta$. The autocoding effect occurs whether or not the propagation matrix is known to the receiver, and $C_a = N \log(1 + \rho)$ in either case, independently of β, where ρ is the expected SNR at each receiver antenna. Lower bounds on the cutoff rate derived from random Unitary Space-Time signals suggest that the autocoding effect manifests itself for relatively small values of T and M. For example, within a single coherence interval of duration $T = 16$, for $M = 7$ transmitter antennas and $N = 4$ receiver antennas, and an 18 dB expected SNR, a total of 80 bits (corresponding to rate $R = 5$) can theoretically be transmitted with a block probability of error less than 10^{-9}, all without any training or knowledge of the propagation matrix.

Index Terms—Space-time coding, multi-element antenna arrays, Unitary Space-Time Modulation, wireless communications, eigenvalues of random matrices

1 Introduction

The piecewise-constant model for flat fading (also called the *block fading* model [1, 2, 3]) has been highly successful in advancing the theory of the multiple-antenna wireless communication link [4, 5, 6, 7, 8]. Under this model, one has an M-element transmitter antenna array, and an N-element receiver antenna array. The propagation between the mth transmitter and nth receiver antennas is described by a complex scalar propagation coefficient h_{mn}; these coefficients are statistically independent, zero-mean, unit-variance complex Gaussian random variables and are unknown to the transmitter. In the piecewise-constant fading model, the $M \times N$ propagation matrix H remains constant for T symbol periods, after which it jumps to a new, independent value which it holds for another T symbol periods, and so on. This mimics, in a tractable manner, the salient features of actual fading channels. During every T symbol coherence interval, a single user transmits a $T \times M$ signal matrix S, and another user receives a $T \times N$ signal X that is multiplied by the propagation matrix, and corrupted by additive noise in the receiver.

The piecewise-constant fading model permits an application of Shannon theory in a simple and natural way, since the two random channel quantities, the propagation matrix and the receiver noise matrix, take on independent values for every coherence interval. If we regard every use of the channel as the transmission of a matrix S over the coherence interval, the channel is memoryless. Thus Shannon capacity, associated with coding over many independent T-symbol coherence intervals, is obtained by maximizing the mutual information between the matrix-valued transmitted and received signals [5, 8]. Shannon theory concludes that arbitrarily reliable communication is possible as the propagation coefficients evolve in time, and the effects of unfavorable fading are mitigated by favorable fading.

This paper describes a less obvious effect: that arbitrarily reliable communication can be achieved within a *single* coherence interval if the duration of the coherence interval and the number of transmitter antennas simultaneously increase without limit. Let the number of receiver antennas N be constant, and let the total transmitted power be constant. Let $R \cdot T$ bits of information, for some constant rate R, be encoded into a signal matrix S which is transmitted within a single T-symbol coherence interval. Let $T/M = \beta$ where β is any positive constant. We show that there is a positive *space-time autocapacity* C_a such that for all $R < C_a$, there is a code such that the block (of length T) probability of error goes to zero as $(T, M) \to \infty$. For brevity, we refer to C_a simply as the *autocapacity*, where the name reflects the fact that the $T \times M$ matrix-valued signals act as their own channel codes.

The autocapacity turns out to be the least upper bound on a certain Shannon capacity. Where Shannon coding relies on temporal diversity, multiple-antenna space-time autocoding relies instead on spatial diversity: the space-time autocoding effect requires both M and T to grow. If either quantity is fixed it is impossible to transmit at a positive rate during a single coherence interval with arbitrary reliability. Any bad propagation coefficients remain bad throughout the coherence interval, but, if T and M both grow, the ever-increasing number of transmitter antennas ensures that there are a compensating number of good propagation coefficients.

Communication links where the coding interval is smaller than the coherence time of the fading [3] have been treated in the past with theories involving an *outage capacity* [2] or *delay-limited capacity* (see, for example, [11]).

The outage capacity [2] involves a tradeoff between the transmission rate and an outage probability. The channel has an instantaneous mutual information that depends on the value of the propagation matrix. If the mutual information happens to be less than the transmission rate, this is called an outage because the channel cannot support this rate. Only by setting the transmission rate to zero is the outage probability also zero. In contrast, autocoding achieves arbitrary reliability for a positive transmission rate.

The delay-limited capacity is used in [9] and [11] to describe the largest rate that can be achieved with a coding delay that is independent of how slowly the fading process changes. It is generally assumed in these references that the transmitter knows the channel or that the instantaneous mutual information has a positive lower bound—neither condition is satisfied with our channel model.

Section 2 reviews the piecewise constant fading model as well as earlier Shannon-theoretic results. Section 3 establishes the existence of the autocoding effect and obtains an explicit formula for autocapacity. Section 4 considers the random coding exponent for a single coherence interval, and obtains useful lower bounds on the cutoff rate. Section 5 illustrates the theory with numerical results. The mathematical machinery required to obtain the results of Sections 4 and 5 is developed in the appendices.

2 Piecewise-Constant Fading and Shannon Capacity

The piecewise-constant fading model constitutes a memoryless channel for matrix-valued transmitted and received signals. This permits a straightforward application of Shannon theory. We consider the cases where the propagation matrix is known and unknown to the receiver. In either case, the matrix is unknown to the transmitter.

2.1 Signal model

A communication link comprising M transmitter antennas and N receiver antennas operates in Rayleigh flat fading. The propagation is characterized by an $M \times N$ matrix of propagation coefficients, which are independent, zero-mean, unit-variance, complex Gaussian $\mathcal{CN}(0,1)$. The propagation matrix remains constant for T symbol periods, after which it jumps to a new, statistically independent value which it holds for another T symbol periods, and so on.

During every T-symbol coherence interval a matrix-valued signal S of dimension $T \times M$ is fed into the array of transmitter antennas. The average variance of the components of S is constrained to be equal to one,

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

The signal that is measured by the array of receive antennas is a $T \times N$ matrix X,

$$X = \sqrt{\frac{\rho}{M}} SH + W, \tag{2}$$

where W is a $T \times N$ matrix of additive receiver noise whose components are independent $\mathcal{CN}(0,1)$. The constraint (1), when combined with the normalization $1/\sqrt{M}$ in (2), implies that the total transmitted power remains constant as the number of transmit antennas changes, and that ρ represents the expected SNR at each receiver antenna.

The channel is completely characterized by the probability density for the received signal given the transmitted signal. This takes one of two forms depending on whether or not the propagation matrix is known to the receiver. If the matrix is known, the appropriate probability density is

$$p(X \mid S, H) = \frac{1}{\pi^{TN}} \exp\left(-\mathrm{tr}\left\{ \left[X - \sqrt{\rho/M} SH \right] \left[X - \sqrt{\rho/M} SH \right]^{\dagger} \right\} \right). \tag{3}$$

If the matrix is unknown, the appropriate probability density is

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

where $\Lambda = I_T + (\rho/M)\, SS^{\dagger}$, I_T denotes the $T \times T$ identity matrix, and "tr" denotes "trace."

2.2 Prior Shannon capacity results

Each channel use (consisting of the transmission of a matrix S and the reception of a matrix X) is independent of every other, so data theoretically can be transmitted at any rate less than the Shannon capacity, which is the least upper bound on the mutual information between the transmitted and the received matrices.

2.2.1 H known to receiver

When the propagation matrix is known to the receiver, the mutual information is

$$I(X, H; S) = \frac{1}{T} \mathrm{E} \left\{ \log \frac{p(X, H \mid S)}{p(X, H)} \right\}, \tag{5}$$

where a base-2 logarithm is used. This expression is examined in detail for $T = 1$ in [14, 13], and for $T > 1$ in [5]. We are interested in the regime $(T, M) \to \infty$, and the normalization $1/T$ implies that the units of mutual information are bits/symbol-period. Throughout the paper we use this convention for rate and capacity as well[1].

Determining capacity involves choosing $p(S)$ to maximize the mutual information (5), subject to a modified power constraint

$$\mathrm{E} \left\{ |s_{tm}|^2 \right\} = 1, \;\; \forall\, t, m.^2 \tag{6}$$

The principal results are:

1. The capacity is achieved when the elements of S are independent $\mathcal{CN}(0, 1)$;

2. The resulting *perfect knowledge* capacity is

$$C = \mathrm{E} \left\{ \log \det \left[I_M + \frac{\rho}{M} H H^\dagger \right] \right\}; \tag{7}$$

3. For fixed ρ and N,

$$\lim_{M \to \infty} C = N \log (1 + \rho). \tag{8}$$

[1]This normalization was *not* used in [5, 6].

[2]In this constraint, which is more restrictive than (1), we implicitly do not allow the allocated power to depend on H—which the transmitter does not know. When H is unknown to the receiver, this power constraint and the power constraint (1) yield the same capacity [5].

Note that the perfect knowledge capacity is independent of T and is an upper bound on the capacity when H is unknown.

2.2.2 H unknown

When the propagation matrix is not known, the mutual information is

$$I(X;S) = \frac{1}{T}\mathrm{E}\log\frac{p(X \mid S)}{p(X)}. \tag{9}$$

Determining capacity involves choosing the density $p(S)$ to maximize the mutual information (9), subject to the power constraint (1). This problem is studied for $T = M = N = 1$ in [15], and for general T, M, N in [5, 6]. The principal results are:

1. The capacity for $M > T$ is equal to the capacity for $M = T$;

2. In choosing $p(S)$ to attain capacity, the transmitted matrix can be constrained without penalty to have the factorization

$$S = \Phi V, \tag{10}$$

   where Φ and V are statistically independent, Φ is an isotropically random $T \times M$ unitary matrix (i.e., the M columns are orthonormal, and the joint probability density of the elements of Φ is unchanged if Φ is premultiplied by any deterministic unitary matrix), and V is an $M \times M$ real, nonnegative, diagonal matrix;

3. For constant values of M and N, as $T \to \infty$ (for fixed ρ) we may choose $V = \sqrt{T} \cdot I$ in the factorization (10).

4. For constant values of ρ, M, and N, the capacity as $T \to \infty$ approaches the capacity obtained for H known to the receiver (see also [1].

Computing capacity remains a difficult problem that has only been solved for some special cases. The third property above motivates the use of *Unitary Space-Time Modulation* (USTM), having a constellation of signals of the form $S = \sqrt{T}\Phi$, where $\Phi^{\dagger}\Phi = I_M$, and $M < T$ [6, 7]. If Φ is chosen randomly and isotropically, this choice of signals achieves capacity for large T and also in some cases for large ρ.

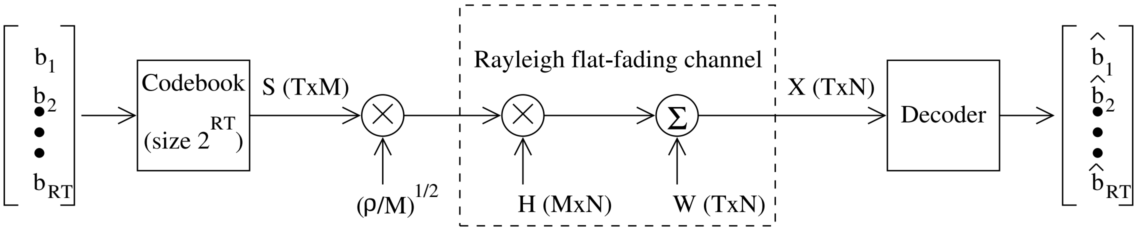

Figure 1: Wireless link comprising M transmitter and N receiver antennas. We wish to transmit $R \cdot T$ bits of information reliably in a *single* coherence interval T, where R is the rate in bits/symbol.

We note, for future reference, two properties of the Shannon capacity $C(\rho, T, M, N)$:

$$C(\rho, T, M_2, N) \geq C(\rho, T, M_1, N), \qquad M_2 \geq M_1, \tag{11}$$

and

$$C(\rho, L \cdot T, M, N) \geq C(\rho, T, M, N), \qquad L = 1, 2, \ldots. \tag{12}$$

3 The Autocoding Effect and Autocapacity

In this section we use the communication link, as displayed in Figure 1, to pass $R \cdot T$ bits of information (where R is the rate in bits/symbol) with arbitrary reliability during a single channel coherence interval T. We let T and M grow simultaneously and fix the total transmit power and number of receiver antennas. Arbitrary reliability cannot be achieved if M remains fixed while T increases because there are $M \cdot N$ propagation coefficients and they may occasionally be too small to support a given coding scheme and rate. Likewise, if T remains fixed while M increases, there are only $T \cdot N$ realizations of the receiver noise so there is always a chance that bad receiver noise will prevent successful communication. We conclude that the autocoding effect can exist only if T and M simultaneously grow large.

We adopt the following operational definition for autocapacity: Let $T = \beta M$, where β is a rational number. The autocapacity $C_a(\rho, \beta, N)$ is the largest nonnegative number such that for any rate $R < C_a(\rho, \beta, N)$ there exists a sequence of codes with this rate, associated with a sequence of pairs (T, M), such that the block probability of error of the codes goes to zero as $(T, M) \to \infty$.

In this section we use a simple construction to establish the existence of a positive-valued autocapacity, and we obtain a lower bound on autocapacity in terms of a least upper bound on a certain Shannon capacity.

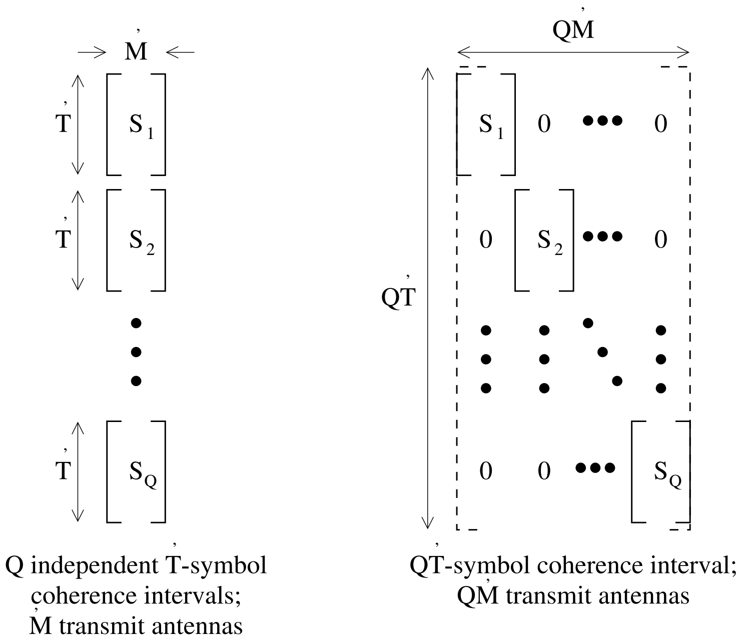

Figure 2: Q uses of an $\acute{M}$-antenna channel with coherence time $\acute{T}$ (shown on left) is equivalent to one use of a $Q\acute{M}$-antenna channel with coherence time $Q\acute{T}$ (shown on right)

We then invoke a particular form of the Converse to the Coding Theorem [17] to show that our lower bound on autocapacity is also an upper bound. Finally, we obtain an explicit simple formula for autocapacity. We examine the autocapacity in the two cases 1) channel known at the receiver, and 2) channel unknown at the receiver. As always, the propagation matrix is unknown to the transmitter.

3.1 Lower bound on autocapacity

Let the number of receiver antennas be N, but let the coherence interval be $\acute{T}$, and the number of transmitter antennas be $\acute{M}$. Consider the conventional encoding of $R \cdot Q \cdot \acute{T}$ bits over Q independent $\acute{T}$-symbol coherence intervals as illustrated on the left in Figure 2. Then there is a coding scheme such that for any $R < C(\rho, \acute{T}, \acute{M}, N)$, the block probability of error decays exponentially with increasing Q, where $C(\cdot)$ is the appropriate Shannon capacity, whose value depends on whether or not H is known to the receiver,

Now suppose that exactly the same coding scheme is utilized within a single extended coherence interval of duration $T = Q\acute{T}$, with a total of $M = Q\acute{M}$ transmit antennas, using the block-diagonal signal structure shown on the right in Figure 2, where the $\acute{T} \times \acute{M}$ matrix that was transmitted during the qth coherence interval is now fed to antennas $m = (q-1) \cdot \acute{M} + 1, \ldots, q \cdot \acute{M}$ during the time interval $t = (q-1) \cdot \acute{T} + 1, \ldots, q \cdot \acute{T}$. In the former situation, the propagation coefficients (and additive noise) are independent from one coherence interval to another. In the latter, the propagation coefficients (and additive noise) are independent from one consecutive group of $\acute{M}$ transmit antennas to another. Hence the block probability of error is exactly equal for the two cases, so the desired exponential decay of the block probability of error occurs within the single coherence interval of duration $T = Q\acute{T}$ with $M = Q\acute{M}$ transmitter antennas. This establishes the existence of the autocoding effect.

This same argument also provides a lower bound on autocapacity. The block diagonal signal structure that is shown on the right in Figure 2 is not the most general possible signal structure; some nondiagonal structure could conceivably be superior. Thus,

$$C(\rho, \acute{T}, \acute{M}, N) \leq C_a(\rho, \beta, N), \quad \beta = \acute{T}/\acute{M}, \tag{13}$$

which, for a given β, holds for all integer values of $(\acute{T}, \acute{M})$ such that $\acute{T} = \beta\acute{M}$. The autocapacity must be greater than, or equal to, the maximum of the left-hand side of (13),

$$\sup_{T,M:\, T/M=\beta} C(\rho, T, M, N) \leq C_a(\rho, \beta, N). \tag{14}$$

3.2 Upper bound on autocapacity

The possibility of the autocapacity C_a exceeding the Shannon capacity, as suggested by the inequality (14), is intuitively disturbing. We now use an inequality that plays a role in the Converse to the Coding Theorem to show that the lower bound on C_a in (14) is actually also an upper bound.

Let us assume that the lower bound (14) is not tight, and we transmit at a rate R that is greater than this lower bound, but less than the autocapacity. We therefore encode $R \cdot T$ bits of information onto a $T \times M$ signal, where $T/M = \beta$, and

$$\sup_{\acute{T},\acute{M}:\, \acute{T}/\acute{M}=\beta} C(\rho, \acute{T}, \acute{M}, N) < R < C_a(\rho, \beta, N). \tag{15}$$

Since the transmission rate is less than the autocapacity, the block probability of error goes to zero as T goes to infinity. We now show that this conclusion is contradicted by an application of the Fano inequality combined with the Data Processing Theorem, equation (4.3.20) of [17], which relates the average bit probability of error to the mutual information between the transmitted signal S and the received signal X. We apply this equation to the case of independent binary message bits $b = \{b_1, \cdots, b_{R \cdot T}\}$ that are equally likely to be zero or one, and for a single channel use (i.e., the transmission of one $T \times M$ signal). The result is

$$-\bar{P}_e \log \bar{P}_e - (1 - \bar{P}_e) \log(1 - \bar{P}_e) \geq \frac{1}{R \cdot T} \mathcal{H}(b) - \frac{T \cdot C(\rho, T, M, N)}{R \cdot T}, \tag{16}$$

where $\bar{P}_e$ is the average bit probability of error,

$$\bar{P}_e = \frac{1}{R \cdot T} \sum_{j=1}^{R \cdot T} P_{ej}, \tag{17}$$

$\mathcal{H}(b) = RT$ is the entropy of the message bits, and $C(\rho, T, M, N)$ is the Shannon capacity of the channel (recall that we normalize $C(\cdot)$ so that $T \cdot C(\cdot)$ is equal to the maximum mutual information between the transmitted signal S and the received signal X). We manipulate (16) to obtain

$$\begin{aligned} -\bar{P}_e \log \bar{P}_e - (1 - \bar{P}_e) \log(1 - \bar{P}_e) &\geq \frac{R - C(\rho, T, M, N)}{R} \\ &\geq \frac{R - \sup_{\acute{T}, \acute{M}:\, \acute{T}/\acute{M} = \beta} C(\rho, \acute{T}, \acute{M}, N)}{R}. \end{aligned} \tag{18}$$

The right-hand side of (18) is positive because the rate is chosen according to (15), and it is independent of T. This implies that the average bit probability of error $\bar{P}_e$ is lower-bounded by a positive constant as $T \to \infty$. But this contradicts the fact that the block probability of error goes to zero. Consequently the assumed rate (15) must be impossible to attain, and

$$C_a(\rho, \beta, N) \leq \sup_{T, M:\, T/M = \beta} C(\rho, T, M, N). \tag{19}$$

3.3 Explicit formula for autocapacity

The combination of (14) and (19) implies that

$$C_a(\rho, \beta, N) = \sup_{T, M:\, T/M = \beta} C(\rho, T, M, N). \tag{20}$$

We now proceed to evaluate this expression when H is known and unknown.

3.3.1 H known

The results cited in Section 2.2.1 immediately enable us to obtain a closed-form expression for the autocapacity (20) when the propagation matrix is known to the receiver. The Shannon capacity is independent of T and increases monotonically with M, which gives

$$C_a(\rho, \beta, N) = N \log(1 + \rho) \qquad (H \text{ known}). \tag{21}$$

3.3.2 H unknown

When H is unknown the Shannon capacity can be computed only for some special cases. In general, one cannot even compute the mutual information for specific densities $p(S)$. It is therefore perhaps somewhat surprising that the autocapacity turns out also to be given by (21).

Starting with (13), we apply the properties of the Shannon capacity (11), (12) to obtain

$$\begin{aligned} C_a(\rho, \beta, N) &\geq C(\rho, L \cdot T, L \cdot M, N) \\ &\geq C(\rho, L \cdot T, M, N), \quad T/M = \beta, \ L = 1, 2, \ldots. \end{aligned} \tag{22}$$

Equation (22) is true for all $L > 0$, and, in particular,

$$C_a(\rho, \beta, N) \geq \sup_{L > 0} C(\rho, L \cdot T, M, N).$$

But this supremum is simply the Shannon capacity for H known to the receiver, since as L grows the Shannon capacity for H unknown approaches the Shannon capacity for H known (see Section 2.2.2 and [5]). Using (7), we thus obtain

$$C_a(\rho, \beta, N) \geq \mathrm{E}\left\{\log\det\left[I_M + \frac{\rho}{M} H H^\dagger\right]\right\}. \tag{23}$$

This formula is true for all $M > 0$, and, in particular, for the supremum over all $M > 0$. This supremum can be evaluated by letting M grow without bound and using (8). We thus conclude that

$$C_a(\rho, \beta, N) \geq N \log(1 + \rho). \tag{24}$$

We see that the right-hand expression is equal to the autocapacity when H is known to the receiver which must be greater than or equal to the autocapacity when H is unknown. Consequently the autocapacity does not depend on the receiver's knowledge or ignorance of the propagation matrix, is independent of β, and it is given by the simple formula

$$C_a(\rho, \beta, N) = N \log(1 + \rho) \qquad (H \text{ known or unknown}). \tag{25}$$

4 Probability of Error and Duration of Coherence Interval

The autocapacity formula (25) suggests the possibility of reliable transmission at a high bit rate, but it does not indicate the number of transmitter antennas or duration of the coherence interval that is needed to achieve low error probability. To examine the error probability in this regime, we reinterpret our autocoding construction of Figure 2 with the help of the familiar random coding exponent and cutoff rate applied to a single coherence interval. We are unable to perform completely the optimization that is needed to compute the random coding exponent, so we resort to lower-bounding the cutoff rate.

4.1 Random coding exponent

Suppose that we have L independent coherence intervals, each of duration T, with M transmitter antennas, N receiver antennas, and an expected SNR ρ. A total of $R \cdot L \cdot T$ bits are transmitted using a signal that is randomly generated with probability density $p(S)$ satisfying power constraint (1), independently across coherence intervals. The block probability of error is upper-bounded [17, 8] as

$$P_e \le \exp\left\{-LT(\ln 2)\left[E_0\left(\mu, p(S)\right) - \mu R\right]\right\}, \tag{26}$$

for arbitrary $0 \le \mu \le 1$, and for arbitrary $p(S)$, where

$$E_0\left(\mu, p(S)\right) = -\frac{1}{T} \log\left[\int_X \left(\mathrm{E}_S\left\{p^{\frac{1}{1+\mu}}(X \mid S)\right\}\right)^{1+\mu} dX\right].^3 \tag{27}$$

For any rate R that is less than the Shannon capacity $C(\rho, T, M, N)$ it is always possible to choose μ and $p(S)$ so that the exponent $E_0(\mu, p(S)) - \mu R$ in (26) is strictly positive. This guarantees that the block probability of error decays exponentially fast as $L \to \infty$.

[3] This expression is for unknown H. When H is known, the received signal also includes H, $p(X, H|S)$ replaces $p(X|S)$, and the integral is over both X and H.

However, our interest is the autocoding regime where $L = 1$ and the coherence time T is large. We wish to show that, *as T goes to infinity* in some manner, $E_0(\mu, p(S)) - \mu R$ remains positive, thus ensuring an exponential decay in T of the block probability of error. Suppose that for some R and (T, M), we have obtained a μ^* and a $p^*(S)$ that yield a positive exponent, and a bound P_e^*. Now increase T and M by an integer factor Q, $(T, M) \to (Q{\cdot}T, Q{\cdot}M)$. We give the expanded $Q{\cdot}T \times Q{\cdot}M$ signal matrix a block-diagonal structure similar to that shown on the right side of Figure 2, so that at times $t = (q-1) \cdot T + 1, \cdots, q \cdot T$ a signal is fed only to antennas $m = (q-1) \cdot M + 1, \cdots, q \cdot M$. We furthermore choose the $T \times M$ signal blocks (of which there are Q) to be statistically independent and distributed as $p^*(S)$, suitably normalized to comply with the power constraint. As a consequence, the expanded $(Q \cdot T) \times N$ received signal comprises $T \times N$ blocks that are statistically independent and that have the same probability density as the original $T \times N$ received signal. Therefore as $(T, M) \to (Q \cdot T, Q \cdot M)$, the log term in (27) increases by the factor Q, so $E_0(\mu^*, p^*(S)) - \mu^* R$ does not change, but the exponent in (26) is Q times larger. Therefore for any rate $R < C(\rho, T, M, N)$ we have a block probability of error for the $Q \cdot T$-symbol coherence interval that decays exponentially with Q. We see that if (T, M) grow simultaneously the error probability can be made as small as desired.

4.2 Cutoff rate

Computing the random coding exponent is even more difficult than computing capacity, and when H is unknown the capacity has been computed in only some special cases [8]. Hence, we resort to computing the so-called cutoff rate R_0 associated with setting $\mu = 1$ in (26), which for $L = 1$ and H unknown gives the error bound

$$\begin{aligned} P_e &\leq \exp\left\{-T(\ln 2)\left(-\frac{1}{T}\log\left[\int_X \left(\mathrm{E}_S\left\{p^{\frac{1}{2}}(X \mid S)\right\}\right)^2 dX\right] - R\right)\right\} \\ &= \exp\left\{-T(\ln 2)\left(-\frac{1}{T}\log\left[\mathrm{E}_{S_1,S_2}\left\{\int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)}\right\}\right] - R\right)\right\}, \end{aligned} \tag{28}$$

where S_1 and S_2 are independent, and identically distributed according to any $p(S)$ that satisfies (1). (An analogous expression is easily obtained when H is known.) The cutoff rate is defined as [16]

$$R_0(T, \beta, N) = \sup_{p(S)} -\frac{1}{T}\log\left[\mathrm{E}_{S_1,S_2}\left\{\int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)}\right\}\right] \geq 0, \tag{29}$$

where we have made the dependence of the cutoff rate on T, β, and N explicit. We are interested in the behavior of $R_0(T, \beta, N)$ in the regime where β and N are fixed and T grows arbitrarily large. In general, we are unable to perform the optimization with respect to $p(S)$, so we instead find lower bounds on the cutoff rate by using an assumed density for S.

The upper bound (28) has a simple interpretation in terms of a Chernoff bound combined with a union bound. In particular the quantity $\ln \int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)}$ is the so-called Bhattacharyya distance between the two conditional densities, and the Chernoff bound (with the "tilting" parameter set equal to 1/2) on the probability of error for deciding between the two signals is [18]

$$P_e \{S_1 \text{ vs. } S_2\} \leq \frac{1}{2} \int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)}. \tag{30}$$

The expectation of the above bound with respect to S_1 and S_2 gives the pairwise probability of error between any two members of the randomly chosen constellation, which when multiplied by the number of signals in the constellation 2^{RT} yields the bound (28) (apart from an extra factor of 1/2 that carries over from (30)).[4]

4.3 Cutoff rate lower bounds: H known

When H is known to the receiver, we use the conditional density (3) and a modified form of (29) to obtain

$$\begin{aligned}
R_0(T, \beta, N) &= \sup_{p(S)} -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \int dH \int dX \sqrt{p(X, H \mid S_1) \cdot p(X, H \mid S_2)} \right\} \right] \\
&= \sup_{p(S)} -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \int dH \frac{1}{\pi^{MN}} \exp \left(-\mathrm{tr} \left\{ HH^\dagger \left[I_M + \frac{\rho}{4M} (S_2 - S_1)^\dagger (S_2 - S_1) \right] \right\} \right) \right\} \right] \\
&= \sup_{p(S)} -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \frac{1}{\det^N \left[I_M + \frac{\rho}{4M} (S_2 - S_1)^\dagger (S_2 - S_1) \right]} \right\} \right].
\end{aligned}$$

Substituting a specific $p(S)$ into the right hand side of this expression yields a lower bound on the cutoff rate

$$R_0(T, \beta, N) \geq -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \frac{1}{\det^N \left[I_M + \frac{\rho}{4M} (S_2 - S_1)^\dagger (S_2 - S_1) \right]} \right\} \right]. \tag{31}$$

We now examine the lower bounds obtained with Gaussian and Unitary Space-Time signals.

[4]This simple derivation of the formula (28) does not permit us to conclude that the probability of error decays exponentially with increasing T, for which the more elaborate machinery of the random coding exponent is needed. Later it will be expedient to replace the Bhattacharyya bound (30) by the exact pairwise probability of error.

4.3.1 Gaussian signals

When the elements of S_1 and S_2 are independent $\mathcal{CN}(0,1)$ then $(S_2 - S_1) = \sqrt{2}S$, where the elements of S are independent $\mathcal{CN}(0,1)$. The cutoff rate bound (31) becomes

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_S\left\{\frac{1}{\det^N\left[I_M + \frac{\rho}{2M}S^\dagger S\right]}\right\}\right]. \tag{32}$$

This expression can be rewritten as a function of the eigenvalues of the $M \times M$ matrix, $S^\dagger S$. When $\beta \geq 1$, the matrix $\frac{1}{M}S^\dagger S$ has, with probability one, M positive eigenvalues that we denote by $\lambda_1, \ldots, \lambda_M$. Equation (32) becomes

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_S\left\{\frac{1}{\prod_{m=1}^{M}\left[1 + \frac{\rho}{2}\lambda_m\right]^N}\right\}\right]. \tag{33}$$

Likewise, when $\beta < 1$, the matrix $\frac{1}{T}SS^\dagger$ has (with probability one) $T < M$ positive eigenvalues $\lambda_1, \ldots, \lambda_T$, so that (32) becomes

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_S\left\{\frac{1}{\prod_{m=1}^{T}\left[1 + \frac{\rho\beta}{2}\lambda_m\right]^N}\right\}\right]. \tag{34}$$

(To avoid unnecessary clutter, we omit the cumbersome phrase "with probability one" in the remainder of the eigenvalue discussions in this paper.)

The asymptotic values of the above cutoff rates are computed in Appendix A where it is shown that

$$\lim_{T\to\infty} R_0(T,\beta,N) \;\geq\; \begin{cases} \frac{N}{\beta}F\left(\beta, \frac{\rho}{2}\right) & \beta \geq 1 \\ NF\left(\frac{1}{\beta}, \frac{\rho\beta}{2}\right) & 0 \leq \beta \leq 1 \end{cases}, \tag{35}$$

where the function $F(\cdot,\cdot)$ is defined in (A.57) as

$$\begin{aligned} F(\beta,\rho) \;&=\; \log\left(1 + \rho(\sqrt{\beta}+1)^2\right) + (\beta+1)\log\left(\frac{1+\sqrt{1-a}}{2}\right) \\ &\quad - (\log e)\sqrt{\beta}\frac{1-\sqrt{1-a}}{1+\sqrt{1-a}} + (\beta-1)\log\left(\frac{1+\gamma}{\gamma+\sqrt{1-a}}\right), \end{aligned} \tag{36}$$

and $a = \frac{4\rho\sqrt{\beta}}{1+\rho(\sqrt{\beta}+1)^2}$ and $\gamma = \frac{\sqrt{\beta}-1}{\sqrt{\beta}+1}$.

4.3.2 Unitary Space-Time signals

For Unitary Space-Time signals, $S_\ell = \sqrt{T}\Phi_\ell$, where the Φ_ℓ are independent, isotropically random unitary (see Section 2.2.2),

$$\begin{aligned}(S_2 - S_1)^\dagger (S_2 - S_1) &= T(\Phi_2 - \Phi_1)^\dagger(\Phi_2 - \Phi_1) \\ &= T\left[2I_M - \left(\Phi_2^\dagger\Phi_1\right) - \left(\Phi_2^\dagger\Phi_1\right)^\dagger\right]. \end{aligned} \tag{37}$$

The product of two independent $T \times M$ isotropically random unitary matrices can be further simplified as follows,

$$\begin{aligned}\Phi_2^\dagger\Phi_1 &= \Phi_2^\dagger \begin{bmatrix} \Phi_2 & \Phi_{2\perp} \end{bmatrix} \begin{bmatrix} \Phi_2^\dagger \\ \Phi_{2\perp}^\dagger \end{bmatrix} \Phi_1 \\ &= \begin{bmatrix} I_M & 0 \end{bmatrix} \Phi, \end{aligned} \tag{38}$$

where $\Phi_{2\perp}$ is any $T \times (T - M)$ matrix such that

$$\Phi_{2\perp}^\dagger\Phi_{2\perp} = I_{T-M}, \qquad \Phi_2^\dagger\Phi_{2\perp} = 0, \tag{39}$$

and

$$\Phi = \begin{bmatrix} \Phi_2^\dagger \\ \Phi_{2\perp}^\dagger \end{bmatrix} \Phi_1. \tag{40}$$

The matrices Φ_2 and $\Phi_{2\perp}$ together form a $T \times T$ unitary matrix. Conditioned on Φ_2, the $T \times M$ matrix Φ is a rotated version of the isotropically random Φ_1. Using the fact that the probability density of an isotropically random unitary matrix is unchanged when this matrix is multiplied by a deterministic unitary matrix, we conclude that Φ is isotropically random and statistically independent of Φ_2. Combining this with (31) gives

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_\Phi\left\{\frac{1}{\det^N\left[\left(1+\frac{\rho T}{2M}\right)I_M - \frac{\rho T}{4M}\left(\Theta+\Theta^\dagger\right)\right]}\right\}\right], \tag{41}$$

where Θ comprises the first M rows of the $T \times M$ isotropically random unitary matrix Φ,

$$\Theta = \begin{bmatrix} I_M & 0 \end{bmatrix} \Phi. \tag{42}$$

Let the M eigenvalues of the $M \times M$ matrix $\frac{1}{2}(\Theta + \Theta^{\dagger})$ be denoted by $\lambda_1, \ldots, \lambda_M$, where the λ_m are real-valued with $-1 \leq \lambda_m \leq 1$. We may write (41) as

$$R_0(T, \beta, N) \geq -\frac{1}{T} \log \mathrm{E}_{\Phi} \left\{ \frac{1}{\prod_{m=1}^{M} \left[1 + \frac{\rho\beta}{2} - \frac{\rho\beta}{2}\lambda_m\right]^N} \right\}. \tag{43}$$

The asymptotic value of the above expression is computed in Appendix A as

$$\lim_{T\to\infty} R_0(T, \beta, N) \ \geq \ \frac{N(2\beta - 1)}{\beta^2 \pi} \int_0^{\pi} \log \left(1 + \frac{\rho\beta}{2} - \frac{\rho\sqrt{2\beta - 1}}{2} \cos\theta\right) \frac{\sin^2\theta}{1 - \frac{2\beta-1}{\beta^2}\cos^2\theta} d\theta. \tag{44}$$

4.4 Cutoff rate lower bounds: H unknown

When H is unknown, we substitute the conditional density (4) into (29) to obtain

$$R_0(T, \beta, N) \geq -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \left[\frac{\det^{1/2}\Lambda_1 \det^{1/2}\Lambda_2}{\det\left(\frac{\Lambda_1 + \Lambda_2}{2}\right)} \right]^N \right\} \right], \tag{45}$$

where

$$\Lambda_\ell = I_T + (\rho/M) S_\ell S_\ell^{\dagger}, \qquad \ell = 1, 2. \tag{46}$$

It follows that

$$\det\left(\frac{\Lambda_1 + \Lambda_2}{2}\right) = \det\left(I_T + \frac{\rho}{2M} S_{12} S_{12}^{\dagger}\right), \quad S_{12} \triangleq \left[\begin{array}{cc} S_1 & S_2 \end{array}\right]. \tag{47}$$

4.4.1 Gaussian signals

When the elements of S_1 and S_2 are independent $\mathcal{CN}(0, 1)$, no simplification of (45) is easily found. However, we can still express the cutoff rate bound in terms of various eigenvalues.

- $\beta \geq 2$: In this case, the matrices $\frac{1}{M} S_\ell^{\dagger} S_\ell$, $\ell = 1, 2$, and $\frac{1}{2M} S_{12}^{\dagger} S_{12}$ have positive eigenvalues $\lambda_{\ell,1}, \ldots, \lambda_{\ell,M}$, $\lambda_{12,1}, \ldots, \lambda_{12,2M}$, respectively, so that

$$R_0(T, \beta, N) \geq -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \frac{\prod_{m=1}^{M} \left[(1 + \rho\lambda_{1,m})(1 + \rho\lambda_{2,m})\right]^{N/2}}{\prod_{m=1}^{2M} \left[1 + \rho\lambda_{12,m}\right]^N} \right\} \right]. \tag{48}$$

- $1 \leq \beta \leq 2$: Here $\frac{1}{M} S_\ell^{\dagger} S_\ell$, $\ell = 1, 2$, still have M positive eigenvalues, whereas $\frac{1}{T} S_{12} S_{12}^{\dagger}$, has $T < 2M$

positive eigenvalues, so that

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_{S_1,S_2}\left\{\frac{\prod_{m=1}^{M}\left[(1+\rho\lambda_{1,m})(1+\rho\lambda_{2,m})\right]^{N/2}}{\prod_{m=1}^{T}\left[1+\frac{\rho\beta}{2}\lambda_{12,m}\right]^{N}}\right\}\right]. \tag{49}$$

- $0 \leq \beta \leq 1$: Here $\frac{1}{T}S_\ell S_\ell^\dagger$, $\ell = 1,2$, has $T < M$ positive eigenvalues, whereas $\frac{1}{T}S_{12}S_{12}^\dagger$ still has $T < 2M$ positive eigenvalues. Thus,

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log\left[\mathrm{E}_{S_1,S_2}\left\{\frac{\prod_{m=1}^{T}\left[(1+\rho\beta\lambda_{1,m})(1+\rho\beta\lambda_{2,m})\right]^{N/2}}{\prod_{m=1}^{T}\left[1+\frac{\rho\beta}{2}\lambda_{12,m}\right]^{N}}\right\}\right]. \tag{50}$$

Appendix A obtains the asymptotic values of the above expressions

$$\lim_{T\to\infty} R_0(T,\beta,N) \;\geq\; \left\{\begin{array}{ll} \frac{2N}{\beta}F\left(\frac{\beta}{2},\rho\right) - \frac{N}{\beta}F\left(\beta,\rho\right) & \beta \geq 2 \\ NF\left(\frac{2}{\beta},\frac{\rho\beta}{2}\right) - \frac{N}{\beta}F\left(\beta,\rho\right) & 1 \leq \beta \leq 2 \\ NF\left(\frac{2}{\beta},\frac{\rho\beta}{2}\right) - NF\left(\frac{1}{\beta},\rho\beta\right) & 0 \leq \beta \leq 1 \end{array}\right. , \tag{51}$$

where $F(\cdot,\cdot)$ is given in (36).

4.4.2 Unitary Space-Time signals

For Unitary Space-Time signals, we have $S_\ell = \sqrt{T}\Phi_\ell$, and

$$\det \Lambda_\ell = (1+\rho\beta)^M . \tag{52}$$

Equation (47) becomes

$$\begin{aligned} \det\left(\frac{\Lambda_1+\Lambda_2}{2}\right) &= \det\begin{bmatrix} \left(1+\frac{\rho\beta}{2}\right)I_M & \frac{\rho T}{2M}\Phi_1^\dagger\Phi_2 \\ \frac{\rho\beta}{2}\Phi_2^\dagger\Phi_1 & \left(1+\frac{\rho\beta}{2}\right)I_M \end{bmatrix} \\ &= \det\left[\left(1+\frac{\rho\beta}{2}\right)^2 I_M - \left(\frac{\rho\beta}{2}\right)^2\Theta\Theta^\dagger\right], \end{aligned} \tag{53}$$

where $\Theta = \Phi_2^\dagger\Phi_1$ has the distribution given in (42). Let the M eigenvalues of the $M\times M$ matrix $\Theta\Theta^\dagger$ be denoted by $\lambda_1,\ldots,\lambda_M$, where the λ_m are real-valued with $0 \leq \lambda_m \leq 1$. The combination of (45) with (52) and (53) yields

$$R_0(T,\beta,N) \geq -\frac{1}{T}\log \mathrm{E}_\Phi \left\{ \frac{1}{\prod_{m=1}^{M}\left[1+\frac{(\rho\beta)^2(1-\lambda_m)}{4(1+\rho\beta)}\right]^N} \right\}. \tag{54}$$

The asymptotic cutoff rate is found in Appendix A to be

$$\lim_{T\to\infty} R_0(T,\beta,N) \;\geq$$

$$\begin{cases} \frac{N}{\beta}\left[\log(1+\alpha)+\beta\log\left(\frac{1+\sqrt{1-\frac{b\alpha}{1+\alpha}}}{2}\right)+(\beta-2)\log\left(\frac{1+\sqrt{1-b}}{\sqrt{1-b}+\sqrt{1-\frac{b\alpha}{1+\alpha}}}\right)\right] & \beta\geq 2 \\ \frac{N}{\beta}\left[(\beta-1)\log(1+\alpha)+\beta\log\left(\frac{1+\sqrt{1-\frac{b\alpha}{1+\alpha}}}{2}\right)+(2-\beta)\log\left(\frac{1+\sqrt{1-b}}{\sqrt{1-b}+\sqrt{1-\frac{b\alpha}{1+\alpha}}}\right)\right] & 1\leq\beta\leq 2 \end{cases}, \tag{55}$$

where $\alpha = \frac{(\rho\beta)^2}{4(1+\rho\beta)}$ and $b = \frac{4(\beta-1)}{\beta^2}$.

5 Computations of Cutoff Rates and Error Probabilities

This section presents numerical evidence that the autocoding effect manifests itself for relatively small values of T and M.

5.1 Convergence of R_0 bounds to their asymptotic values

In Section 4, we present analytical formulas for the asymptotic $(T,M)\to\infty$ cutoff rate bounds. These bounds can be conveniently computed for finite T,M through Monte-Carlo integration of equations (31) and (45) for Gaussian and USTM (isotropically) distributed S. (These can also be computed in closed-form; see Appendix B.) The rate of convergence of these bounds to their asymptotic values is of practical interest. Figure 3 shows the cutoff rate bounds as functions of M (solid curves), for $\beta=2$, $\rho=18$ dB, and $N=1$, for the four cases (S USTM or Gaussian, H unknown or known). The dashed lines indicate their limiting values as $M\to\infty$, given by equations (35), (44), (51), and (55). Even for small values of M the R_0 bounds are close to their asymptotic values.

5.2 Asymptotic cutoff rate bounds versus β

Figure 4 shows the asymptotic cutoff rate bounds versus β, for $\rho=18$ dB and $N=1$ for the four cases. Of necessity the USTM signals require that $\beta\geq 1$.

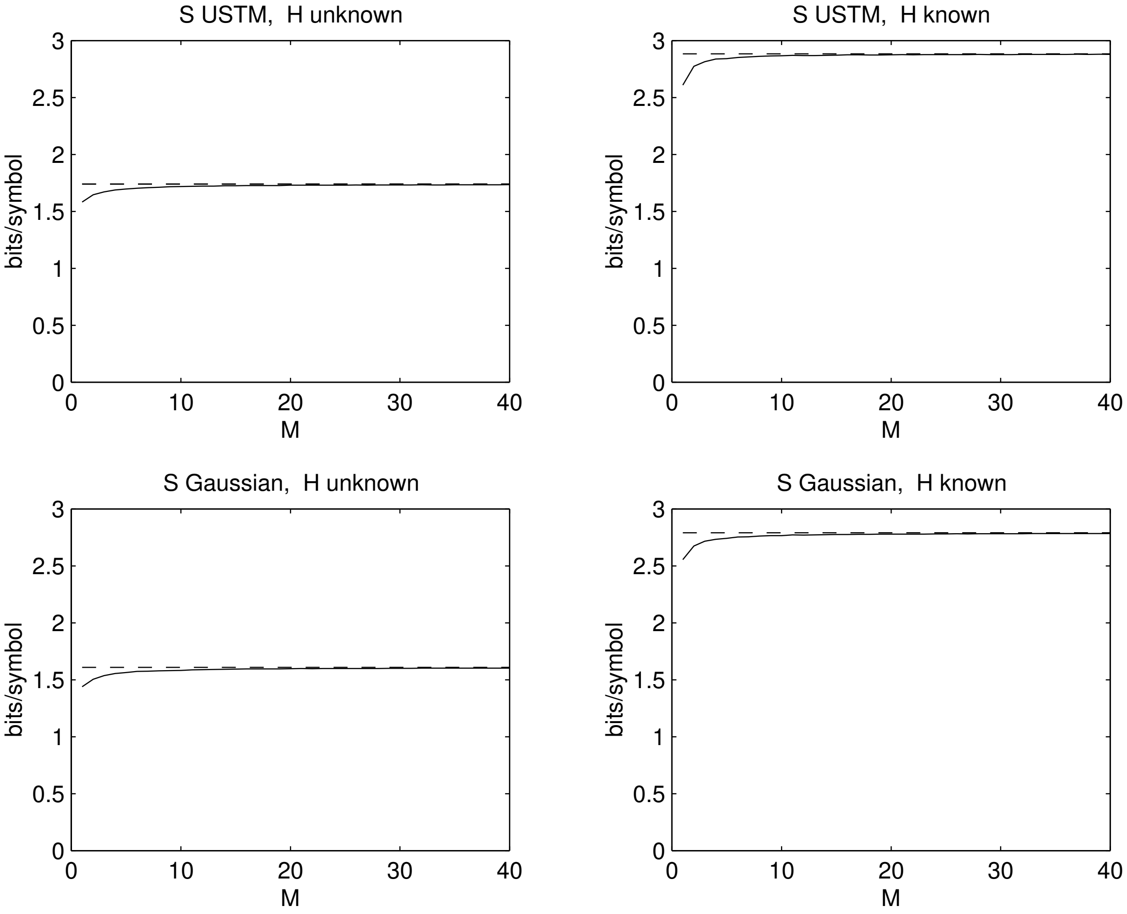

Figure 3: Cutoff rate bounds (bits/symbol) obtained by Monte-Carlo integration, as a function of M for $\rho = 18$ dB, $\beta = 2$, and $N = 1$. The dashed lines are the asymptotic cutoff rate bounds ($M \to \infty$).

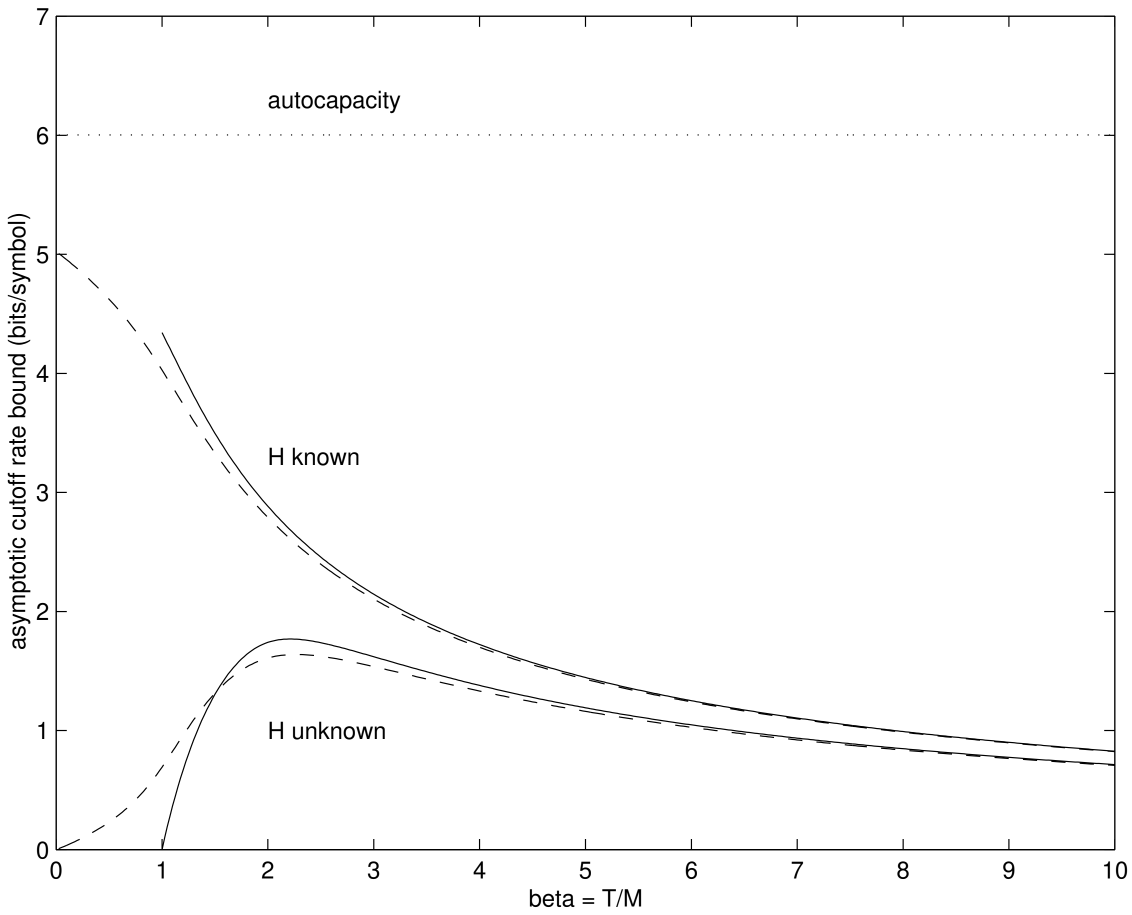

Figure 4: Asymptotic cutoff rate bounds (bits/symbol) versus β, for $\rho = 18$ dB, and $N = 1$. The solid curves are S USTM; the dashed curves are S Gaussian; the dotted line is the autocapacity

There are a number of features of interest. The cutoff rate bounds can be a significant fraction of the autocapacity. The bound for S USTM and H unknown is zero for $\beta = 1$ because all $\beta = 1$ USTM signals are indistinguishable when H is unknown [6]. The bounds for H known are greater than the corresponding bounds for H unknown; however for large β (where a small fraction of the coherence interval could be used for training [5]) this difference vanishes. For the cases where H is unknown, the R_0 bounds peak in the neighborhood of $\beta = 2$. The USTM signals yield larger R_0 bounds than the Gaussian signals near $\beta = 2$. The region where H is known and $\beta < 1$ is probably impractical because at least M training symbols are generally needed ($T \geq M$ or $\beta \geq 1$), even for large ρ, to estimate the propagation coefficients [19].

Similar plots for other SNR's (not shown) indicate the same qualitative behavior. As the SNR increases, the bounds on R_0, when H is unknown, peak more sharply near $\beta = 2$ (or $T = 2M$), which emerges as a desirable operating point.

5.3 Dependence on number of receiver antennas

The formulas for the asymptotic cutoff rate bound (35), (44), (51), (55) have a linear dependence on the number of receiver antennas N. However for finite M, other considerations suggest that this linear relation cannot hold for arbitrary values of N. In fact, when $N \gg M$, the asymptotic Shannon capacity (7) for H known to the receiver is [14]

$$C = M \log\left(1 + \frac{\rho N}{M}\right),$$

with only logarithmic dependence on N.

Figure 5 illustrates the behavior of the cutoff rate bound as a function of N when H is unknown and the signals are USTM. We generated the curves by evaluating the expression (54) in closed-form, rather than through Monte-Carlo integration, with SNR 18 dB and $\beta = 2$. See Appendix B for details. The dashed line is the asymptotic $M \to \infty$ cutoff rate bound. We see that for $N < M$, the cutoff rate bounds increase approximately linearly with N, but for $N \gg M$, the increase is only logarithmic; the excess number of receiver antennas serves only to increase the effective received SNR.

5.4 Transmission rate versus probability of error for H unknown and Unitary Space-Time signals

The formula (28) can translate the cutoff rate bound into an upper bound on the block probability of error as a function of the transmission rate. We obtain tighter bounds by combining the union bound with the exact pairwise probability of error as follows,

$$P_e \leq 2^{RT} \mathrm{E}_{S_1,S_2}\left\{P_e\left\{S_1 \text{ vs. } S_2\right\}\right\}, \tag{56}$$

where the pairwise probability of error is (see Appendix B)

$$\mathrm{E}_{S_1,S_2}\left\{P_e\left\{S_1 \text{ vs. } S_2\right\}\right\} = \mathrm{E}_\lambda\left\{\frac{1}{\pi}\int_0^{\pi/2} d\theta \prod_{m=1}^{M}\left[\frac{\cos^2\theta}{\cos^2\theta + \frac{(\rho T/M)^2(1-\lambda_m)}{4(1+\rho T/M)}}\right]^N\right\}, \tag{57}$$

where $\lambda_1, \ldots, \lambda_M$ are the eigenvalues of $\Theta\Theta^\dagger$ (42). This expression accounts for all sources of randomness: the choice of the signals that comprise the random codebook, the propagation matrix H, and the receiver

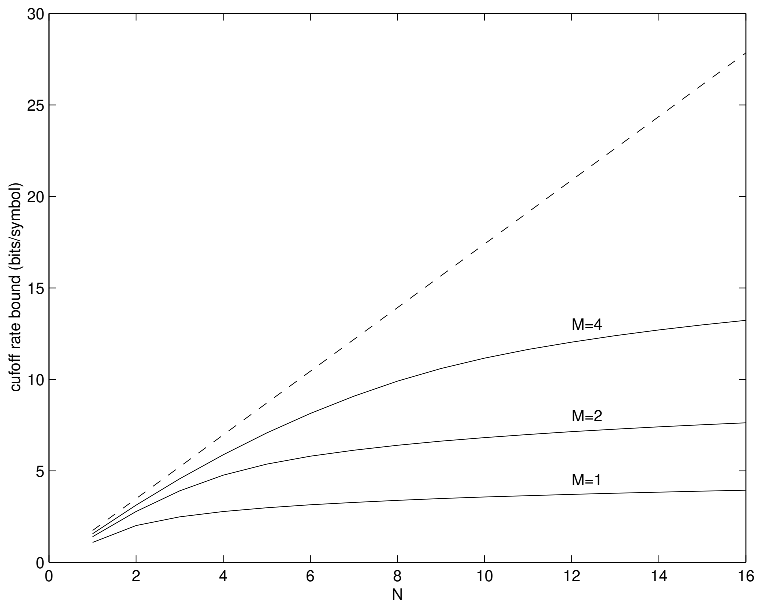

Figure 5: Cutoff rate bound versus the number of receiver antennas for $M = 1, 2, 4$, where H is unknown and the signals are Unitary Space-Time. The SNR is 18 dB and $\beta = 2$. The dashed line is the asymptotic $M \to \infty$ cutoff rate bound.

noise matrix W. We take the expectation with respect to the eigenvalues analytically (see Appendix B) to obtain the pairwise probability of error as a function of T and M as shown in Figure 6. There are $N = 4$ receiver antennas and the SNR is 18 dB.

The resulting upper bounds on the block probability of error as a function of the transmission rate are shown in Figure 7, utilizing the number of transmitter antennas for each value of T that minimizes the pairwise probability of error, which is approximately equal to $T/2$. It is worth noting that, in this regime the receiver cannot obtain a very good estimate for the propagation matrix. As we can see from the figure, one can transmit information during a single short coherence interval at rates that are an appreciable fraction of autocapacity with extremely high reliability. For example, within the coherence interval $T = 16$, for $M = 7$ transmitter antennas a total of 80 bits (corresponding to rate $R = 5$) can be transmitted with a block probability of error less than 10^{-9}. Space-time autocoding is thus a very real effect that manifests itself even for small values of T and M.

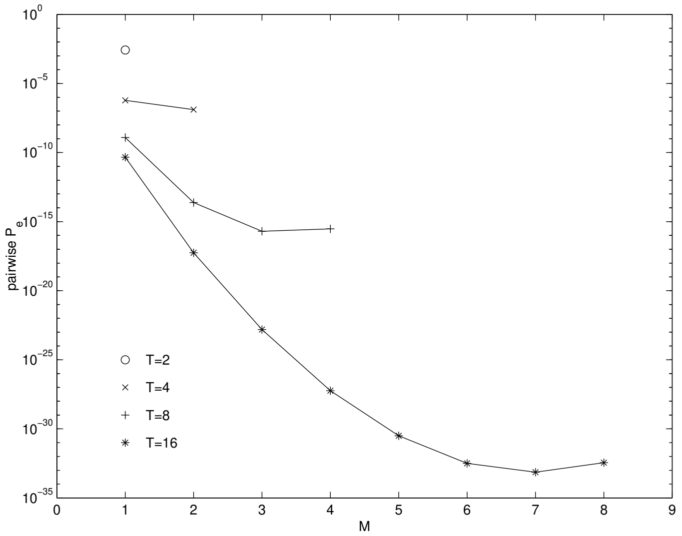

Figure 6: Pairwise probability of error for random Unitary Space-Time signals with H unknown for $N = 4$ receiver antennas and $\rho = 18$ dB.

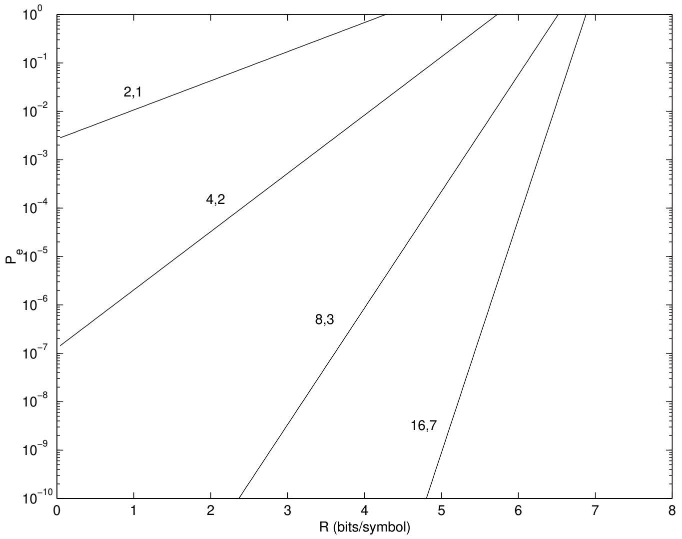

Figure 7: Upper bound on block probability of error versus transmission rate (bits/symbol) for random constellation of Unitary Space-Time signals, for $N = 4$, $\rho = 18$ dB, and $(T, M) = (2, 1), (4, 2), (8, 3), (16, 7)$. Autocapacity is equal to 24.01 bits/symbol. The propagation matrix is unknown.

6 Conclusions

In the past, it was believed that arbitrarily reliable transmission at positive rates in fading channels could only be achieved through redundant signaling over many independent coherence intervals. We have described an effect, called space-time autocoding, where arbitrary reliability can be obtained within a *single* coherence interval. In space-time autocoding, spatial diversity replaces temporal diversity while simultaneously yielding the high throughput promised by multiple antennas. The resulting autocapacity does not depend on whether or not the propagation coefficients are known.

Unitary Space-Time signals are promising candidates for autocoding, and for typical SNR's the number of transmitter antennas should be about half of the symbol duration of the coherence interval. For reasonably short coherence intervals, rates of up to 25% of autocapacity are sustained with extremely low probabilities of error.

Space-time autocoding eliminates the complexity of multiple-antenna channel coding over several coherence intervals, but it still leaves us with the problem of how to create and decode the large constellations that are typically needed to realize the autocoding effect at high data rates.

A Appendix: Asymptotics and Eigenvalue Distributions

In Section 4 we obtain lower bounds on the cutoff rate (equations (33)–(34), (43), (48)–(50), and (54)) as expectations over the eigenvalues of Gaussian and isotropically distributed unitary random matrices. Computing these expectations requires multidimensional integrations involving the joint distribution of these eigenvalues. Although we do such a multidimensional integration for USTM signals when H is unknown in Appendix B, the resulting expressions generally are complicated and offer little insight.

Of considerable interest is the behavior of the cutoff rate when β and N are fixed and T (and simultaneously M) grows arbitrarily large, because this is the regime in which the autocoding effect takes place. In this regime it so happens that the cutoff rate bounds involve integrations over the asymptotic *marginal* distribution of a single eigenvalue, rather than over the joint distribution of all the eigenvalues, and are therefore relatively simple to compute. Moreover, as we show, the asymptotic marginal eigenvalue distributions can be found in simple closed forms.

A.1 Asymptotic cutoff rates

We derive the asymptotic cutoff rate bounds for β and N fixed, and $T \to \infty$. Because the derivations are rather intricate, we begin by outlining the argument for the known channel and Gaussian signals (Section 4.3.1). The remaining three cases (Sections 4.3.2–4.4.2) follow similar lines of reasoning.

A.1.1 H known and Gaussian signals

Define

$$x_T \triangleq \frac{N}{\beta M}\sum_{m=1}^{M}\log\left[1+\frac{\rho}{2}\lambda_m\right], \qquad \bar{x}_T \triangleq Ex_T = \frac{N}{\beta}\mathrm{E}_\lambda \log\left(1+\frac{\rho}{2}\lambda\right),$$

where λ denotes any of the M eigenvalues. Using equation (33) for $\beta > 1$, we can write

$$\begin{aligned} R_0(T,\beta,N) &\geq -\log\left[\mathrm{E}_S 2^{-Tx_T}\right]^{1/T}, \\ &= -\log\left[\mathrm{E}_S 2^{-T(x_T-\bar{x}_T)}\right]^{1/T} 2^{-\bar{x}_T}, \\ &= \frac{N}{\beta}\mathrm{E}_\lambda \log\left(1+\frac{\rho}{2}\lambda\right) - \log\left[\mathrm{E}_S 2^{-T(x_T-\bar{x}_T)}\right]^{1/T}. \end{aligned}$$

As $T \to \infty$ (equivalently, as $M \to \infty$ because β is fixed), by the law of large numbers, x_T converges to its mean value $\bar{x}_T$. (Even though the $\{\lambda_m\}$ are not independent, it can be verified that they satisfy an appropriate mixing condition so that the law of large number holds.) Moreover, if the convergence of x_T to $\bar{x}_T$ is "fast" enough, i.e., if the moments $E(x_T - \bar{x}_T)^k$ vanish fast enough, then it can be shown that

$$\lim_{T\to\infty}\left[\mathrm{E}_S 2^{-T(x_T-\bar{x}_T)}\right]^{1/T} = 1. \tag{A.1}$$

For example, by expanding $\mathrm{E}_S 2^{-T(x_T-\bar{x}_T)}$ in terms of its moments, it can be seen that a sufficient condition for this to hold is that $E(x_T-\bar{x}_T)^k = O(\frac{1}{T^{\alpha k}})$, for some $\alpha > 0$. Although we do not explicitly compute the moments for the eigenvalue distribution here, we believe (A.1) to be true. This contention is supported by the Monte-Carlo integrations presented in Section 5.1, as well as by the closed-form expressions for the cut-off rate bounds that can be evaluated for finite (T, M) using the method of Appendix B.

We thus have the following results.

- $\beta \geq 1$:

$$\lim_{T\to\infty} R_0(T,\beta,N) \geq \frac{N}{\beta}\mathrm{E}_\lambda \log\left(1+\frac{\rho}{2}\lambda\right), \tag{A.2}$$

where λ is an arbitrary eigenvalue of the $M \times M$ random matrix $\frac{1}{M}S^\dagger S$, as $M \to \infty$.

- $0 \le \beta \le 1$:
$$\lim_{T\to\infty} R_0(T,\beta,N) \;\ge\; N\mathrm{E}_\lambda \log\left(1+\frac{\rho\beta}{2}\lambda\right), \tag{A.3}$$
where λ is an arbitrary eigenvalue of the $T \times T$ random matrix $\frac{1}{T}SS^\dagger$, as $T \to \infty$.

Note that (A.2) and (A.3) essentially show that to compute the asymptotic cutoff rate we may interchange the order of the $\log$ and E_S in the formulae (33) and (34). A similar argument leads to the same conclusion for the remaining three cutoff rate expressions.

A.1.2 H known: Unitary space-time signals

$$\lim_{T\to\infty} R_0(T,\beta,N) \;\ge\; \frac{N}{\beta}\mathrm{E}_\lambda \log\left(1+\frac{\rho\beta}{2}-\frac{\rho\beta}{2}\lambda\right), \tag{A.4}$$

where λ is an arbitrary eigenvalue of the $M \times M$ random matrix $\frac{1}{2}(\Theta+\Theta^\dagger)$ (with Θ as in Sec. 4.3.2), as $M \to \infty$.

A.1.3 H unknown: Gaussian signals

- $\beta \ge 2$:
$$\lim_{T\to\infty} R_0(T,\beta,N) \;\ge\; \frac{2N}{\beta}\mathrm{E}_{\lambda_{12}} \log\left(1+\rho\lambda_{12}\right) - \frac{N}{\beta}\mathrm{E}_{\lambda_1}\log\left(1+\rho\lambda_1\right), \tag{A.5}$$
where λ_1 and λ_{12} are arbitrary eigenvalues of the $M \times M$ and $2M \times 2M$ random matrices, $\frac{1}{M}S^\dagger S$ and $\frac{1}{2M}S_{12}^\dagger S_{12}$, as $M \to \infty$, respectively.

- $1 \le \beta \le 2$:
$$\lim_{T\to\infty} R_0(T,\beta,N) \;\ge\; N\mathrm{E}_{\lambda_{12}} \log\left(1+\frac{\rho\beta}{2}\lambda_{12}\right) - \frac{N}{\beta}\mathrm{E}_{\lambda_1}\log\left(1+\rho\lambda_1\right), \tag{A.6}$$
where λ_1 and λ_{12} are arbitrary eigenvalues of the $M \times M$ and $T \times T$ random matrices, $\frac{1}{M}S^\dagger S$ and $\frac{1}{T}S_{12}S_{12}^\dagger$, as $M \to \infty$, respectively.

- $0 \le \beta \le 1$:
$$\lim_{T\to\infty} R_0(T,\beta,N) \;\ge\; N\mathrm{E}_{\lambda_{12}} \log\left(1+\frac{\rho\beta}{2}\lambda_{12}\right) - N\mathrm{E}_{\lambda_1}\log\left(1+\rho\beta\lambda_1\right), \tag{A.7}$$

where λ_1 and λ_{12} are arbitrary eigenvalues of the $T \times T$ random matrices, $\frac{1}{T}SS^\dagger$ and $\frac{1}{T}S_{12}S_{12}^\dagger$, as $T \to \infty$, respectively.

A.1.4 H unknown: Unitary space-time signals

$$\lim_{T\to\infty} R_0(T,\beta,N) \ \geq \ \frac{N}{\beta}\mathrm{E}_\lambda \log\left(1 + \frac{(\rho\beta)^2(1-\lambda)}{4(1+\rho\beta)}\right), \tag{A.8}$$

where λ is an arbitrary eigenvalue of the $M \times M$ random matrix $\Theta\Theta^\dagger$ (with Θ as in Sec. 4.4.2), as $M \to \infty$.

A.2 Asymptotic eigenvalue distributions

In this section we shall obtain the asymptotic marginal eigenvalue distributions necessary to compute the cutoff rate bounds given in (A.2)–(A.8). The asymptotic marginal distributions for the Gaussian signals are well known and can be found, for example, in [20, 21, 22]. The marginal eigenvalue distributions that we obtain for the unitary space-time signals appear to be new, and so we outline their derivation in some detail. We use the following procedure, which can also be used to obtain the asymptotic marginal distribution for a wide class of random matrices (including those arising from the Gaussian signals):

1. Obtain the joint probability distribution of the elements of the matrix of interest.

2. Obtain the joint probability distribution of the eigenvalues by "integrating out" the eigenvectors. This can be readily done when the distribution of the elements is invariant to unitary similarity transformations.

3. Obtain the marginal distribution of a single eigenvalue from the joint distribution using a technique of Wigner [23]. This applies to a special class of eigenvalue distributions, and the resulting marginal distribution is expressed as a Christoffel-Darboux sum of orthogonal polynomials.

4. Obtain the asymptotic marginal distribution by studying the asymptotics of the Christoffel-Darboux sum using a saddlepoint integration. (This approach is used for Wishart matrices in [22].)

A.3 The distribution for $\Theta\Theta^\dagger$

We begin by considering the asymptotic marginal eigenvalue distribution required to compute the cutoff rate bound when H is unknown for unitary space-time signals. (This is perhaps the most interesting of the four cases.)

Recall from (38) and (53) that in this case the random matrix of interest is

$$\Theta\Theta^\dagger = \begin{bmatrix} I_M & 0 \end{bmatrix} \Phi\Phi^\dagger \begin{bmatrix} I_M \\ 0 \end{bmatrix},$$

where Φ is a $T \times M$ ($T \geq M$) isotropically distributed random unitary matrix. By this we mean that

(a) $\Phi\Phi^\dagger = I_M$.

(b) $p_{\mathbf{\Phi}}(\Phi) = p_{\mathbf{\Phi}}(U\Phi V)$, for any square unitary matrices $U \in \mathcal{C}^{T\times T}$ and $V \in \mathcal{C}^{M\times M}$.

A.3.1 Relation to ratio of Wishart matrices

The following lemma provides a simple characterization of isotropically distributed random unitary matrices.

Lemma 1 (Isotropically distributed random matrices) *Let G_1 and G_2 be $M \times M$ and $(T-M)\times M$ random matrices with independent $\mathcal{CN}(0,1)$ entries. Then the matrix*

$$\Phi = \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \left(G_1^\dagger G_1 + G_2^\dagger G_2\right)^{-\dagger/2}, \tag{A.9}$$

is a $T \times M$ isotropically distributed random unitary matrix, where $(G_1^\dagger G_1 + G_2^\dagger G_2)^{1/2}(G_1^\dagger G_1 + G_2^\dagger G_2)^{\dagger/2} = G_1^\dagger G_1 + G_2^\dagger G_2$.

Proof: We can readily confirm that $\Phi^\dagger\Phi = I_M$. Moreover, the invariance of the distribution of Φ to pre- and post-multiplication by unitary matrices follows because for any unitary matrix U,

$$\begin{bmatrix} G_1' \\ G_2' \end{bmatrix} = U \begin{bmatrix} G_1 \\ G_2 \end{bmatrix},$$

is a $T \times M$ matrix of independent $\mathcal{CN}(0,1)$ entries, $G_1^\dagger G_1 + G_2^\dagger G_2 = (G_1')^\dagger G_1' + (G_2')^\dagger G_2'$, and, for any unitary V, $(G_1^\dagger G_1 + G_2^\dagger G_2)^{-\dagger/2}V$ is still a square root of $(G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}$. ∎

We also note that since the distribution function of a $T \times M$ isotropically distributed random unitary matrix is unique, the above lemma gives a complete characterization of such matrices. Moreover, we observe that our matrix of interest is given by

$$\Theta\Theta^\dagger = \Phi_1\Phi_1^\dagger = G_1(G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}G_1^\dagger. \tag{A.10}$$

This leads us to the following result.

Lemma 2 (Ratio of Wishart matrices) *Let $\beta \geq 2$. Then λ is an eigenvalue of $\Theta\Theta^\dagger$ if and only if $\mu = \frac{\lambda}{1-\lambda}$ is an eigenvalue of $G_1^\dagger G_1 \left(G_2^\dagger G_2\right)^{-1}$.*

Proof: For any two nonsingular square matrices A and B, AB and BA have identical eigenvalues. Since G_1 is $M \times M$ and nonsingular (with probability one), we conclude that $\Theta\Theta^\dagger = G_1(G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}G_1^\dagger$ and $(G_1^\dagger G_1 + G_2^\dagger G_2)^{-1}G_1^\dagger G_1$ have the same eigenvalues. Now let λ be one such eigenvalue with corresponding eigenvector X. Then

$$\begin{aligned}
\left(G_1^\dagger G_1 + G_2^\dagger G_2\right)^{-1} G_1^\dagger G_1 X = \lambda X \quad &\Leftrightarrow \quad G_1^\dagger G_1 X = \left(G_1^\dagger G_1 + G_2^\dagger G_2\right)\lambda X \\
&\Leftrightarrow \quad G_1^\dagger G_1 X(1-\lambda) = G_2^\dagger G_2 X \lambda \\
&\Leftrightarrow \quad \left(G_2^\dagger G_2\right)^{-1} G_1^\dagger G_1 X = X\frac{\lambda}{1-\lambda} = X\mu,
\end{aligned}$$

where in the last step we use the fact that $G_2^\dagger G_2$ is invertible because $(T-M) \geq M$ (recall that $\beta \geq 2$). ■

This lemma tells us that when $\beta \geq 2$ the eigenvalue distribution for $\Theta\Theta^\dagger$ can be found from the eigenvalue distribution of the *ratio of two independent Wishart matrices*, $G_1^\dagger G_1$ and $G_2^\dagger G_2$. For $1 \leq \beta < 2$, we have the following result.

Lemma 3 *Let $1 \leq \beta < 2$. Then $\Theta\Theta^\dagger$ has $2M - T$ unit eigenvalues. The remaining $T - M$ eigenvalues have the same distribution as the eigenvalues of a $(T-M)\times(T-M)$ matrix $\Theta\Theta^*$, with β replaced by $\frac{\beta}{1-\beta} > 2$.*

Proof: Let us augment the matrix Φ in (A.9) with $T-M$ additional columns to obtain the $T\times T$ isotropically distributed unitary matrix

$$\Phi_T = \begin{bmatrix} \Phi_1 & \Phi_3 \\ \Phi_2 & \Phi_4 \end{bmatrix}.$$

Since $\Phi_T\Phi_T^\dagger = I_T$, we have $\Phi_1\Phi_1^\dagger + \Phi_3\Phi_3^\dagger = I_M$. Moreover, since Φ_3 is $M \times (T-M)$, the matrix $\Phi_3\Phi_3^\dagger$ has rank $T - M < M$. Thus, $\Phi_1\Phi_1^\dagger$ has $M - (T-M) = 2M - T$ eigenvalues at unity. This establishes our first claim.

To establish our second claim, we note that Φ_T is unitary and hence

$$\Phi_1\Phi_2^\dagger + \Phi_3\Phi_4^\dagger = 0 \quad \text{and} \quad \Phi_1^\dagger\Phi_3 + \Phi_2^\dagger\Phi_4 = 0.$$

Post-multiplying the first equation by Φ_4, and using the second equation to replace $\Phi_2^\dagger\Phi_4$ by $-\Phi_1^\dagger\Phi_3$, we obtain

$$\Phi_1\Phi_1^\dagger\Phi_3 = \Phi_3\Phi_4^\dagger\Phi_4. \tag{A.11}$$

The above equation shows that Φ_3 is an invariant subspace of both $\Phi_1\Phi_1^\dagger$ and $\Phi_4^\dagger\Phi_4$. This implies that every eigenvalue of $\Phi_4^\dagger\Phi_4$ must be an eigenvalue of $\Phi_1\Phi_1^\dagger$. Indeed, let $\Phi_4^\dagger\Phi_4 = U\Lambda U^\dagger$, where U is $(T-M)\times(T-M)$ and unitary, and Λ is $(T-M)\times(T-M)$ and diagonal. Then,

$$\Phi_1\Phi_1^\dagger\Phi_3 U = \Phi_3 U\Lambda.$$

Since $\Phi_3 U$ has $T-M$ independent columns, we conclude that every eigenvalue of $\Phi_4^\dagger\Phi_4$ must be an eigenvalue of $\Phi_1\Phi_1^\dagger$.

Finally, since $\Phi_4^\dagger\Phi_4$ has no eigenvalues at unity, we conclude that all the non-unit eigenvalues of $\Phi_1\Phi_1^\dagger$ are distributed as the eigenvalues of the $(T-M)\times(T-M)$ matrix $\Phi_4^\dagger\Phi_4$, or equivalently $\Phi_4\Phi_4^\dagger$, which has an effective β of

$$\frac{T}{T-M} = \frac{\beta}{\beta-1} > 2.$$

■

The above lemma shows how to obtain the eigenvalue distribution for $\beta < 2$ from the eigenvalue distribution for $\beta > 2$. Thus, in what follows, we assume that $\beta \geq 2$.

A.3.2 The distribution of the elements of $G_1^\dagger G_1\left(G_2^\dagger G_2\right)^{-1}$

The elements of the matrix $A = G_1^\dagger G_1\left(G_2^\dagger G_2\right)^{-1}$, where G_1 and G_2 are $M\times M$ and $(T-M)\times M$ matrices of independent $\mathcal{CN}(0,1)$ entries is given by the so-called multivariate F-distribution

$$p_{\mathbf{A}}(A) = \frac{\tilde{\Gamma}_M(T)}{\tilde{\Gamma}_M(M)\tilde{\Gamma}_M(T-M)}\cdot\frac{1}{\det\left[I_M + A\right]^T}, \tag{A.12}$$

where $\tilde{\Gamma}_m(n)$ is the *complex multivariate gamma function*:

$$\tilde{\Gamma}_m(n) = \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(n-i+1). \tag{A.13}$$

The multivariate F distribution was first obtained by Constantine [24], and can also be found in [25, 26]. The distribution can be derived by considering the joint distribution of the independent Wishart matrices $W_1 = G_1^\dagger G_1$ and $W_2 = G_2^\dagger G_2$:

$$p_{\mathbf{W}_1,\mathbf{W}_2}(W_1, W_2) = \frac{1}{2^{MT}\tilde{\Gamma}_M(M)\tilde{\Gamma}_M(T-M)} e^{-\frac{1}{2}\mathrm{tr}W_1} e^{-\frac{1}{2}\mathrm{tr}W_2} \det W_2^{T-2M},$$

showing that the Jacobian in going from (W_1, W_2) to $(A = W_1 W_2^{-1}, W_2)$ is given by $\det W_2^M$, and finally evaluating the integral

$$p_{\mathbf{A}}(A) = \int_{W_2} dW_2 \frac{1}{2^{MT}\tilde{\Gamma}_M(M)\tilde{\Gamma}_M(T-M)} e^{-\frac{1}{2}\mathrm{tr}AW_2} e^{-\frac{1}{2}\mathrm{tr}W_2} \det W_2^{T-2M} \det W_2^M.$$

A.3.3 The distribution of the eigenvalues of $G_1^\dagger G_1 \left(G_2^\dagger G_2\right)^{-1}$

Note that the multivariate F-distribution (A.12) depends only on the eigenvalues of A. This simplifies considerably the transition from the distribution of the elements of A to the distribution of its eigenvalues.

Consider the eigenvalue decomposition

$$A = UDU^\dagger, \tag{A.14}$$

with U unitary and $D = \mathrm{diag}(d_1, \ldots, d_M)$. The above eigenvalue decomposition can be regarded as a change of variables from A to (U, D). The Jacobian corresponding to this change of variables, assuming the $\{d_i\}$ are unordered, is given by (see, e.g., [26, 27])

$$dA = \frac{1}{M!} \prod_{i<j} (d_i - d_j)^2 dD dU. \tag{A.15}$$

This Jacobian depends only on the eigenvalues of A. Using (A.12) and (A.15) we may write

$$p_{\mathbf{D},\mathbf{U}}(D, U) = \frac{\tilde{\Gamma}_M(T)}{M!\tilde{\Gamma}_M(M)\tilde{\Gamma}_M(T-M)} \cdot \prod_{m=1}^{M} \left(\frac{1}{1+d_m}\right)^T \prod_{i<j} (d_i - d_j)^2.$$

Since the above distribution is independent of U, integrating over U is straightforward and we need only scale by the volume contribution

$$\int_U dU = \frac{\pi^{M(M-1)}}{\tilde{\Gamma}_M(M)}. \tag{A.16}$$

Thus,

$$p_{\mathbf{D}}(d_1, \ldots, d_M) = \frac{\pi^{M(M-1)}\tilde{\Gamma}_M(T)}{M!\tilde{\Gamma}^2_M(M)\tilde{\Gamma}_M(T-M)} \cdot \prod_{m=1}^{M} \left(\frac{1}{1+d_m}\right)^T \prod_{i<j}(d_i - d_j)^2. \tag{A.17}$$

A.3.4 The distribution of the eigenvalues of $\Theta\Theta^\dagger$

In view of Lemma 2, the eigenvalues of $\Theta\Theta^\dagger$ are related to those of A via

$$d_m = \frac{\lambda_m}{1-\lambda_m}, \quad m = 1, \ldots M.$$

The Jacobian corresponding to this transformation is given by

$$\prod_{m=1}^{M} \frac{1}{(1-\lambda_m)^2}.$$

Thus,

$$p_\Lambda(\lambda_1, \ldots, \lambda_M) = \frac{\pi^{M(M-1)}\tilde{\Gamma}_M(T)}{M!\tilde{\Gamma}^2_M(M)\tilde{\Gamma}_M(T-M)} \cdot \prod_{m=1}^{M} \left(\frac{1}{1+\frac{\lambda_m}{1-\lambda_m}}\right)^T \prod_{i<j}(\frac{\lambda_i}{1-\lambda_i} - \frac{\lambda_j}{1-\lambda_j})^2 \prod_{m=1}^{M} \frac{1}{(1-\lambda_m)^2},$$

which upon simplification yields

$$p_\Lambda(\lambda_1, \ldots, \lambda_M) = \frac{\pi^{M(M-1)}\tilde{\Gamma}_M(T)}{M!\tilde{\Gamma}^2_M(M)\tilde{\Gamma}_M(T-M)} \cdot \prod_{m=1}^{M} (1-\lambda_m)^{T-2M} \prod_{i<j}(\lambda_i - \lambda_j)^2. \tag{A.18}$$

A.3.5 The marginal eigenvalue distribution

To obtain the marginal distribution, we need to integrate out the variables, $\lambda_2, \ldots, \lambda_M$. An effective way to do this for any distribution of the form:

$$p(\lambda_1, \ldots, \lambda_M) = K \prod_{m=1}^{M} f(\lambda_m) \prod_{i<j}(\lambda_i - \lambda_j)^2, \tag{A.19}$$

was introduced by Wigner [23].[5]

Lemma 4 (Wigner [23]) *Let the random variables $\lambda_1, \ldots, \lambda_M$ be distributed according to (A.19). Then the marginal distribution for an arbitrary λ is given by*

$$p(\lambda) = \frac{f(\lambda)}{M} \sum_{i=0}^{M-1} \phi_i^2(\lambda), \tag{A.20}$$

where the $\{\phi_m(\lambda)\}$ are polynomials orthonormal with respect to the weighting function $f(\lambda)$.

Outline of Proof:

- Regard $\prod_{i<j}(\lambda_i - \lambda_j)$ as the determinant of a Vandermonde matrix:

$$V = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_M \\ \vdots & \vdots & \ldots & \vdots \\ \lambda_1^{M-1} & \lambda_2^{M-1} & \ldots & \lambda_M^{M-1} \end{bmatrix} \tag{A.21}$$

- Apply row operations to the Vandermonde matrix, to express it in terms of $\{\phi_m(\lambda)\}$, polynomials orthonormal with respect to the weighting function $f(\lambda)$. Thus,

$$V = L \begin{bmatrix} \phi_0(\lambda_1) & \phi_0(\lambda_2) & \ldots & \phi_0(\lambda_M) \\ \phi_1(\lambda_1) & \phi_1(\lambda_2) & \ldots & \phi_1(\lambda_M) \\ \vdots & \vdots & \ldots & \vdots \\ \phi_{M-1}(\lambda_1) & \phi_{M-1}(\lambda_2) & \ldots & \phi_{M-1}(\lambda_M) \end{bmatrix},$$

  with L lower triangular.

- Expand the squared Vandermonde determinant in terms of the $\{\phi_m(\lambda)\}$, and note that upon integrating over $\lambda_2, \ldots, \lambda_M$ all the cross-terms integrate to zero, thus yielding the desired result.

■

[5] Wigner introduced this method in his study of the so-called *Gaussian unitary ensemble*, $A = \frac{1}{2}(G + G^\dagger)$, where G is a square matrix of independent $\mathcal{CN}(0, 1)$ elements, and for which $f(\lambda) = e^{-\lambda}$. Of course, his method readily generalizes to any weighting function $f(\cdot)$.

The sum in (A.20) has a Christoffel-Darboux form and can be further simplified as

$$p(\lambda) = \frac{f(\lambda)}{M} \cdot \frac{a_{M-1}}{a_M} \left[\phi_M^{'}(\lambda)\phi_{M-1}(\lambda) - \phi_{M-1}^{'}(\lambda)\phi_M(\lambda) \right]. \tag{A.22}$$

where a_m is the coefficient of λ^m in $\phi_m(\lambda)$ (see, e.g., [28]).

In our case, since

$$f(\lambda) = \begin{cases} (1-\lambda)^{T-2M} = (1-\lambda)^{(\beta-2)M} & 0 \le \lambda \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{A.23}$$

the $\phi_m(\lambda)$ are (shifted and scaled) Jacobi polynomials. In fact, (A.22) can be written as

$$\begin{aligned} p(\lambda) \quad = \quad & \frac{2(1-\lambda)^{(\beta-2)M}\,((\beta-1)M)!}{\beta M\,((\beta-2)M-1)!} \cdot \left[P_M^{'((\beta-2)M,0)}(2\lambda-1) P_{M-1}^{((\beta-2)M,0)}(2\lambda-1) - \right. \\ & \left. P_{M-1}^{'((\beta-2)M,0)}(2\lambda-1) P_M^{((\beta-2)M,0)}(2\lambda-1) \right], \end{aligned} \tag{A.24}$$

where $P_m^{(\mu,\nu)}(x)$ is the mth order Jacobi polynomial, orthogonal over $[-1,1]$, with respect to the weighting function $(1-x)^{\mu}(1+x)^{\nu}$, $\mu, \nu > -1$.

A.3.6 The asymptotic marginal distribution

Computing the asymptotic behavior, as $M \to \infty$, of the density (A.24) appears to be a formidable task, since both the *order* (the subscript), and the *kind* (the superscript), of the Jacobi polynomials are growing with M.

Computing the asymptotic marginal density of the eigenvalues of a Wishart matrix by studying the limiting behavior of a Christoffel-Darboux sum is done by Mazo in [22]. We mimic his approach, essentially replacing the associated Laguerre polynomials that arise in the context of Wishart matrices, with the Jacobi polynomials that arise here.

The main idea is to use the contour integral representation of the Jacobi polynomials

$$P_m^{((\beta-2)M,0)}(2\lambda-1) = \frac{(-1)^m}{2^{(\beta-2)M+m+1}\pi j} \oint \left(\frac{1-z^2}{z-2\lambda+1} \right)^m \left(\frac{1-z}{1-\lambda} \right)^{(\beta-2)M} dz, \tag{A.25}$$

and then to evaluate the resulting integrals in (A.24) using the saddlepoint method.

Differentiation of $P_m^{((\beta-2)M,0)}(x)$ with respect to its argument in the above contour integral yields

$$\begin{aligned} P_m^{'((\beta-2)M,0)}(2\lambda-1) &= \frac{m(-1)^m}{2^{(\beta-2)M+m+1}\pi j}\oint \frac{(1-z^2)^m}{(z-2\lambda+1)^{m+1}}\cdot\left(\frac{1-z}{1-\lambda}\right)^{(\beta-2)M} dz + \\ &\quad \frac{(\beta-2)M}{2(1-\lambda)}P_m^{((\beta-2)M,0)}(2\lambda-1). \end{aligned}$$

Defining

$$A_{kl} \triangleq \oint \left[\left(\frac{1-z^2}{z-2\lambda+1}\right)\left(\frac{1-z}{1-\lambda}\right)^{\beta-2}\right]^M \frac{(1-z^2)^k}{(z-2\lambda+1)^l}dz, \tag{A.26}$$

we may write (A.24) as

$$\begin{aligned} p(\lambda) &= \frac{2(1-\lambda)^{(\beta-2)M}\,((\beta-1)M)!}{\beta M\,((\beta-2)M-1)!}\cdot\frac{1}{2^{2(\beta-1)M}\pi^2} \\ &\quad \left[\left(MA_{01}+\frac{(\beta-2)M}{2(1-\lambda)}A_{00}\right)A_{-1-1}-\left((M-1)A_{-10}++\frac{(\beta-2)M}{2(1-\lambda)}A_{-1-1}\right)A_{00}\right] \\ &= \frac{2(1-\lambda)^{(\beta-2)M}\,((\beta-1)M)!}{\beta M\,((\beta-2)M-1)!}\cdot\frac{1}{2^{2(\beta-1)M}\pi^2}\left(MA_{01}A_{-1-1}-(M-1)A_{-10}A_{00}\right). \end{aligned} \tag{A.27}$$

To obtain the limit of the above expression as $M\to\infty$ we need to compute the asymptotic value of the integrals A_{kl} as given by (A.26). We may note that these integrals may be rewritten as

$$A_{kl} = \oint e^{Mh(z)}\frac{(1-z^2)^k}{(z-2\lambda+1)^l}dz, \tag{A.28}$$

where

$$h(z) = \ln(1-z^2) - \ln(z-2\lambda+1) + (\beta-2)\ln(1-z) - (\beta-2)\ln(1-\lambda). \tag{A.29}$$

The standard saddlepoint evaluation yields, for large M,

$$A_{kl} \approx \sum_q je^{-\frac{j}{2}\arg h''(z_q)}\sqrt{\frac{2\pi}{M|h''(z_q)|}}\left[\left(\frac{1-z_q^2}{z_q-2\lambda+1}\right)\left(\frac{1-z_q}{1-\lambda}\right)^{\beta-2}\right]^M\frac{(1-z_q^2)^k}{(z_q-2\lambda+1)^l}, \tag{A.30}$$

where the $\{z_q\}$ represent the saddlepoints of $h(z)$, i.e., the roots of $h^{'}(z)=0$. In our case,

$$h^{'}(z) = -\frac{2z}{1-z^2} - \frac{1}{z-2\lambda+1} - \frac{\beta-2}{1-z}.$$

Upon simplification, the equation $h^{'}(z) = 0$ reduces to the quadratic equation:

$$(\beta - 1)z^2 - [\beta(2\lambda - 1) - (\beta - 2)]\, z + 1 - (\beta - 2)(2\lambda - 1) = 0,$$

so the saddlepoints are

$$z_{q_1,q_2} = \frac{\beta(2\lambda - 1) - (\beta - 2) \pm \sqrt{\Delta}}{2(\beta - 1)}, \tag{A.31}$$

where

$$\Delta = 4\lambda \left[\beta^2\lambda - 4(\beta - 1)\right]. \tag{A.32}$$

For $\lambda \geq \frac{4(\beta-1)}{\beta^2}$ or $\lambda \leq 0$, the two saddlepoints z_{q_1} and z_{q_2} are real. In this case, evaluating the saddlepoint formula (A.30) and substituting it into (A.27) shows that

$$\lim_{M\to\infty} p(\lambda) = 0, \quad \lambda \geq \frac{4(\beta - 1)}{\beta^2} \;\text{ or }\; \lambda \leq 0. \tag{A.33}$$

On the other hand, when $0 < \lambda < \frac{4(\beta-1)}{\beta^2}$, the two saddlepoints z_{q_1} and z_{q_2} are complex. In this case, evaluating the saddlepoint formula (A.30) and substituting it into (A.27) shows that

$$\lim_{M\to\infty} p(\lambda) = \frac{\beta}{2\pi}\sqrt{\frac{\frac{4(\beta-1)}{\beta^2} - \lambda}{\lambda(1-\lambda)^2}}, \quad 0 < \lambda < \frac{4(\beta - 1)}{\beta^2} \tag{A.34}$$

We can formalize our result in the following theorem.

Theorem 1 *Let Φ be a $T \times M$ ($T \geq M$) isotropically distributed unitary random matrix. Then the asymptotic marginal distribution of any eigenvalue of*

$$\Theta\Theta^\dagger = \begin{bmatrix} I_M & 0 \end{bmatrix} \Phi\Phi^\dagger \begin{bmatrix} I_M \\ 0 \end{bmatrix},$$

as $M \to \infty$, with $\beta = T/M$ fixed, is given by

(i) when $\beta = T/M \geq 2$:

$$p(\lambda) = \begin{cases} \frac{\beta}{2\pi}\sqrt{\frac{\frac{4(\beta-1)}{\beta^2} - \lambda}{\lambda(1-\lambda)^2}} & 0 < \lambda < \frac{4(\beta-1)}{\beta^2} \\ 0 & \textit{otherwise} \end{cases}, \tag{A.35}$$

(ii) and when $\beta = T/M < 2$:

$$p(\lambda) = \begin{cases} (2-\beta)\delta(\lambda-1) + \frac{\beta}{2\pi}\sqrt{\frac{\frac{4(\beta-1)}{\beta^2}-\lambda}{\lambda(1-\lambda)^2}} & 0 < \lambda < \frac{4(\beta-1)}{\beta^2} \\ 0 & \textit{otherwise} \end{cases} . \tag{A.36}$$

Proof: The proof of part (i) follows from the arguments leading to the statement of the theorem. To prove part (ii), we use Lemma 3 which relates the eigenvalue distribution for $\beta < 2$ to the eigenvalue distribution for $\beta^{'} = \frac{\beta}{\beta-1} > 2$. For $\beta < 2$, the proportion $\frac{2M-T}{M} = 2-\beta$ of the eigenvalues are unit, whereas the remaining proportion $1-(2-\beta) = \beta-1$ of the eigenvalues are distributed with $\beta^{'} = \frac{\beta}{\beta-1} > 2$. Thus,

$$p(\lambda) = (2-\beta)\delta(\lambda-1) + (\beta-1)\frac{\beta^{'}}{2\pi}\sqrt{\frac{\frac{4(\beta^{'}-1)}{(\beta^{'})^2}-\lambda}{\lambda(1-\lambda)^2}}, \quad 0 < \lambda < \frac{4(\beta^{'}-1)}{(\beta^{'})^2}.$$

Replacing $\beta^{'}$ by $\frac{\beta}{\beta-1} > 2$ yields (A.36). ∎

Plots of the above asymptotic density, along with the empirical marginal density for $M = 16$, and various values of β are given in Figure 8. As can be seen, we obtain a close match between the empirical and asymptotic distributions even for this moderate value of M.

A.3.7 A trigonometric interpretation

The density of Theorem 1, part (i), has an interesting interpretation. Let

$$\lambda = \frac{4(\beta-1)}{\beta^2}\cos^2\theta, \quad 0 \le \theta \le \frac{\pi}{2} \tag{A.37}$$

then

$$p(\theta) = \frac{4(\beta-1)}{\beta\pi} \cdot \frac{\sin^2\theta}{1-\frac{4(\beta-1)}{\beta^2}\cos^2\theta}, \quad 0 \le \theta \le \frac{\pi}{2}. \tag{A.38}$$

In particular, when $\beta = 2$, we have

$$\lambda = \cos^2\theta \quad , \quad p(\theta) = \frac{2}{\pi}, \tag{A.39}$$

which is uniform over θ. This trigonometric representation of the density is useful for explicitly computing of the cutoff rate bounds.

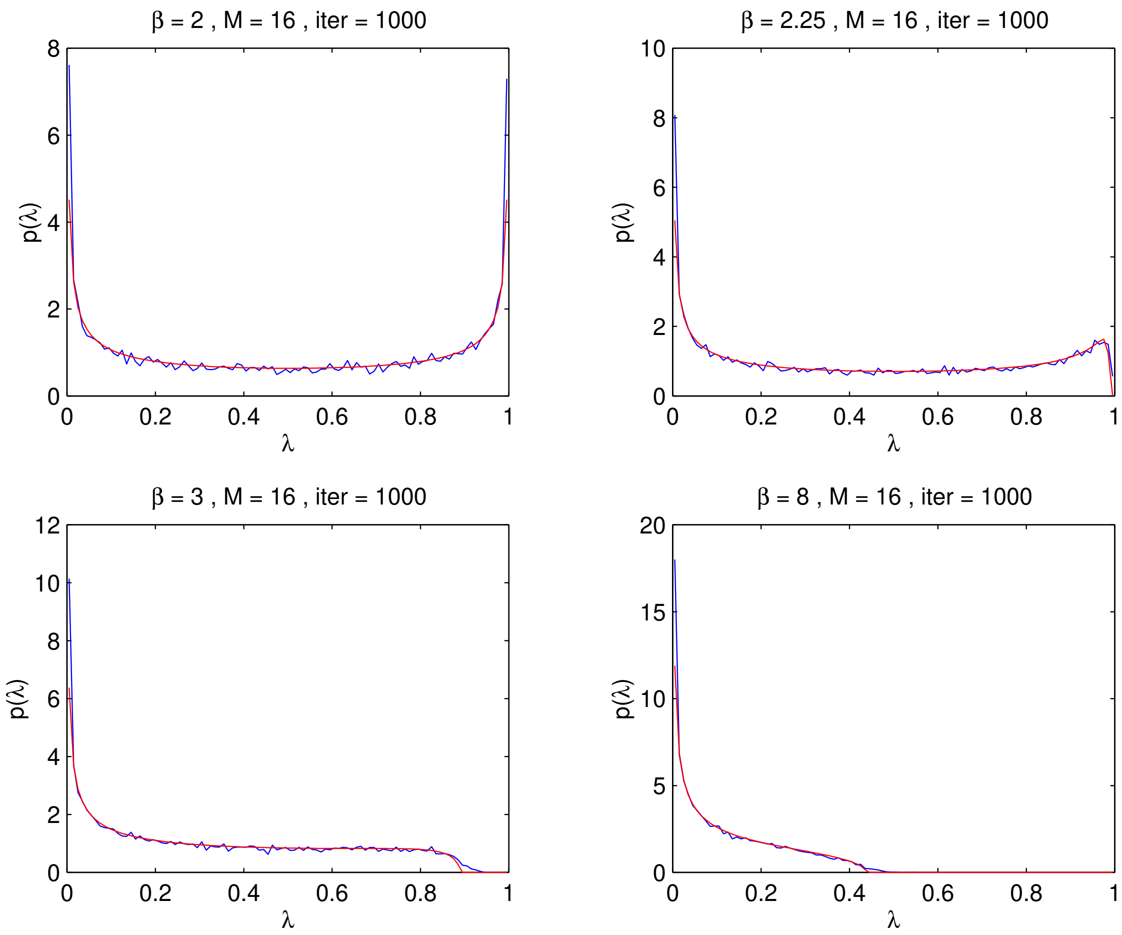

Figure 8: The asymptotic and empirical densities.

A.4 The distribution for $SS^\dagger$

To compute the cutoff rate bounds (A.2)–(A.3) and (A.5)–(A.7), corresponding to Gaussian signals, we require the asymptotic marginal eigenvalue distribution of the matrix $\frac{1}{m}S^\dagger S$, where S is a $n \times m$ ($n \geq m$) matrix of independent $\mathcal{CN}(0,1)$ elements, as $m \to \infty$, with $n/m = \beta$ fixed. This asymptotic marginal distribution is well known (see, e.g., [21, 22]) and can, for example, be computed using the method presented above for the matrix $\Theta\Theta^\dagger$. The result is given in the next theorem.

Theorem 2 *Let S be an $n \times m$ ($n \geq m$) matrix of independent $\mathcal{CN}(0,1)$ elements. Then the asymptotic marginal distribution of any eigenvalue of $\frac{1}{m}S^\dagger S$, as $m \to \infty$, with $n/m = \beta$ fixed, is given by*

$$p(\lambda) = \begin{cases} \frac{1}{\pi}\sqrt{\frac{\beta}{\lambda} - \frac{1}{4}\left(1 + \frac{\beta-1}{\lambda}\right)^2} & (\sqrt{\beta}-1)^2 \leq \lambda \leq (\sqrt{\beta}+1)^2 \\ 0 & \textit{otherwise} \end{cases}. \tag{A.40}$$

A.4.1 A trigonometric interpretation

The density of Theorem 2 has a trigonometric representation. Let

$$\lambda = 4\sqrt{\beta}\sin^2\theta + (\sqrt{\beta}-1)^2, \quad 0 \leq \theta \leq \frac{\pi}{2} \tag{A.41}$$

then

$$p(\theta) = \frac{16\beta}{\pi} \cdot \frac{\sin^2\theta\cos^2\theta}{4\sqrt{\beta}\sin^2\theta + (\sqrt{\beta}-1)^2}, \quad 0 \leq \theta \leq \frac{\pi}{2}. \tag{A.42}$$

A.5 The distribution for $\frac{1}{2}(\Theta + \Theta^\dagger)$

An argument similar to the one used to obtain the asymptotic marginal eigenvalue distribution for the matrix $\Theta\Theta^\dagger$ can be used to obtain the following result.

Theorem 3 *Let Φ be a $T \times M$ ($T \geq M$) isotropically distributed unitary random matrix. Then the asymptotic marginal distribution of any eigenvalue of*

$$\frac{1}{2}\left(\Theta + \Theta^\dagger\right),$$

where $\Theta = \left[\begin{array}{cc} I_M & 0 \end{array} \right] \Phi$, *as* $M \to \infty$, *with* $\beta = T/M$ *fixed, is given by*

$$p(\lambda) = \begin{cases} \frac{\beta}{\pi} \frac{\sqrt{\frac{2\beta-1}{\beta^2} - \lambda^2}}{1-\lambda^2} & 0 < \lambda < \sqrt{\frac{2\beta-1}{\beta^2}} \\ 0 & \textit{otherwise} \end{cases} . \tag{A.43}$$

A.5.1 A trigonometric interpretation

Let

$$\lambda = \frac{\sqrt{2\beta - 1}}{\beta} \cos\theta, \quad 0 \leq \theta \leq \pi. \tag{A.44}$$

Then

$$p(\theta) = \frac{2\beta - 1}{\beta\pi} \cdot \frac{\sin^2\theta}{1 - \frac{2\beta-1}{\beta^2}\cos^2\theta}, \quad 0 \leq \theta \leq \pi. \tag{A.45}$$

In particular, when $\beta = 1$ then

$$\lambda = \cos\theta \quad , \quad p(\theta) = \frac{1}{\pi}, \tag{A.46}$$

which is uniform over θ.

A.6 Briefly on Stieltjes transforms

The simplicity of the formulas for the asymptotic marginal eigenvalue densities suggests that there may be a more direct way of obtaining them. In the case of Wishart matrices that arise for Gaussian signals (see Sec. A.4), it is well known that the Stieltjes transform provides such a way. In what follows, we briefly describe this method and how it can be used to obtain the asymptotic marginal eigenvalue distribution for $\Theta\Theta^\dagger$, which appears to be new. At this time, however, we do not know how to use the Stieltjes transform method to obtain the corresponding distribution for $\frac{1}{2}(\Theta + \Theta^\dagger)$.

The *Stieltjes transform* of a probability density function $p(\cdot)$ is the function

$$m(z) = \int_{-\infty}^{\infty} \frac{p(t)}{t - z} dt, \tag{A.47}$$

defined for all nonreal z. Conversely, $p(\cdot)$ can be recovered from its Stieltjes transform via the inverse transform formula

$$p(t) = \lim_{\epsilon \to 0} \frac{1}{\pi} \mathrm{Im}\left[m(t + j\epsilon)\right] = \frac{1}{\pi} \mathrm{Im}\left[m(t + j0)\right]. \tag{A.48}$$

Theorem 4 (Stieltjes transform) *Consider the matrix $A = \frac{1}{N} X^\dagger X D$, where X is an $N \times N$ matrix of independent, zero-mean, unit variance random variables, and D is an $N \times N$ random diagonal matrix whose diagonal entries have the same marginal distribution but are otherwise arbitrary. Then the Stieltjes transform of the asymptotic marginal distribution of any eigenvalue of A, as $N \to \infty$, satisfies*

$$m(z) = \frac{1}{E_d\left[\frac{d}{1+m(z)d}\right] - z}, \tag{A.49}$$

where the expectation is over d, an arbitrary diagonal entry of D.

The above result is given in [21] and can be considered a strengthening of an earlier result of Marcenko and Pastur [20] that assumed that the diagonal entries of D are independent.

When $D = I_N$, for example, Theorem 4 implies that the Stieltjes transform of the asymptotic marginal eigenvalue distribution for the Wishart matrix $A = \frac{1}{N} X^\dagger X$ satisfies

$$m(z) = \frac{1}{\frac{1}{1+m(z)} - z}.$$

The above equation can be used to solve for $m(z)$, and then (A.48) can be employed to obtain the distribution of Theorem 2, for $\beta = 1$. We now show how Theorem 4 can be used to compute the asymptotic marginal eigenvalue distribution for $\Theta\Theta^\dagger$.

Recall from Lemma 2 that the eigenvalues of $\Theta\Theta^\dagger$ are related to the eigenvalues of

$$G_1^\dagger G_1 \left(G_2^\dagger G_2\right)^{-1} = \frac{1}{M} G_1^\dagger G_1 \left(\frac{1}{M} G_2^\dagger G_2\right)^{-1} = \frac{1}{M} G_1^\dagger G_1 \left(U \Sigma U^\dagger\right)^{-1}, \tag{A.50}$$

where U is an $M \times M$ unitary matrix, and Σ is an $M \times M$ diagonal matrix representing the eigenvalues of the Wishart matrix $\frac{1}{M} G_2^\dagger G_2$. Defining $G = G_1 U$, we see that the eigenvalues of the matrix in (A.50) are the same as those of

$$A = \frac{1}{M} G^\dagger G \Sigma^{-1}.$$

Since the elements of G are independent $\mathcal{CN}(0,1)$, we can appeal to Theorem 4 to write

$$m(z) = \frac{1}{\mathrm{E}_\sigma\left[\frac{\sigma^{-1}}{1+\sigma^{-1}m(z)}\right] - z} = \frac{1}{\mathrm{E}_\sigma\left[\frac{1}{\sigma+m(z)}\right] - z}, \tag{A.51}$$

where σ is an arbitrary eigenvalue of $\frac{1}{M} G_2^\dagger G_2$.

Using the distribution (A.41)–(A.42) for σ (with $\beta \to \frac{T-M}{M} = \beta - 1 \stackrel{\Delta}{=} \rho$), we can write:

$$\mathrm{E}_\sigma\left[\frac{1}{\sigma + m(z)}\right] = \frac{16\rho}{\pi}\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta\cos^2\theta}{(4\sqrt{\rho}\sin^2\theta + (\sqrt{\rho}-1)^2 + m(z))(4\sqrt{\rho}\sin^2\theta + (\sqrt{\rho}-1)^2)} d\theta.$$

This integral has the closed form solution

$$\mathrm{E}_\sigma\left[\frac{1}{\sigma + m(z)}\right] = \frac{1}{2}\sqrt{\frac{(m(z) + (\sqrt{\rho}-1)^2)(m(z) + (\sqrt{\rho}+1)^2)}{m^2(z)}} - \frac{\rho - 1}{2m(z)} - \frac{1}{2}.$$

Inserting this last expression into (A.51) yields

$$m(z) = \frac{1}{\frac{1}{2}\sqrt{\frac{(m(z)+(\sqrt{\rho}-1)^2)(m(z)+(\sqrt{\rho}+1)^2)}{m^2(z)}} - \frac{\rho-1}{2m(z)} - \frac{1}{2} - z},$$

which simplifies to

$$z(1+z)m^2(z) + (1+\rho)m(z) + \rho = 0. \tag{A.52}$$

The solution to this quadratic equation is

$$m(z) = \frac{1}{2z(1+z)}\left[-(1+\rho)z \pm \sqrt{(1-\rho)^2z^2 - 4\rho z}\right]. \tag{A.53}$$

To compute the inverse Stieltjes transform using (A.48) we need to find the imaginary part of $m(t + j0)$. Clearly, $m(t + j0)$ will have an imaginary part if, and only if,

$$(1-\rho)^2t^2 - 4\rho t < 0 \;\Leftrightarrow\; 0 \le t \le \frac{4\rho}{(1-\rho)^2}$$

in which case

$$p(t) = \frac{1}{\pi}\cdot\frac{\sqrt{4\rho t - (1-\rho)^2t^2}}{2t(1+t)}, \quad 0 \le t \le \frac{4\rho}{(1-\rho)^2}. \tag{A.54}$$

Otherwise, $p(t) = 0$. Now in view of Lemma 2, we know that $t = \frac{\lambda}{1-\lambda}$. Using this change of variable, and noting that $\rho = \beta - 1$, we obtain the desired distribution (A.35) for λ.

A.7 Closed-form expressions for the cutoff rate bounds

Having computed the asymptotic marginal distributions, we are in a position to give explicit expressions for the cutoff rate bounds. These require the evaluation of certain definite integrals, as explained below.

A.7.1 Gaussian signals; both H known and H unknown

Inspecting Eqs. (A.2)–(A.3) and (A.5)–(A.7) and using the distribution of Theorem 2, shows that we need to compute the definite integral

$$\begin{aligned} F(\beta,\rho) &\triangleq \frac{1}{\pi}\int_{(\sqrt{\beta}-1)^2}^{(\sqrt{\beta}+1)^2} \log(1+\rho\lambda)\sqrt{\frac{\beta}{\lambda}-\frac{1}{4}\left(1+\frac{\beta-1}{\lambda}\right)^2}\,d\lambda, && \text{(A.55)}\\ &= \frac{16\beta}{\pi}\int_0^{\frac{\pi}{2}} \log\left(1+\rho(\sqrt{\beta}-1)^2+4\rho\sqrt{\beta}\sin^2\theta\right)\frac{\sin^2\theta\cos^2\theta}{4\sqrt{\beta}\sin^2\theta+(\sqrt{\beta}-1)^2}d\theta, && \text{(A.56)}\end{aligned}$$

where in the second step, we used the trigonometric representations (A.41)–(A.42). This integral has the closed-form solution (see, e.g., [29])

$$\begin{aligned} F(\beta,\rho) &= \log\left(1+\rho(\sqrt{\beta}+1)^2\right)+(\beta+1)\log\left(\frac{1+\sqrt{1-a}}{2}\right)\\ &\quad -(\log e)\sqrt{\beta}\frac{1-\sqrt{1-a}}{1+\sqrt{1-a}}+(\beta-1)\log\left(\frac{1+\gamma}{\gamma+\sqrt{1-a}}\right), && \text{(A.57)}\end{aligned}$$

where $a=\frac{4\rho\sqrt{\beta}}{1+\rho(\sqrt{\beta}+1)^2}$ and $\gamma=\frac{\sqrt{\beta}-1}{\sqrt{\beta}+1}$.

A.7.2 H known: unitary space-time signals

In this case, (A.4), Theorem 3, and (A.44)–(A.45) show that we need to evaluate

$$\begin{aligned} &\frac{\beta}{\pi}\int_0^{\sqrt{\frac{2\beta-1}{\beta^2}}} \log\left(1+\frac{\rho\beta}{2}-\frac{\rho\beta}{2}\lambda\right)\frac{\sqrt{\frac{2\beta-1}{\beta^2}-\lambda^2}}{1-\lambda^2}d\lambda\\ &\quad=\frac{2\beta-1}{\beta\pi}\int_0^{\pi}\log\left(1+\frac{\rho\beta}{2}-\frac{\rho\sqrt{2\beta-1}}{2}\cos\theta\right)\frac{\sin^2\theta}{1-\frac{2\beta-1}{\beta^2}\cos^2\theta}d\theta. && \text{(A.58)}\end{aligned}$$

A.7.3 H unknown: unitary space-time signals

In this case, (A.8), Theorem 1, and (A.37)–(A.38) show that we need to evaluate

$$\begin{aligned} &\frac{\beta}{2\pi}\int_0^{\frac{4(\beta-1)}{\beta^2}}\log\left(1+\frac{(\rho\beta)^2(1-\lambda)}{4(1+\rho\beta)}\right)\sqrt{\frac{\frac{4(\beta-1)}{\beta^2}-\lambda}{\lambda(1-\lambda)^2}}d\lambda\\ &\quad=\frac{4(\beta-1)}{\beta\pi}\int_0^{\frac{\pi}{2}}\log\left(1+\frac{\rho^2(\beta-1)\sin^2\theta}{1+\rho\beta}\right)\frac{\sin^2\theta}{1-\frac{4(\beta-1)}{\beta^2}\cos^2\theta}d\theta\end{aligned}$$

$$= \log(1+\alpha) + \beta \log\left(\frac{1+\sqrt{1-\frac{b\alpha}{1+\alpha}}}{2}\right) + (\beta-2)\log\left(\frac{1+\sqrt{1-b}}{\sqrt{1-b}+\sqrt{1-\frac{b\alpha}{1+\alpha}}}\right), \quad \text{(A.59)}$$

where we have defined $\alpha = \frac{(\rho\beta)^2}{4(1+\rho\beta)}$ and $b = \frac{4(\beta-1)}{\beta^2}$, and where we have used the integral tables in [29].

B Appendix: Pairwise P_e for H unknown and Unitary Space-Time Signals

We are interested in computing (56)

$$P_e \leq 2^{RT} \mathrm{E}_{S_1,S_2}\left\{P_e\left\{S_1 \text{ vs. } S_2\right\}\right\}.$$

The conditional pairwise probability of error for two Unitary Space-Time signals (averaging over H and W) is [6]

$$\begin{aligned} P_e\left\{S_1 \text{vs.} S_2\right\} &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\omega \, \frac{1}{\omega^2+1/4} \mathrm{E} \prod_{m=1}^{M}\left[\frac{1}{1+\frac{(\rho T/M)^2(1-\lambda_m)(\omega^2+1/4)}{(1+\rho T/M)}}\right]^N \\ &= \frac{1}{\pi}\int_0^{\pi/2} d\theta \, \mathrm{E} \prod_{m=1}^{M}\left[\frac{\cos^2\theta}{\cos^2\theta + \frac{(\rho T/M)^2(1-\lambda_m)}{4(1+\rho T/M)}}\right]^N, \end{aligned} \quad \text{(B.1)}$$

where we have made the coordinate transformation $\omega = \frac{1}{2}\tan\theta$. When the two signals are independent Unitary Space-Time, the joint probability density for the eigenvalues $p_\Lambda(\lambda_1,\ldots,\lambda_M)$ is given by (A.18).

An effective way to compute the expectation over the eigenvalues in the integrand of (B.1) is given by the following lemma.

Lemma 5 (A special expectation) *Let the random variables $\lambda_1,\ldots,\lambda_M$ be distributed according to*

$$p(\lambda_1,\ldots,\lambda_M) = K\prod_{m=1}^{M} f(\lambda_m)\prod_{i<j}(\lambda_i-\lambda_j)^2.$$

Then, for any positive function $g(\cdot)$, we have

$$E_\lambda \prod_{m=1}^{M} g(\lambda_m) = KM!\det R, \quad \text{(B.2)}$$

where R is an $M \times M$ Hankel matrix whose (k,l) entry is given by

$$R_{kl} = \int \lambda^{k+l-2} f(\lambda) g(\lambda) d\lambda. \tag{B.3}$$

Outline of Proof: We employ a technique similar to the one presented in Lemma 4. We thus regard $\prod_{i<j}(\lambda_i - \lambda_j)$ as the determinant of the Vandermonde matrix, V, and apply row operations to express it in terms of $\{\phi_m(\lambda)\}$, polynomials orthonormal with respect to the weighting function $f(\lambda)g(\lambda)$:

$$V = L \begin{bmatrix} \phi_0(\lambda_1) & \phi_0(\lambda_2) & \dots & \phi_0(\lambda_M) \\ \phi_1(\lambda_1) & \phi_1(\lambda_2) & \dots & \phi_1(\lambda_M) \\ \vdots & \vdots & \dots & \vdots \\ \phi_{M-1}(\lambda_1) & \phi_{M-1}(\lambda_2) & \dots & \phi_{M-1}(\lambda_M) \end{bmatrix},$$

with L lower triangular and real. Expanding the squared Vandermonde determinant in terms of the $\{\phi_m(\lambda)\}$, and noting that upon integration over $\lambda_1, \dots, \lambda_M$ all the cross-terms with non-identical indices integrate to zero, we obtain

$$E_{\lambda_m} \prod_{m=1}^{M} g(\lambda_m) = KM!(\det L)^2. \tag{B.4}$$

(The coefficient $M!$ comes from the facts that, by the orthonormality of the $\{\phi_m(\lambda)\}$, the cross-terms with identical indices integrate to unity, and that there are $M!$ such terms.) To compute $\det^2 L$, let us define

$$y \triangleq \begin{bmatrix} 1 \\ \lambda \\ \vdots \\ \lambda^{M-1} \end{bmatrix} \quad \text{and} \quad z \triangleq \begin{bmatrix} \phi_0(\lambda) \\ \phi_1(\lambda) \\ \vdots \\ \phi_{M-1}(\lambda) \end{bmatrix},$$

and note that $y = Lz$. Thus,

$$R = \int yy^\dagger f(\lambda)g(\lambda)d\lambda = \int Lzz^\dagger L^\dagger f(\lambda)g(\lambda)d\lambda = L\left(\int zz^\dagger f(\lambda)g(\lambda)d\lambda\right)L^\dagger = LL^\dagger,$$

where in the last step we used $\int zz^\dagger f(\lambda)g(\lambda)d\lambda = I_M$, since the $\{\phi_m(\lambda)\}$ are orthonormal. We thus conclude that $\det R = \det^2 L$, which along with (B.4) yields the desired result. ■

We see that the expectation in the integrand of (B.1) can be computed using Lemma 5 and taking

$$g(\lambda) = \left[\frac{\cos^2\theta}{\cos^2\theta + \frac{(\rho T/M)^2(1-\lambda_m)}{4(1+\rho T/M)}} \right]^N, \quad f(\lambda) = (1-\lambda)^{T-2M} \text{ and } K = \frac{\pi^{M(M-1)}\tilde{\Gamma}_M(T)}{M!\tilde{\Gamma}_M^2(M)\tilde{\Gamma}_M(T-M)}.$$

To complete the computation, we integrate over θ numerically. The Chernoff bound, required for the cutoff rate bound, is obtained by upper-bounding the integrand by its value at $\theta = 0$.

We can also use Lemma 5 to compute the cut-off rate and probability of error bounds in the Gaussian case by replacing the above $f(\lambda)$ and K with those given by the Wishart distribution. We do not pursue this approach, however.

Acknowledgements

We thank A. O. Hero, J. E. Mazo, S. Shamai, and J. Ziv for helpful comments during this research.

References

[1] R. McEliece and W. Stark, “Channels with block interference,” *IEEE Trans. Info. Theory*, vol. 30, no. 1, pp. 44–53, Jan. 1984.

[2] 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.

[3] E. Biglieri, J. Proakis, and S. Shamai (Shitz), “Fading Channels: Information-Theoretic and Communications Aspects”, *IEEE Trans. Info. Theory*, vol 44, no. 6, pp. 2619–2992, Oct. 1998.

[4] 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.

[5] 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, no. 1, pp. 139–157, January 1999.

[6] B. M. Hochwald and T. L. Marzetta, “Unitary Space-Time Modulation for Multiple-Antenna Communications in Rayleigh Flat Fading”, *IEEE Trans. Info. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[7] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, “Systematic Design of Unitary Space-Time Constellations”, *IEEE Trans. Info. Theory*, vol. 46, no. 6, pp. 1962–1973, Sept. 2000.

[8] I. Abou-Faycal and B. M. Hochwald, “Coding Requirement for Multiple-Antenna Channels with Unknown Rayleigh Fading”, *Bell Labs. Tech. Memo; submitted to IEEE Trans. Info. Theory*, 1999.

[9] S. V. Hanly and D. N. Tse, “The Multi-access Fading Channel: Shannon and Delay Limited Capacities”, *Proc. 33d Allteron Conf.*, Monticello, IL, pp. 786–795, 1995.

[10] E. Biglieri, G. Caire, and G. Taricco, “Minimum Outage Probability for Slowly-Varying Fading Channels”, *Proc. Int. Symp. Info. Theory*, Cambridge, MA, p. 7, August, 1998.

[11] S. V. Hanly and D. N. Tse, “Multiaccess Fading Channels-Part II: Delay-Limited Capacities”, *IEEE Trans. Info. Theory*, vol. 44, pp. 2816–2831, 1998.

[12] G. Caire, G. Taricco and E. Biglieri, “Optimum Power Control Over Fading Channels”, *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1468–1489, 1999.

[13] 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.

[14] I. E. Telatar, “Capacity of multi-antenna Gaussian channels, *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.

[15] I. C. Abou Faycal, M. D. Trott and S. Shamai (Shitz), “The capacity of discrete-time Rayleigh fading channels,” in *Proc. Int. Symp. Info. Th.*, Ulm, Germany, 1997, p. 473.

[16] R. Blahut, *Principles and Practice of Information Theory*, Reading, MA: Addison-Wesley, 1987.

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

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

[19] T. L. Marzetta, “BLAST Training: Estimating Channel Characteristics for High-Capacity Space-Time Wireless”, *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Sept. 22–24, 1999.

[20] V. A. Marcenko and L. A. Pastur, "Distribution for some sets of random matrices," *Math. USSR-Sb*, vol. 1, pp. 457–483, 1967.

[21] J. W. Silverstein and Z. D. Bai, "On the emprical distribution of eigenvalues of a class of large dimensional random matrices," *J. Mult. Anal.*, vol. 54, pp. 175–192, 1995.

[22] J. E. Mazo, "Asymptotic formula for the eigenvalue distribution of certain random matrices", *Bell Labs. Tech. Mem.*, 1995.

[23] E. P. Wigner, "Random matrices in physics", *SIAM Review*, vol. 9, pp. 1–23, 1967.

[24] A. G. Constantine, "Some noncentral distribution problems in multivariate analysis", *Ann. Math. Statist.*, vol. 34, pp. 1270–1285, 1963.

[25] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples", *Ann. Math. Statist.*, vol. 35, pp. 40–75, 1964.

[26] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, New York: John Wiley and Sons, 1982.

[27] M. L. Mehta, *Random Matrices and Statistical Theory*, New York: Academic Press, 1967.

[28] P. Beckmann, *Orthogonal Polynomials for Engineers and Physicists*, Boulder, CO: The Golem Press, 1973.

[29] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series and Products* New York: Academic Press, 1994.