Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading

THOMAS L. MARZETTA

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

BERTRAND M. HOCHWALD

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

Abstract

We analyze a mobile wireless link comprising M transmitter and N receiver antennas operating in a Rayleigh flat-fading environment. The propagation coefficients between every pair of transmitter and receiver antennas are statistically independent and unknown; they remain constant for a coherence interval of T symbol periods, after which they change to new independent values which they maintain for another T symbol periods, and so on.

Computing the link capacity, associated with channel coding over multiple fading intervals, requires an optimization over the joint density of $T \cdot M$ complex transmitted signals. We prove that there is no point in making the number of transmitter antennas greater than the length of the coherence interval: the capacity for $M > T$ is equal to the capacity for $M = T$. Capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix times a certain $T \times M$ random matrix that is diagonal, real, and nonnegative. This result enables us to determine capacity for many interesting cases. We conclude that, for a fixed number of antennas, as the length of the coherence interval increases, the capacity approaches the capacity obtained as if the receiver knew the propagation coefficients.

Index Terms—Multi-element antenna arrays, wireless communications, space-time modulation

1 Introduction

It is likely that future breakthroughs in wireless communication will be driven largely by high data rate applications. Sending video rather than speech, for example, increases the data rate by two or three orders-of-magnitude. Increasing the link or channel bandwidth is a simple but costly—and ultimately unsatisfactory—remedy. A more economical solution is to exploit propagation diversity through multiple-element transmitter and receiver antenna arrays.

It has been shown [3, 7] that, in a Rayleigh flat-fading environment, a link comprising multiple-element antennas has a theoretical capacity that increases linearly with the smaller of the number of transmitter and receiver antennas, provided that the complex-valued propagation coefficients between all pairs of transmitter and receiver antennas are statistically independent and known to the receiver (but not the transmitter). The independence of the coefficients provides diversity, and is often achieved by physically separating the antennas at the transmitter and receiver by a few carrier wavelengths. With such wide antenna separations, the traditional adaptive array concepts of beam pattern and directivity do not directly apply.

If the time between signal fades is sufficiently long—often a reasonable assumption for a fixed wireless environment—then the transmitter can send training signals that allow the receiver to estimate the propagation coefficients accurately, and the results of [3, 7] are applicable. With a mobile receiver, however, the time between fades may be too short to permit reliable estimation of the coefficients. A 60 mile/hour mobile operating at 1.9 GHz has a fading interval of about 3 ms, which for a symbol rate of 30 kHz, corresponds to only about 100 symbol periods. We approach the problem of determining the capacity of a time-varying multiple-antenna communication channel, using the tools of information theory, and without any ad hoc training schemes in mind.

The propagation coefficients, which neither the transmitter nor the receiver knows, are assumed to be constant for T symbol periods, after which they change to new independent random values which they maintain for another T symbol periods, and so on. This piecewise constant fading process approximates, in a tractable manner, the behavior of a continuously fading process such as Jakes' [5]. Furthermore, it is a very accurate representation of many TDMA, frequency hopping, or block-interleaved systems. The random propagation coefficients are modelled as independent, identically distributed, zero-mean, circularly symmetric complex Gaussian random variables. Thus there are two sources of noise at work: multiplicative noise that is associated with the Rayleigh fading, and the usual additive receiver noise.

Suppose that there are M transmitter and N receiver antennas. Then the link is completely described by the conditional probability density of the $T \cdot N$ complex received signals given the $T \cdot M$ complex transmitted signals. Although this conditional density is complex Gaussian, the transmitted signals affect only the conditional covariance (rather than the mean) of the received signals—a source of difficulty in the problem.

If one performs channel coding over multiple independent fading intervals, information theory tells us that it is theoretically possible to transmit information reliably at a rate that is bounded by the channel capacity [4]. Computing the capacity involves finding the joint probability density function of the TM-dimensional transmitted signal that maximizes the mutual information between it and the TN-dimensional received signal. The special case $M = N = T = 1$ is addressed in [9], where it is shown that the maximizing transmitted signal density is discrete and has support only on the nonnegative real axis. The maximization appears, in general, to be computationally intractable for $M > 1$ or $T > 1$.

Nevertheless, we show that the dimensionality of the maximization can be reduced from $T \cdot M$ to $\min(M, T)$, and that the capacity can therefore be easily computed for many nontrivial cases. In the process, we determine the signal probability densities that achieve capacity and find the asymptotic dependences of the capacity on T. The signaling structures turn out to be surprisingly simple and provide practical insight into communicating over a multi-element link. Although we approach this communication problem with no training schemes in mind, as a by-product of our analysis we are able to provide an asymptotic upper bound on the number of channel uses that one could devote to training and still achieve capacity.

There are four main theorems proven in the paper that can be summarized as follows. Theorem 1 states that there is no point in making the number of transmitter antennas greater than T. Theorem 2 gives the general structure of the signals that achieves capacity. Theorem 3 derives the capacity, asymptotically in T, for $M = N = 1$. Theorem 4 gives the signal density that achieves capacity, asymptotically in T, for $M = N = 1$. Various implications and generalizations of the theorems are mentioned as well.

The following notation is used throughout the paper: $\log x$ is the base-two logarithm of x, while $\ln x$ is base e. Given a sequence $b_1, b_2, \ldots,$ of positive real numbers, we say that $a_n = O(b_n)$ as $n \to \infty$ if $|a_n|/b_n$ is bounded by some positive constant for sufficiently large n; we say that $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. The sequence δ_{mn} for integer m and n is defined to be one when $m = n$ and zero otherwise, and $\delta(z)$ is Dirac's δ-function, which, when z is complex, is defined as $\delta(z) \stackrel{\text{def}}{=} \delta(\operatorname{Re} z) \cdot \delta(\operatorname{Im} z)$. Two complex vectors, a and b, are *orthogonal* if $a^\dagger b = 0$, where the superscript † denotes "conjugate transpose." The mean-zero, unit-variance, circularly-symmetric, complex Gaussian distribution is denoted $\mathcal{CN}(0, 1)$.

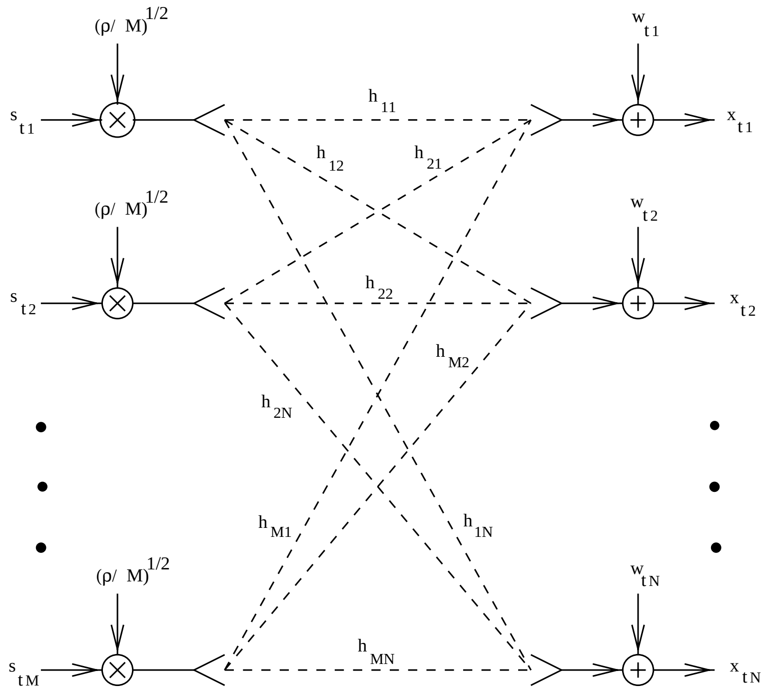

Figure 1: Wireless link comprising M transmitter and N receiver antennas. Every receiver antenna is connected to every transmitter antenna through an independent, random, unknown propagation coefficient having Rayleigh distributed magnitude and uniformly distributed phase. Normalization ensures that the total expected transmitted power is independent of M for a fixed ρ.

2 Multiple-Antenna Link

2.1 Signal model

Figure 1 displays a communication link or channel comprising M transmitter antennas and N receiver antennas that operates in a Rayleigh flat-fading environment. Each receiver antenna responds to each transmitter antenna through a statistically independent fading coefficient that is constant for T symbol periods. The fading coefficients are not known by either the transmitter or the receiver. The received signals are corrupted by additive noise that is statistically independent among the N receivers and the T symbol periods.

The complex-valued signal x_{tn} that is measured at receiver antenna n, and discrete time t, is given by

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

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

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

The complex-valued signal that is fed at time t into transmitter antenna m is denoted s_{tm}, and its average (over the M antennas) expected power is equal to one. This may be written

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

The additive noise at time t and receiver antenna n is denoted w_{tn}, and is independent (with respect to both t and n), identically distributed $\mathcal{CN}(0,1)$. The quantities in the signal model (1) are normalized so that ρ represents the expected signal-to-noise ratio (SNR) at each receiver antenna, independently of M. (It is easy to show that the channel capacity is unchanged if we replace the equality constraint in (2) with an upper bound constraint.) We later show that the constraint (2) can be strengthened or weakened in certain convenient ways without changing the channel capacity.

2.2 Conditional probability density

Both the fading coefficients and the receiver noise are complex Gaussian distributed. As a result, conditioned on the transmitted signals, the received signals are jointly complex Gaussian. Let

$$S = \begin{bmatrix} s_{11} & \dots & s_{1M} \\ \vdots & & \\ s_{T1} & \dots & s_{TM} \end{bmatrix}, \qquad X = \begin{bmatrix} x_{11} & \dots & x_{1N} \\ \vdots & & \\ x_{T1} & \dots & x_{TN} \end{bmatrix},$$

$$H = \begin{bmatrix} h_{11} & \dots & h_{1N} \\ \vdots & & \\ h_{M1} & \dots & h_{MN} \end{bmatrix}, \qquad W = \begin{bmatrix} w_{11} & \dots & w_{1N} \\ \vdots & & \\ w_{T1} & \dots & w_{TN} \end{bmatrix},$$

where S is the $T \times M$ matrix of transmitted signals, X is the $T \times N$ matrix of received signals, H is the $M \times N$ matrix of propagation coefficients, and W is the $T \times N$ matrix of additive noise components. Then

$$X = \sqrt{\frac{\rho}{M}}SH + W. \tag{3}$$

It is clear that

$$\mathrm{E}\left\{X \mid S\right\} = 0$$

and

$$\mathrm{E}\left\{x_{t_1 n_1} x^*_{t_2 n_2} \mid S\right\} = \delta_{n_1 n_2} \cdot \left[\delta_{t_1 t_2} + \left(\frac{\rho}{M}\right) \sum_{m=1}^{M} s_{t_1 m} s^*_{t_2 m}\right].$$

Thus, the conditional probability density of the received signals given the transmitted signals is

$$p(X \mid S) = \frac{\exp\left(-\mathrm{tr}\left\{\left[I_T + (\rho/M)\, SS^\dagger\right]^{-1} XX^\dagger\right\}\right)}{\pi^{TN} \det^N \left[I_T + (\rho/M)\, SS^\dagger\right]}, \tag{4}$$

where I_T denotes the $T \times T$ identity matrix and "tr" denotes "trace."

The channel is completely described by this conditional probability density. Note that the propagation coefficients do not appear in this expression. Although the received signals are conditionally Gaussian, the transmitted signals only affect the covariance of the received signals, in contrast to the classical additive Gaussian noise channel where the transmitted signals affect the mean of the received signals.

2.3 Special properties of the conditional probability

The conditional probability density of the received signals given the transmitted signals (4) has a number of special properties that are easy to verify.

Property 1 *The $T \times T$ matrix, $XX^\dagger$, is a sufficient statistic.*

When the number of receiver antennas is greater than the duration of the fading interval ($N > T$), then this sufficient statistic is a more economical representation of the received signals than the $T \times N$ matrix X.

Property 2 *The conditional probability density $p(X \mid S)$ depends on the transmitted signals S only through the $T \times T$ matrix $SS^\dagger$.*

Property 3 *For any $M \times M$ unitary matrix Ψ, $p(X \mid S\Psi^\dagger) = p(X \mid S)$.*

Property 4 *For any $T \times T$ unitary matrix Φ, $p(\Phi X \mid \Phi S) = p(X \mid S)$.*

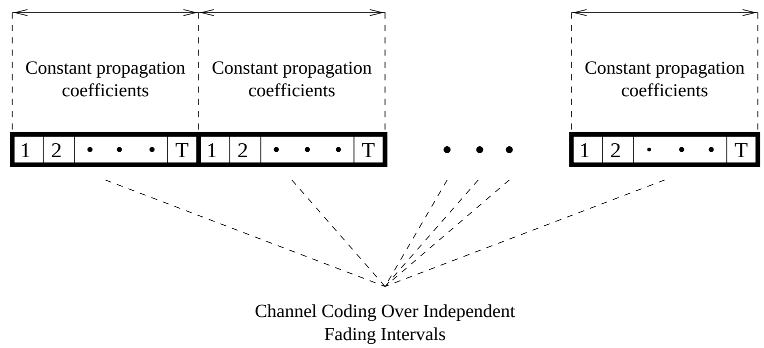

Figure 2: Propagation coefficients change randomly every T symbol periods. Channel coding is performed over multiple independent fading intervals.

3 Channel Coding Over Multiple Fading Intervals

We assume that the fading coefficients change to new independent realizations every T symbol periods. By performing channel coding over multiple fading intervals, as in Figure 2, the intervals of favorable fading compensate for the intervals of unfavorable fading.

Each channel use (consisting of a block of T transmitted symbols) is independent of every other, and (4) is the conditional probability density of the output X, given the input S. Thus, data can theoretically be transmitted reliably at any rate less than the channel capacity, where the capacity is the least upper bound on the mutual information between X and S, or

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

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

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

Thus, C is measured in bits per block of T symbols. We will often find it convenient to normalize C by dividing by T.

The next section uses (5) and the special properties of the conditional density (4) to derive some properties of the transmitted signals that achieve capacity.

4 Properties of Transmitted Signals that Achieve Capacity

Direct analytical or numerical maximization of the mutual information in (5) is hopelessly difficult whenever $T \cdot M$, the number of components of the transmitted signal matrix S, is much greater than one. This section shows that the maximization effort can be reduced to a problem in $\min(M, T)$ dimensions, making it possible to compute capacity easily for many significant cases.

Lemma 1 *(Invariance of $I(X;S)$ to rotations of S) Suppose that S has a probability density $p_0(S)$ that generates some mutual information I_0. Then, for any $M \times M$ unitary matrix Ψ and for any $T \times T$ unitary matrix Φ, the "rotated" probability density, $p_1(S) = p_0(\Phi^\dagger S \Psi)$, also generates I_0.*

Proof: We prove this result by substituting the rotated density $p_1(S)$ into (5); let I_1 be the mutual information thereby generated. Changing the variables of integration from S to $\Phi^\dagger S \Psi$, and from X to $\Phi^\dagger X$ (note that the Jacobian determinant of any unitary transformation is equal to one), and using Properties 3 and 4, we obtain

$$\begin{aligned} I_1 &= \int dS\, p_0(\Phi^\dagger S \Psi) \int dX\, p(X \mid S) \log \left\{ \frac{p(X \mid S)}{\int d\acute{S}\, p_0(\Phi^\dagger \acute{S} \Psi) p(X \mid \acute{S})} \right\} \\ &= \int dS\, p_0(S) \int dX\, p(\Phi X \mid \Phi S \Psi^\dagger) \log \left\{ \frac{p(\Phi X \mid \Phi S \Psi^\dagger)}{\int d\acute{S}\, p_0(\acute{S}) p(\Phi X \mid \Phi \acute{S} \Psi^\dagger)} \right\} \\ &= \int dS\, p_0(S) \int dX\, p(X \mid S) \log \left\{ \frac{p(X \mid S)}{\int d\acute{S}\, p_0(\acute{S}) p(X \mid \acute{S})} \right\} \\ &= I_0. \end{aligned}$$

□

Lemma 1 implies that we can interchange rows or columns of S—since this is equivalent to pre- or post-multiplying S by a permutation matrix—without changing the mutual information.

Lemma 2 *(Symmetrization of signaling density) For any transmitted signal probability density $p_0(S)$, there is a probability density $p_1(S)$ that generates at least as much mutual information and is unchanged by rearrangements of its arguments.*

Proof: There are $T!$ distinct permutations of the rows of S, and $M!$ distinct permutations of the columns. We let $p_1(S)$ be a mixture density involving all distinct permutations of the rows and columns, namely,

$$p_1(S) = \frac{1}{T!M!} \sum_{k=1}^{T!} \sum_{\ell=1}^{M!} p_0(P_{Tk}^\dagger S P_{M\ell}), \tag{6}$$

where $P_{Tk},\ k = 1, \ldots, T!$ are the $T \times T$ permutation matrices, and $P_{M\ell},\ \ell = 1, \ldots, M!$ are the $M \times M$ permutation matrices. Plainly, $p_1(S)$ is unchanged by rearrangements of its arguments. The concavity of mutual information as a functional of $p(S)$, Lemma 1, and Jensen's inequality imply that

$$I_1 \geq \frac{1}{T!M!} \sum_{k=1}^{T!} \sum_{\ell=1}^{M!} I(X; P_{Tk}^\dagger S P_{M\ell}) = \frac{1}{T!M!} \sum_{k=1}^{T!} \sum_{\ell=1}^{M!} I_0 = I_0.$$

□

Lemma 2 is consistent with one's intuition that all transmission paths are, on average, equally good. With respect to the mixture density $p_1(S)$ of (6), the expected power of the (tm)th component of S is

$$\int dS \frac{1}{T!M!} \sum_{k=1}^{T!} \sum_{\ell=1}^{M!} p_0(P_{Tk}^\dagger S P_{M\ell}) \cdot |s_{tm}|^2 = \frac{1}{T!M!} \sum_{k=1}^{T!} \sum_{\ell=1}^{M!} \int dS\, p_0(S) \cdot |[P_{Tk} S P_{M\ell}^\dagger]_{tm}|^2, \tag{7}$$

where we have substituted $P_{Tk}^\dagger S P_{M\ell}$ for S as the variable of integration. Over all possible permutations, $[P_{Tk} S P_{M\ell}^\dagger]_{tm}$ takes on the value of every S component $T!M!/(TM)$ times. Consequently, the expected power (7) becomes

$$\int dS\, p_1(S) \cdot |s_{tm}|^2 = \frac{1}{TM} \sum_{k=1}^{T} \sum_{\ell=1}^{M} \int dS\, p_0(S) \cdot |s_{k\ell}|^2 = 1$$

for all t and m, where the second equality follows from (2).

The constraint (2) requires that the expected power, spatially averaged over all antennas, be one at all times. As we have just seen, Lemma 2 implies that the same capacity is obtained by enforcing the stronger constraint that the expected power for each transmit element be one at all times. We obtain the following Corollary.

Corollary 1 *The following power constraints all yield the same channel capacity:*

a)

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

b)

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

c)

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

d)

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

The last condition is the weakest and says that, without changing capacity, one could impose the constraint that the expected power, averaged over both space and time, be one. This can equivalently be expressed as $(1/TM)\mathrm{E}\,\mathrm{tr}\,SS^\dagger = 1$.

4.1 Increasing number of transmitter antennas beyond T does not increase capacity

We observe in Property 2 that the effect of the transmitted signals on the conditional probability density of the received signals is through the $T \times T$ matrix $SS^\dagger$. It is therefore reasonable to expect that any possible joint probability density of the elements of $SS^\dagger$ can be realized with at most $M = T$ transmitter antennas.

Theorem 1 *(Capacity for $M > T$ equals capacity for $M = T$) For any coherence interval T and any number of receiver antennas, the capacity obtained with $M > T$ transmitter antennas is the same as the capacity obtained with $M = T$ transmitter antennas.*

Proof: Suppose that a particular joint probability density of the elements of $SS^\dagger$ achieves capacity with $M > T$ antennas. We can perform the Cholesky factorization $SS^\dagger = LL^\dagger$, where L is a $T \times T$ lower triangular matrix. Using T transmitter antennas, with a signal matrix that has the same joint probability density as the joint probability density of L, we may therefore also achieve the same probability density on $SS^\dagger$. If S satisfies power condition d) of Corollary 1, then so does L. $\square$

This result, for which we have no simple physical interpretation, contrasts sharply with the capacity obtained when the receiver knows the propagation coefficients, which grows linearly with $\min(M, N)$, independently of T; see [7] and Appendix C.

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

4.2 Structure of signal that achieves capacity

In this section, we will be concerned with proving the following theorem.

Theorem 2 *(Structure of signal that achieves capacity) The signal matrix that achieves capacity can be written as $S = \Phi V$, where Φ is an $T \times T$ isotropically distributed unitary matrix, and V is an independent $T \times M$ real, nonnegative, diagonal matrix. Furthermore, we can choose the joint density of the diagonal elements of V to be unchanged by rearrangements of its arguments.*

In calling the oblong matrix V diagonal, we mean that only the elements along its main diagonal may be nonzero. An *isotropically distributed* unitary matrix has a probability density that is unchanged when the matrix is multiplied by any deterministic unitary matrix. In a natural way, an isotropically distributed unitary matrix is the $T \times T$ counterpart of a complex scalar having unit magnitude and uniformly distributed phase. More details, including the probability density of these matrices, may be found in Appendix A. The theorem relies on the following lemma, which is proven first.

Lemma 3 *(Singular value decomposition of S) Suppose that S, with singular value decomposition $S = \Phi V \Psi^\dagger$, has an arbitrary distribution that generates some mutual information I_0. Then the signal matrix formed from the first two factors, $S_1 = \Phi V$, also generates I_0.*

Proof: The singular value decomposition (SVD) says that the $T \times M$ signal matrix can always be decomposed into the product of three jointly distributed random matrices, $S = \Phi V \Psi^\dagger$, where Φ is a $T \times T$ unitary matrix, V is a $T \times M$ nonnegative real matrix whose only nonzero elements are on the main diagonal, and Ψ is an $M \times M$ unitary matrix.

We write the mutual information I_0 in terms of the three SVD factors and then apply Property 3 to obtain

$$\begin{aligned} I_0 &= \int d\Phi\, dV\, d\Psi\, p(\Phi, V, \Psi) \int dX\, p(X \mid \Phi V \Psi^\dagger) \\ &\quad \cdot \log \left\{ \frac{p(X \mid \Phi V \Psi^\dagger)}{\int d\acute{\Phi}\, d\acute{V}\, d\acute{\Psi}\, p(\acute{\Phi}, \acute{V}, \acute{\Psi}) p(X \mid \acute{\Phi}\acute{V}\acute{\Psi}^\dagger)} \right\} \\ &= \int d\Phi\, dV\, d\Psi\, p(\Phi, V, \Psi) \int dX\, p(X \mid \Phi V) \\ &\quad \cdot \log \left\{ \frac{p(X \mid \Phi V)}{\int d\acute{\Phi}\, d\acute{V}\, d\acute{\Psi}\, p(\acute{\Phi}, \acute{V}, \acute{\Psi}) p(X \mid \acute{\Phi}\acute{V})} \right\} \end{aligned}$$

$$= \int d\Phi\, dV\, p(\Phi, V) \int dX\, p(X \mid \Phi V) \log \left\{ \frac{p(X \mid \Phi V)}{\int d\acute{\Phi}\, d\acute{V}\, p(\acute{\Phi}, \acute{V}) p(X \mid \acute{\Phi}\acute{V})} \right\},$$

where the last expression is immediately recognized as the mutual information generated by $S_1 = \Phi V$. Finally, if S satisfies power constraint d) of Corollary 1, then so does S_1. □

Ostensibly, maximizing the mutual information with respect to the joint probability density of Φ and V is even more difficult than the problem that it replaces. However, as we now show, capacity can be achieved by making Φ and V independent, with Φ isotropically distributed.

Proof of Theorem 2: Using Lemma 3, we write the transmitted signal matrix as $S = \Phi V$, where Φ and V are jointly distributed, Φ is unitary, and V is diagonal, nonnegative and real. Suppose S has probability density $p(S)$ and generates mutual information I_0. Let Θ be an isotropically distributed unitary matrix that is statistically independent of Φ and V, and define a new signal matrix, $S_1 = \Theta S$, generating mutual information I_1. It follows from Lemma 1 that, conditioned on Θ, the mutual information generated by S_1 equals I_0. The concavity of mutual information as a functional of $p(S)$, and Jensen's inequality, then imply that $I_1 \geq I_0$.

From the definition of an isotropically distributed unitary matrix (see Appendix A), the product $\Theta\Phi$, conditioned on Φ, is also isotropically distributed. Since the conditional probability density does not depend on Φ, it follows that the product is independent of Φ and V. Consequently S_1 is equal to the product of an isotropically distributed unitary matrix and V, with the two matrices statistically independent. If S satisfies power condition d) of Corollary 1, then so does S_1.

The expression for mutual information (5) becomes

$$I(X;S) = \int dV\, p(V) \int d\Phi\, p(\Phi) \int dX\, p(X \mid \Phi V) \log \left\{ \frac{p(X \mid \Phi V)}{\int d\acute{V}\, p(\acute{V}) \int d\acute{\Phi}\, p(\acute{\Phi}) p(X \mid \acute{\Phi}\acute{V})} \right\}, \qquad (8)$$

where $p(V)$ is the probability density of the diagonal elements of V. The probability density $p(\Phi)$ is given in (A.5), and the maximization of the mutual information $I(X;S)$ needed to calculate capacity now takes place only with respect to $p(V)$. The mutual information is a concave functional of $p(V)$, because it is concave in $p(S)$ and $p(S)$ is linear in $p(V)$.

The conclusion that there exists a capacity-achieving joint density on V that is unchanged by rearrangements of its arguments does not follow automatically from Lemma 2, because the symmetry of the signal that achieves capacity in Lemma 2 does not obviously survive the above dropping of the right-hand SVD factor and premultiplication by an isotropically distributed unitary matrix. Nevertheless, we follow some of the same techniques presented in the proof of Lemma 2.

There are $M!$ ways of arranging the diagonal elements of V, each corresponding to pre- and post-multiplying V by appropriate permutation matrices, say P_{Tk} and $P_{Mk},\ k = 1, \ldots, M!$. The permutation does not change the mutual information; this can be verified by plugging the reordered V into (8), substituting $P_{Tk}^{\dagger} V P_{Mk}$ for V, and ΦP_{Tk} for Φ, as variables of integration, and then using Property 3 and the fact that multiplying Φ by a permutation matrix does not change its probability density. Now, as in the proof of Lemma 2, using an equally-weighted mixture density for V, involving all $M!$ arrangements, and exploiting the concavity of $I(X;S)$ as a functional of $p(V)$, we conclude that the mutual information for the mixture density is at least as large as the mutual information for the original density. But, clearly, this mixture density is invariant to rearrangements of its arguments. □

We remark that the mixture density in the above proof symmetrizes the probability density for S. Hence, $\mathrm{E}\,|s_{tm}|^2 = 1$ for all t and m. Let $v_1, \ldots, v_M$ denote the diagonal elements of V (recall that $M \leq T$). Then

$$1 = \mathrm{E}\,|s_{tm}|^2 = \mathrm{E}\,|[\Phi V]_{tm}|^2 = \mathrm{E}\,|\phi_{tm}|^2 \cdot \mathrm{E}\,v_m^2 = \frac{1}{T} \cdot \mathrm{E}\,v_m^2,$$

where the last equality is a consequence of (A.1); therefore $\mathrm{E}\,v_m^2 = T$ for $m = 1, \ldots, M$.

Thus, the problem of maximizing $I(X;S)$ with respect to the joint probability density of the $T \cdot M$ complex elements of S reduces to the simpler problem of maximizing $I(X;S)$ with respect to the joint probability density of the M nonnegative real diagonal elements of V. This joint probability can be constrained to be invariant to rearrangements of its arguments, and thus the marginal densities on $v_1, \ldots, v_M$ can be made identical, with $\mathrm{E}\,v_1^2 = \ldots = \mathrm{E}\,v_M^2 = T$. But we do not know if $v_1, \ldots, v_M$ are independent.

The mth column of S, representing the T complex signals that are fed into the mth transmitter antenna, is equal to the real nonnegative scalar v_m times an independent T-dimensional isotropically distributed complex unit vector ϕ_m. Since ϕ_m is the mth column of the $T \times T$ isotropically distributed unitary matrix Φ, the M signal vectors $v_1\phi_1, \ldots, v_M\phi_M$ are mutually orthogonal. Figure 3 shows the signal vectors associated with the M transmitter antennas. Each signal vector is a T-dimensional complex vector (comprising $2T$ real components). The solid sphere demarcates the root mean-square values of the vector lengths; that is, $\mathrm{E}\,v_m^2 = T$. Later we argue that, for $T \gg M$, the magnitudes of the signal vectors are approximately $\sqrt{T}$ with very high probability.

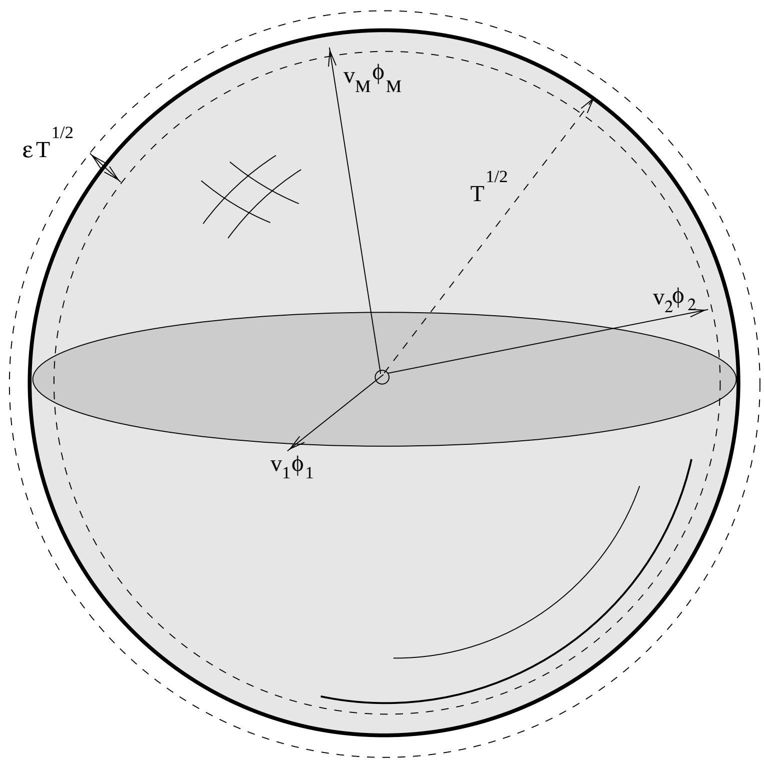

Figure 3: The transmitted signals that achieve capacity are mutually orthogonal with respect to time. The constituent orthonormal unit vectors are isotropically distributed (see Appendix A), and independent of the signal magnitudes, which have mean-square value T. The solid sphere of radius $T^{1/2}$ demarcates the root mean-square. For $T \gg M$, the vectors all lie approximately on the surface of this sphere. The shell of thickness $\varepsilon T^{1/2}$ is discussed in Section 5.

5 Capacity and Capacity Bounds

The simplification provided by Theorem 2 allows us to compute capacity easily for many cases of interest. The mutual information expression (8) requires integrations with respect to Φ, X, and the diagonal elements of V. Although the maximization of (8) is only over the M diagonal elements of V, the dimensionality of integration is still high. We reduce this dimensionality in Appendix B, resulting in the expression (B.10). Integration over the TN complex components of X is reduced to integration over $\min(N,T)$ real eigenvalues, $\{\lambda_1,\ldots,\lambda_{\min(N,T)}\}$, and integration over the T^2 complex elements of Φ is reduced to $M\cdot\min(N,T)$ complex elements. In fact, as we show in Appendix B, closed-form expressions for the integral over Φ can sometimes be obtained.

In this section, we calculate the capacity in some simple but nontrivial cases that sometimes require optimization of a scalar probability density. Where needed, any numerical optimization was performed using the Blahut-Arimoto algorithm [1]. Where instructive, we also include upper and lower bounds on capacity.

5.1 Capacity upper bound

An upper bound is obtained if we assume that the receiver is provided with a noise-free measurement of the propagation coefficients H. This *perfect-knowledge* upper bound, obtained under power constraint a) of Corollary 1, is

$$C_u = T\cdot \mathrm{E}\,\log\det\left[I_N+\frac{\rho}{M}H^\dagger H\right], \tag{9}$$

and is derived in Appendix C. Equation (9) gives the upper bound per block of T symbols. The normalized bound, C_u/T, is independent of T. When H is known to the receiver, the perfect-knowledge capacity bound is achieved with transmitted signals S that are independent $\mathcal{CN}(0,1)$ (see also [7]).

In our model H is unknown to the receiver, but we intuitively expect C to approach C_u as T becomes large because a small portion of the coherence interval can be reserved for sending training data from which the receiver can estimate the propagation coefficients. When H is unknown and T is large, we therefore also expect the joint probability density of the capacity-achieving transmitted signals S to become approximately independent $\mathcal{CN}(0,1)$. We reconcile this intuition with the structure that is demanded in Theorem 2,

$$S=\Phi V=\left[\begin{array}{cccc} v_1\phi_1 & v_2\phi_2 & \ldots & v_M\phi_M\end{array}\right],$$

where $\{\phi_m\}_{m=1}^{M}$ are the column vectors of Φ, by observing that, for fixed M, as T grows large two interesting things happen. First, the M complex random orthogonal unit vectors that comprise S become virtually independent. Second, for any $\varepsilon > 0$, the magnitude of a vector of T independent $\mathcal{CN}(0,1)$ random variables is contained in a shell of radius $\sqrt{T}$ and width $\varepsilon\sqrt{T}$ with probability that approaches one as $T \to \infty$ (Figure 3 displays this so-called sphere-hardening phenomenon).

Hence, the ΦV and $\mathcal{CN}(0,1)$ structures for S are reconciled if $v_1 = \cdots = v_M = \sqrt{T}$ with high probability as $T \to \infty$ (see also Appendix A for a method to generate Φ from a matrix of $\mathcal{CN}(0,1)$ random variables). This intuition is formalized in Section 5.3.

5.2 Capacity lower bound

By substituting an arbitrary density for $p(V)$ in (B.10), one obtains a lower bound on capacity. The arguments of Section 5.1 suggest that by assigning unit probability mass to $v_1 = \cdots = v_M = \sqrt{T}$, one should obtain a lower bound that becomes tight as $T \to \infty$. The result is

$$\begin{aligned} C_l \stackrel{\text{def}}{=} & -TN\log e - NM\log\Big(1+\frac{\rho T}{M}\Big) \\ & - \int d\lambda \cdot p(\lambda)\cdot f_l(\lambda)\cdot\Big[\log f_l(\lambda) - (\log e)\cdot\sum_{\ell=1}^{\min(N,T)}\lambda_\ell\Big], \end{aligned} \tag{10}$$

where $p(\lambda)$ is given by (B.8), and $f_l(\lambda)$ is obtained by evaluating (B.6) with $v_1 = \cdots = v_M = \sqrt{T}$, yielding

$$f_l(\lambda) = \left(1+\frac{\rho T}{M}\right)^{-NM}\int d\Phi\, p(\Phi)\cdot\exp\left\{\sum_{n=1}^{\min(N,T)}\sum_{m=1}^{M}\lambda_n\cdot\left(\frac{\rho T}{M+\rho T}\right)\cdot|\phi_{nm}|^2\right\}. \tag{11}$$

By the reasoning of Section 5.1, this lower bound should be most useful when $T \gg M$, and least useful when $M \approx T$. In fact, when $M = T$, it is easy to show that assigning unit mass to $v_1 = \cdots = v_M = \sqrt{T}$ implies that $SS^\dagger = T\cdot\Phi\Phi^\dagger = T\cdot I_T$, so $p(X \mid S) = p(X)$ and $C_l = 0$.

When $M = N = 1$, the integration over Φ in (11) can be performed analytically as shown in Appendix B.1, and (10) becomes

$$\begin{aligned} & C_l = -T\log e - \log(1+\rho T) \\ & -\int_0^\infty \frac{(T-1)e^{-\lambda/(1+\rho T)}\gamma\Big(T-1,\frac{\rho T\lambda}{1+\rho T}\Big)}{\Gamma(T)(1+\rho T)\left[\frac{\rho T}{1+\rho T}\right]^{T-1}}\log\left[\frac{(T-1)e^{-\lambda/(1+\rho T)}\gamma\Big(T-1,\frac{\rho T\lambda}{1+\rho T}\Big)}{(1+\rho T)\left[\frac{\rho T\lambda}{1+\rho T}\right]^{T-1}}\right]d\lambda, \end{aligned} \tag{12}$$

where $\gamma(T,z) \stackrel{\text{def}}{=} \int_0^z q^{T-1}e^{-q}dq$ is the *incomplete gamma* function (see also (B.3)). We now justify the use of (10) for large T by proving that (12) is a tight lower bound as $T \to \infty$.

5.3 Capacity, asymptotically in T

For $M = N = 1$ the perfect-knowledge capacity upper bound (9) is

$$C_u = T \cdot \mathrm{E}\ \log\Big(1 + \rho|h_{11}|^2\Big) = T\int_0^\infty dy\ e^{-y}\log\Big(1+\rho y\Big) = T(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho),$$

where

$$\mathrm{E}_1(x) = \int_x^\infty \frac{e^{-y}}{y}\,dy \tag{13}$$

is the *exponential integral*. Hence, for $M = N = 1$, we expect $C/T \to (\log e)e^{1/\rho}\mathrm{E}_1(1/\rho)$ as $T \to \infty$. This intuition is made precise in the following theorem.

Theorem 3 *(Capacity, asymptotically in T) Let $M = N = 1$. Then*

$$(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho) - O\Big(\sqrt{\frac{\log T}{T}}\Big) = C_l/T \le C/T \le C_u/T = (\log e)e^{1/\rho}\mathrm{E}_1(1/\rho)$$

as $T \to \infty$.

Proof: It remains to show that C_l/T approaches C_u/T at the indicated rate, and this is proven in Appendix D.

The remainder term $O(\sqrt{(\log T)/T})$ can be viewed as a penalty for having to learn h_{11} at the receiver. One possible way to learn h_{11} is to have the transmitter send, say, τ training symbols per block that are known to the receiver. Clearly, when the transmitter sends a training symbol, no message information is sent. Even if h_{11} is thereby learned perfectly at the receiver, the remaining $T - \tau$ symbols cannot communicate more than $(T-\tau)(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho)$ bits. We therefore have the following corollary.

Corollary 2 *Let $M = N = 1$. Of the T symbols transmitted over the link per block, one cannot devote more than $O(\sqrt{T\log T})$ to training and still achieve capacity, as $T \to \infty$.*

Since we have shown that, for large T, the capacity of our communication link approaches the perfect-knowledge upper bound, we also expect the joint probability density of the elements of $S = \Phi V$ to become approximately independent $\mathcal{CN}(0,1)$ as part of the sphere-hardening phenomenon described in Section 5.1. This is the content of the next theorem.

Theorem 4 *(Distribution that achieves capacity, asymptotically in T) Let $M = N = 1$. For the signal that achieves capacity, $v_1/\sqrt{T}$ converges in distribution to a unit mass at 1, as $T \to \infty$.*

Proof: See Appendix E.

When $M > 1$, it is reasonable to expect that the joint distribution of $v_1, \ldots, v_M$ becomes a unit mass at $v_1 = v_2 = \ldots = v_M = \sqrt{T}$ (see Figure 3) as $T \to \infty$, but we do not include a formal proof. Consequently, when $T \gg M$ the diagonal components of V that yield capacity should all be $\sqrt{T}$, and C_l given by (10) should be a tight bound. Furthermore, as $T \to \infty$, C_l and, hence, the capacity, should approach the perfect-knowledge upper bound.

5.4 Capacity and capacity bounds for $M = N = 1$ and $T \geq 1$

There is a single transmitter antenna, and the transmitted signal is $s_t = v_1 \cdot \phi_{t1},\ t = 1, \ldots, T$, where ϕ_{t1} is the tth element of the isotropically distributed unit vector ϕ_1. Figures 4–6 display the capacity, along with the perfect-knowledge upper bound (9), and lower bound (12), (all normalized by T) as functions of T for three SNR's. The optimum probability density of v_1, obtained from the Blahut-Arimoto algorithm and yielding the solid capacity curve, turns out to be discrete; as T becomes sufficiently large, the discrete points become a single mass at $\sqrt{T}$, and the lower bound and capacity coincide exactly. For still greater T, the capacity approaches the perfect-knowledge upper bound. This observed behavior is consistent with Theorems 3 and 4. The effect of increasing SNR is to accelerate the convergence of the capacity with its upper and lower bounds.

When $T = 1$, all of the transmitted information is contained in the magnitude of the signal, which is v_1. When T is large enough so that $C = C_l$, then all of the information is contained in ϕ_1, which is, of course, completely specified by its direction.

5.5 Capacity and capacity bounds for $M \geq 1$, $N \geq 1$, and $T = 1$

In this case, there are multiple transmitter and receiver antennas and the coherence interval is $T = 1$, corresponding to a very rapidly changing channel. Theorem 1 implies that the capacity for $M \geq T$ is the same as for $M = T$, so we assume, for computational purposes, that $M = 1$. The transmitted signal is then a complex scalar having a uniformly distributed phase and a magnitude that has a discrete probability density.

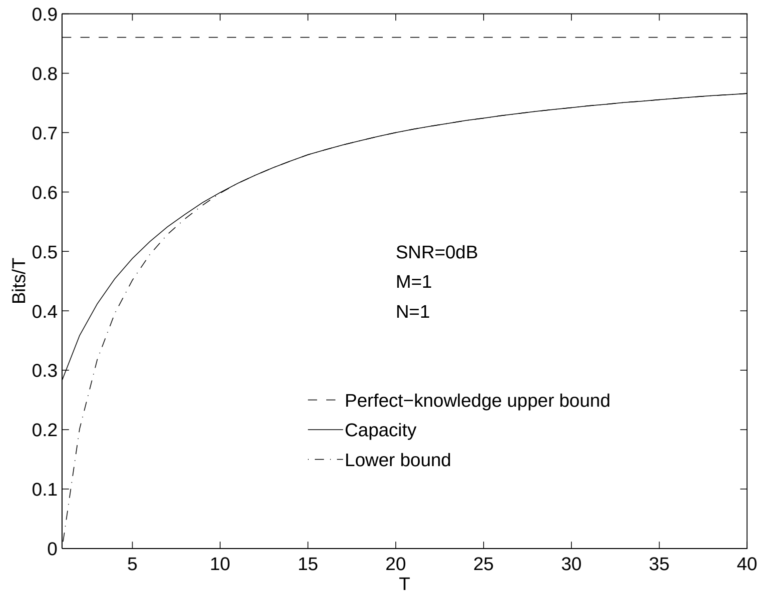

Figure 4: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR=0dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 12$. As per Theorem 3, the capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

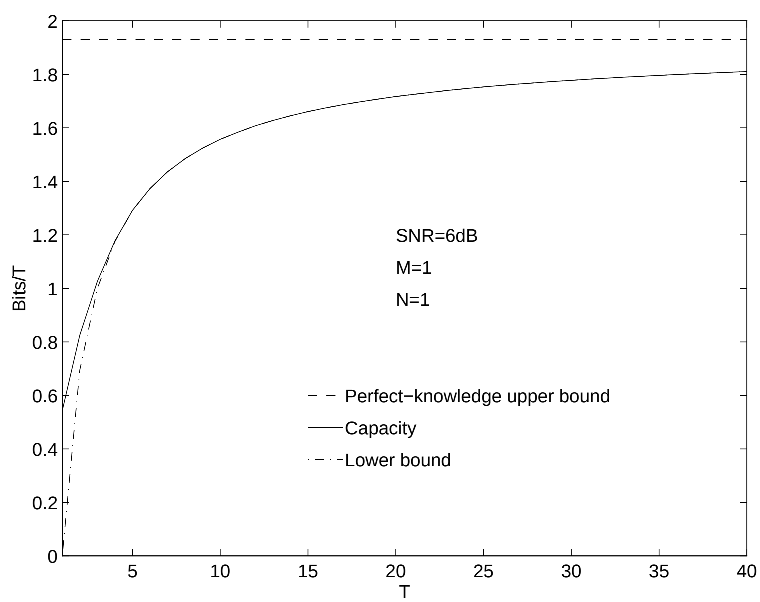

Figure 5: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR=6dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 4$. The capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

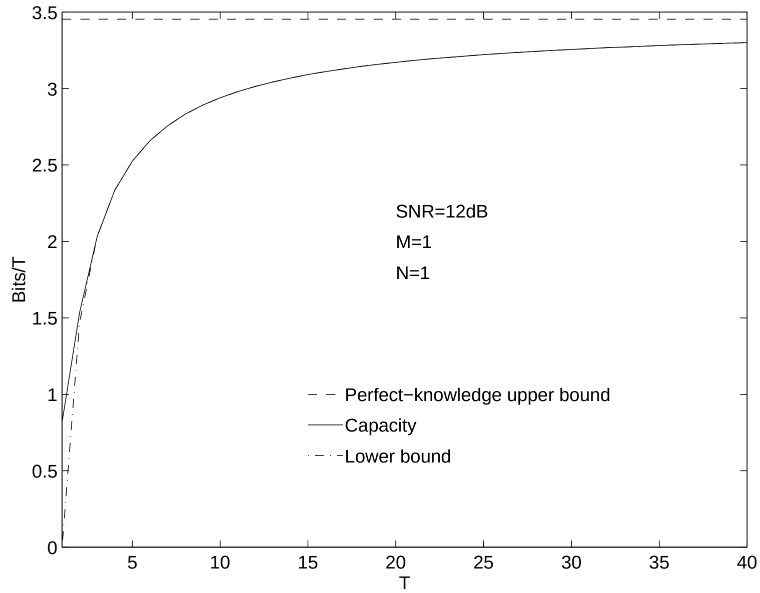

Figure 6: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR=12dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 3$. The capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

All of the transmitted information is contained in the magnitude. Because $M = T$, the capacity lower bound is trivially zero. The receiver cannot estimate the propagation coefficients reliably, so the capacity is far less than the perfect-knowledge upper bound.

Figure 7 displays the capacity as a function of N for arbitrary M. In the rapidly fading channel the difference between the capacity and the perfect-knowledge upper bound becomes especially dramatic as M and N both increase since, as shown in Appendix C (see also [7]), the upper bound grows approximately linearly with the minimum of M and N, while in Figure 7 the growth of capacity with N (recall that the capacity in this example does not grow with M) is very moderate and appears to be logarithmic.

5.6 Capacity bounds for $M \leq 20$, $N = 1$, and $T = 100$

In this case there are multiple transmitter antennas, a single receiver antenna, and the coherence interval is $T = 100$ symbols. Figure 8 illustrates the utility of the upper and lower capacity bounds, C_u and C_l, since it becomes cumbersome to compute capacity directly for large values of M.

We argue in Section 5.2 that C_l in equation (10) is most useful when $T \gg M$ since $C_l = 0$ when $M = T$. We see in Figure 8 that C_l/T peaks at $M = 3$. Nevertheless, the peak value of the lower bound,

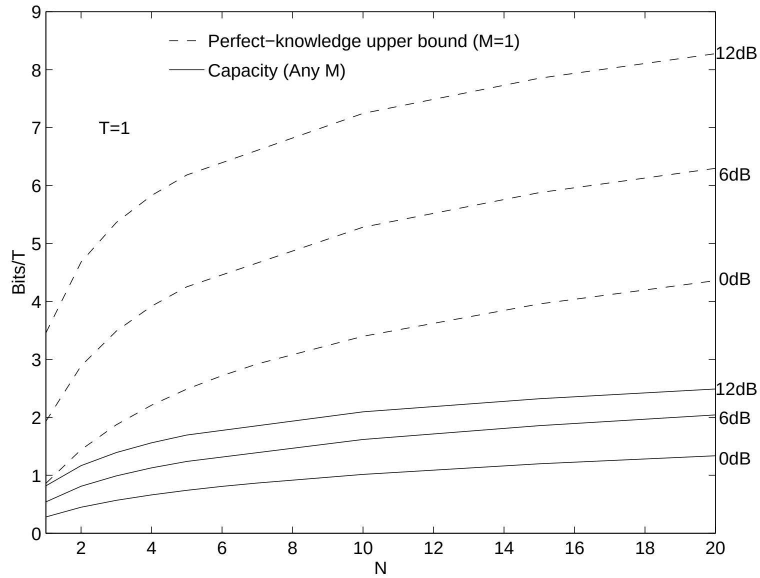

Figure 7: Capacity and perfect-knowledge upper bound versus number of receiver antennas N (SNR=0dB, 6dB, 12dB, arbitrary number of transmitter antennas, coherence interval equal to one). The gap between capacity and upper bound only widens as $N \to \infty$.

approximately 6.1 bits/T, remains a valid lower bound on capacity for all $M > 3$. (One could always ignore all but three of the transmitter antennas.) This gives us the modified lower bound also displayed in the figure. The uppermost dashed line is the limit of the perfect-knowledge upper bound as $M \to \infty$.

6 Conclusions

We have taken a fresh look at the problem of communicating over a flat-fading channel using multiple-antenna arrays. No knowledge about the propagation coefficients and no ad hoc training schemes were assumed. Three key findings emerged from our research.

First, there is no point in making the number of transmitter antennas greater than the length of the coherence interval. In a very real sense, the ultimate capacity of a multiple-antenna wireless link is determined by the number of symbol periods between fades. This is somewhat disappointing since it severely limits the ultimate capacity of a rapidly fading channel. For example, in the extreme case where a fresh fade occurs every symbol period, only one transmitter antenna can be usefully employed. Strictly speaking, one could increase capacity indefinitely by employing a large number of receiver antennas, but the capacity appears to increase only logarithmically in this number—not a very effective way to boost capacity.

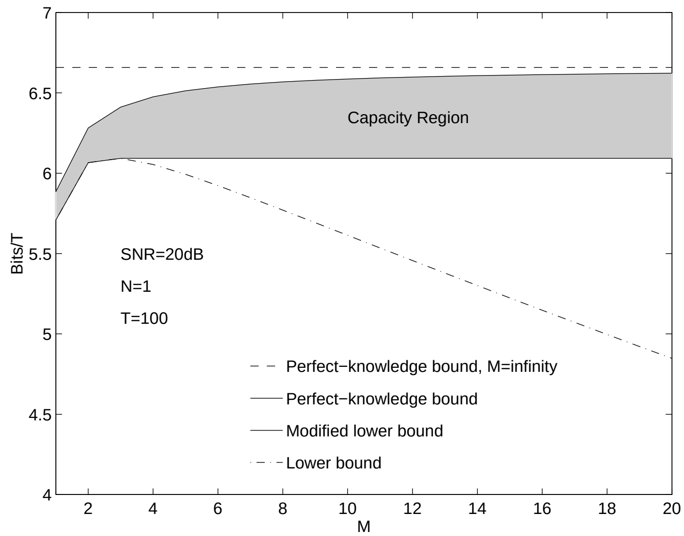

Figure 8: Normalized capacity lower bounds and perfect-knowledge upper bounds versus number of transmitter antennas M (SNR=20dB, one receiver antenna, coherence interval equal to 100). The actual channel capacity lies in the shaded region. Lower bound peaks at $M = 3$; this peak is a valid lower bound for $M \geq 3$, giving us the modified lower bound.

Second, the transmitted signals that achieve capacity are mutually orthogonal with respect to time among the transmitter antennas. The constituent orthonormal unit vectors are isotropically distributed and statistically independent of the signal magnitudes. This result provides insight for the design of efficient signaling schemes, and it greatly simplifies the task of determining capacity, since the dimensionality of the optimization problem is equal only to the number of transmitter antennas.

Third, when the coherence interval becomes large compared with the number of transmitter antennas, the normalized capacity approaches the capacity obtained as if the receiver knew the propagation coefficients. The magnitudes of the time-orthogonal signal vectors become constants that are equal for all transmitter antennas. In this regime, all of the signaling information is contained in the directions of the random orthogonal vectors, the receiver learns the propagation coefficients, and the channel becomes similar to the classical Gaussian channel.

We have computed capacity and upper and lower bounds for some nontrivial cases of interest. Clearly our methods can be extended to many others. The methods require an optimization on the order of the number of transmitter antennas. Hence, we are still hard pressed to compute the capacity, for example, when there are fifty transmitter and receiver antennas and the coherence interval is fifty symbols. Further work in simplifying such a computation is a possible next step.

Acknowledgement

We thank L. Shepp for helping us find closed-form expressions for the integrals in Appendix B.1. We are also grateful for the helpful comments of J. Mazo, R. Urbanke, E. Telatar, and especially the late A. Wyner.

A Appendix: Isotropically Distributed Unit Vectors and Unitary Matrices

Random unit vectors and unitary matrices figure extensively in this research. This appendix summarizes their key properties.

A.1 Isotropically distributed unit vectors

The intuitive idea of an isotropically distributed (i.d.) complex unit vector is that it is equally likely to point in any direction in complex space. Equivalently, multiplying such a vector by any deterministic unitary matrix results in a random unit vector that has exactly the same probability density function. We define a T-dimensional complex random unit vector ϕ to be i.d. if its probability density is invariant to all unitary transformations; that is,

$$p(\phi) = p(\Theta^\dagger \phi), \quad \forall\, \Theta : \; \Theta^\dagger \Theta = I.$$

This property implies that the probability density depends on the magnitude but not the direction of ϕ, so $p(\phi) = f(\phi^\dagger \phi)$, for some nonnegative function $f(\cdot)$. The fact that the magnitude of ϕ must equal one leads directly to the required probability density:

$$p(\phi) = f(1) \cdot \delta(\phi^\dagger \phi - 1).$$

The constant $f(1)$ is such that the integral of $p(\phi)$ over T-dimensional complex space is unity. Thus, $f(1) = \Gamma(T)/\pi^T$, and

$$p(\phi) = \frac{\Gamma(T)}{\pi^T} \cdot \delta(\phi^\dagger \phi - 1).$$

Successively integrating the probability density gives the joint probability density of any L of the elements of ϕ; denoting the L-dimensional vector by $\phi^{(L)}$, we obtain

$$p(\phi^{(L)}) = \frac{\Gamma(T)}{\pi^L \cdot \Gamma(T-L)} \cdot \left(1 - \phi^{(L)\dagger}\phi^{(L)}\right)^{T-1-L}, \qquad L = 1, \ldots, T-1. \tag{A.1}$$

An i.d. unit vector ϕ can be conveniently generated by letting z be a T-dimensional vector of independent $\mathcal{CN}(0,1)$ random variables, and $\phi = z/\sqrt{z^\dagger z}$.

A.2 Isotropically distributed unitary matrices

We define a $T \times T$ unitary matrix Φ to be i.d. if its probability density is unchanged when premultiplied by a deterministic unitary matrix, or

$$p(\Phi) = p(\Theta^\dagger \Phi), \quad \forall \Theta : \ \Theta^\dagger \Theta = I. \tag{A.2}$$

The real-valued counterpart to this distribution is sometimes called "random orthogonal" or "Haar measure" [6].

Multiplying any deterministic unit vector by an i.d. unitary matrix results in an i.d. unit vector. To see this, let Φ be an i.d. unitary matrix, and let a be a deterministic unit vector. Then $b = \Phi a$ is a random unit vector. Multiplying b by a deterministic unitary matrix Θ gives $\Theta^\dagger b = \Theta^\dagger \Phi a = \acute{\Phi} a$. But, by definition, $\acute{\Phi}$ has the same probability density as Φ, so $\Theta^\dagger b$ has the same probability density as b. Therefore b is an i.d. unit vector.

The column vectors of Φ are themselves i.d. unit vectors. However, because they are orthogonal they are statistically dependent. The joint probability density of the first two columns is $p(\phi_1, \phi_2) = p(\phi_1) \cdot p(\phi_2 \mid \phi_1)$, and (A.2) implies that $p(\Theta^\dagger \phi_2 \mid \Theta^\dagger \phi_1) = p(\phi_2 \mid \phi_1)$ for all unitary Θ. Now condition on ϕ_1, and let the first column of Θ be equal to ϕ_1; then

$$p(\phi_2 \mid \phi_1) = p(\Theta^\dagger \phi_2 \mid \Theta^\dagger \phi_1) = p(\Theta^\dagger \phi_2 \mid \phi_1 = [1, 0, \ldots, 0]). \tag{A.3}$$

The last $T-1$ columns of Θ can be chosen arbitrarily with the constraint that they are orthogonal to the first column and to each other. This imparts an arbitrary direction to the vector consisting of the last $T-1$ elements of $\Theta^\dagger \phi_2$. Therefore the product, $z = \Theta^\dagger \phi_2$ is a random unit vector, which, when conditioned on ϕ_1, has first component equal to zero and last $T-1$ components comprising a $(T-1)$-dimensional i.d.

vector. Hence,

$$p(z \mid \phi_1 = [1, 0, \ldots, 0]) = \delta(z_1) \cdot \frac{\Gamma(T-1)}{\pi^{T-1}} \cdot \delta(z^\dagger z - |z_1|^2 - 1) = \delta(z_1) \cdot \frac{\Gamma(T-1)}{\pi^{T-1}} \cdot \delta(z^\dagger z - 1), \quad \text{(A.4)}$$

where z_1 is the first component of z. Substituting $\Theta^\dagger \phi_2$ for z in (A.4) and using (A.3) give the desired conditional probability density,

$$p(\phi_2 \mid \phi_1) = \frac{\Gamma(T-1)}{\pi^{T-1}} \cdot \delta(\phi_2^\dagger \phi_2 - 1) \cdot \delta(\phi_1^\dagger \phi_2).$$

Continuing in this fashion we obtain the probability density of Φ in nested form,

$$\begin{aligned} p(\Phi) \quad = \quad & \left[\frac{\Gamma(T)}{\pi^T} \cdot \delta(\phi_1^\dagger \phi_1 - 1) \right] \cdot \left[\frac{\Gamma(T-1)}{\pi^{T-1}} \cdot \delta(\phi_2^\dagger \phi_2 - 1) \cdot \delta(\phi_1^\dagger \phi_2) \right] \\ & \cdots \left[\frac{\Gamma(1)}{\pi} \cdot \delta(\phi_T^\dagger \phi_T - 1) \cdot \delta(\phi_1^\dagger \phi_T) \cdots \delta(\phi_{T-1}^\dagger \phi_T) \right]. \end{aligned} \quad \text{(A.5)}$$

If Φ is an i.d. unitary matrix then the transpose is also i.d., that is, the formula for the probability density of Φ is unchanged if Φ is postmultiplied by a unitary matrix. To demonstrate this, let $\Psi = \Phi\Upsilon$, where Υ is a deterministic unitary matrix. Then Ψ is clearly unitary, and we need to show that it is i.d. Premultiplying Ψ by still another deterministic unitary matrix Θ gives $\Theta^\dagger \Psi = \Theta^\dagger \Phi \Upsilon = \acute{\Phi} \Upsilon$. Both $\acute{\Phi}$ and Φ are i.d., so $\Theta^\dagger \Psi = \acute{\Phi} \Upsilon$ has the same probability density as $\Psi = \Phi\Upsilon$, so Ψ is i.d.

A straightforward way to generate an i.d. unitary matrix is first to generate a $T \times T$ random matrix Y whose elements are independent $\mathcal{CN}(0,1)$, and then perform the QR (Gram-Schmidt) factorization $Y = \Phi R$, where Φ is unitary and R is upper triangular, yielding $\Phi = Y R^{-1}$. The triangular matrix is a function of Y and, in fact, depends only on the inner products between the columns of Y. We denote this explicit dependence on Y with a subscript, $R_Y = R$. To show that Φ is i.d., we premultiply by an arbitrary unitary matrix to get

$$\Theta^\dagger \Phi = \Theta^\dagger Y R_Y^{-1} = \Theta^\dagger Y R_{\Theta^\dagger Y}^{-1} = Z R_Z^{-1},$$

where $Z = \Theta^\dagger Y$. But because Z has the same probability density as Y it follows that $\Theta^\dagger \Phi$ has the same probability density as Φ, and therefore Φ is i.d.

B Appendix: Simplified Expression for $I(X;S)$

The representation of the transmitted signal matrix $S = \Phi V$ in Theorem 2 does not automatically lead to an easily computed expression for mutual information; one must still integrate with respect to both Φ and X in (8). Some simplification is both necessary and possible.

B.1 Integrating with respect to Φ

Consider first the conditional covariance matrix appearing in (4),

$$I_T + \frac{\rho}{M} \cdot SS^\dagger = I_T + \frac{\rho}{M} \cdot \Phi V V^\dagger \Phi^\dagger = \Phi \left(I_T + \frac{\rho}{M} \cdot V V^\dagger \right) \Phi^\dagger.$$

We represent V, which has dimensions $T \times M$, in partitioned form,

$$V = \begin{pmatrix} \bar{V} \\ 0 \end{pmatrix},$$

where $\bar{V}$ is an $M \times M$ real diagonal matrix. With this notation, the determinant and the inverse of the conditional covariance become

$$\det \left(I_T + \frac{\rho}{M} \cdot SS^\dagger \right) = \det \left(I_M + \frac{\rho}{M} \cdot \bar{V}^2 \right) = \prod_{m=1}^{M} \left(1 + \frac{\rho \cdot v_m^2}{M} \right),$$

$$\left(I_T + \frac{\rho}{M} \cdot SS^\dagger \right)^{-1} = I_T - \Phi \begin{bmatrix} I_M - (I_M + \frac{\rho}{M} \cdot \bar{V}^2)^{-1} & 0 \\ 0 & 0_{T-M} \end{bmatrix} \Phi^\dagger, \qquad \text{(B.1)}$$

where $v_1, \ldots, v_M$ are the diagonal elements of $\bar{V}$.

The X-dependence in $p(X \mid \Phi V)$ appears only as $XX^\dagger$. We use the eigenvector-eigenvalue decomposition,

$$XX^\dagger = \Psi \Lambda \Psi^\dagger,$$

where Ψ is $T \times T$ and unitary, and Λ has the structure

$$\Lambda = \begin{bmatrix} \bar{\Lambda} & 0 \\ 0 & 0 \end{bmatrix},$$

where $\bar{\Lambda}$ is $\min(N,T) \times \min(N,T)$, diagonal, real and nonnegative.

Using (B.1), we have

$$\begin{aligned}
&\operatorname{tr}\left\{\left(I_T + \frac{\rho}{M}\cdot SS^\dagger\right)^{-1} XX^\dagger\right\} \\
&\quad = \operatorname{tr} XX^\dagger - \operatorname{tr}\left\{\begin{bmatrix} I_M - (I_M + \frac{\rho}{M}\cdot \bar{V}^2)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \Phi^\dagger XX^\dagger \Phi\right\}.
\end{aligned}$$

We may now compute one of the innermost integrals in (8) to obtain

$$\begin{aligned}
\int d\Phi\, p(\Phi) p(X \mid \Phi V) &= \frac{1}{\pi^{TN} \det^N (I_M + \frac{\rho}{M}\cdot \bar{V}^2)} \\
&\quad \cdot \int d\Phi\, p(\Phi)\cdot \exp\left\{\operatorname{tr}\left[\left(I_T + \frac{\rho}{M}\cdot VV^\dagger\right)^{-1} \Phi^\dagger \Psi\Lambda\Psi^\dagger\Phi\right]\right\} \\
&= \frac{1}{\pi^{TN} \det^N (I_M + \frac{\rho}{M}\cdot \bar{V}^2)} \\
&\quad \cdot \int d\Phi\, p(\Phi)\cdot \exp\left\{\operatorname{tr}\left[\left(I_T + \frac{\rho}{M}\cdot VV^\dagger\right)^{-1} \Phi^\dagger \Lambda\Phi\right]\right\} \\
&= \frac{\exp\left\{-\operatorname{tr} XX^\dagger\right\}}{\pi^{TN} \det^N (I_M + \frac{\rho}{M}\cdot \bar{V}^2)} \\
&\quad \cdot \int d\Phi\, p(\Phi)\cdot \exp\left\{\sum_{n=1}^{\min(N,T)} \sum_{m=1}^{M} \lambda_n \cdot \left(\frac{\rho v_m^2}{M + \rho v_m^2}\right) \cdot |\phi_{nm}|^2\right\} \qquad \text{(B.2)}
\end{aligned}$$

where $p(\Phi)$ is given in (A.5), and $\{\lambda_n\}_{n=1}^{\min(N,T)}$ denote the diagonal elements of $\bar{\Lambda}$. In the above, we change integration variable from Φ to $\Psi^\dagger\Phi$, and use the fact that $\Psi^\dagger\Phi$ has the same probability density as Φ.

There is, at present, no general expression available for the expectation with respect to $p(\Phi)$ that appears in (B.2). However, a closed-form expression can be obtained for the special cases where either $M = 1$, or $N = 1$, or $T = 1$. In any of these cases, the argument of the $\exp\{\cdot\}$ is a function of only a single column or row of Φ, taking the form $\sum_{\ell=1}^{T} a_\ell \cdot |\phi_\ell|^2$ for some real $\{a_\ell\}_{\ell=1}^{T}$, where ϕ is an i.d. unit vector. Three possible forms for the integral are:

a) When $a_1, \ldots, a_T$ are all positive and distinct, then

$$\mathrm{E}\left\{\exp\left(\sum_{\ell=1}^{T} a_\ell \cdot |\phi_\ell|^2\right)\right\} = \Gamma(T)\cdot \sum_{i=1}^{T} \frac{e^{a_i} - e^{a_1}}{\prod_{j\neq i}(a_i - a_j)}, \quad a_1 > a_2 > \cdots > a_T > 0.$$

b) When $0 < L < T$ of the a_ℓ's are positive and distinct, and the remainder zero, then

$$\begin{aligned} \mathrm{E}\left\{\exp\left(\sum_{\ell=1}^{L} a_\ell \cdot |\phi_\ell|^2\right)\right\} &= \frac{\Gamma(T)}{\Gamma(T-L)} \cdot \sum_{i=1}^{L} \frac{1}{\prod_{j\neq i}(a_i - a_j)} \\ &\quad \cdot \left[\frac{e^{a_i}\cdot\gamma(T-L,a_i)}{a_i^{T-L}} - \frac{e^{a_1}\cdot\gamma(T-L,a_1)}{a_1^{T-L}}\right], \\ &\quad a_1 > \cdots > a_L > a_{L+1} = \cdots = a_T = 0, \end{aligned}$$

where

$$\gamma(T,z) \stackrel{\text{def}}{=} \int_0^z q^{T-1} e^{-q}\, dq \tag{B.3}$$

is the *incomplete gamma* function.

c) When $L < T$ of the a_ℓ's are positive and equal and the remainder zero, we have

$$\begin{aligned} \mathrm{E}\left\{\exp\left(a\sum_{\ell=1}^{L} |\phi_\ell|^2\right)\right\} &= \frac{\Gamma(T)e^a}{\Gamma(T-L)a^{T-L}} \cdot \sum_{i=0}^{L-1} \frac{(-1)^i a^{-i}\gamma(T-L+i,a)}{i!(L-1-i)!}, \\ &\quad a = a_1 = \cdots = a_L > a_{L+1} = \cdots = a_T = 0. \end{aligned}$$

Finally, we note that when either $M \ll T$ or $N \ll T$, the M column vectors of Φ (in the former case) or the N row vectors of Φ (in the latter case) that are in the argument of the $\exp\{\cdot\}$ in (B.2) are approximately independent. In this case the expectation is approximately equal to a product of expectations involving the individual column or row vectors of Φ.

B.2 Integrating with respect to X

From (4) and Theorem 2, the expectation of the logarithm of the numerator that appears in (8) reduces to an expectation with respect to V,

$$\mathrm{E}\,\log p(X \mid \Phi V) = -TN(\log e + \log \pi) - N \cdot \mathrm{E}\left\{\sum_{m=1}^{M} \log\left(1 + \frac{\rho v_m^2}{M}\right)\right\}. \tag{B.4}$$

The marginal probability density of X, obtained by taking the expectation of (B.2) with respect to V, depends only the eigenvalues of $XX^\dagger$ and can be written in the form

$$p(X) = \frac{\exp\left(-\mathrm{tr}XX^\dagger\right)}{\pi^{TN}} \cdot f(\lambda), \tag{B.5}$$

where λ is the vector of $\min(N,T)$ eigenvalues, and

$$\begin{aligned} f(\lambda) &= \int dV \cdot \frac{p(V)}{\det^N\left(I_M + \frac{\rho}{M}\cdot \bar{V}^2\right)} \\ &\quad \cdot \int d\Phi\, p(\Phi)\cdot \exp\left\{\sum_{n=1}^{\min(N,T)} \sum_{m=1}^{M} \lambda_n \cdot \left(\frac{\rho v_m^2}{M+\rho v_m^2}\right)\cdot |\phi_{nm}|^2\right\}. \end{aligned} \tag{B.6}$$

The density $p(\Phi)$ is given in (A.5). Using (B.5), we have

$$\mathrm{E}\,\log p(X) = \int dX \cdot \frac{\exp\left(-\mathrm{tr}XX^\dagger\right)}{\pi^{TN}}\cdot f(\lambda)\cdot \log\left[\frac{\exp\left(-\mathrm{tr}XX^\dagger\right)}{\pi^{TN}}\cdot f(\lambda)\right]. \tag{B.7}$$

Observe that the expression $(1/\pi^{TN})\exp(-\mathrm{tr}\,XX^\dagger)$ takes the form of the joint probability density function on X, as if the components of X were independent $\mathcal{CN}(0,1)$. Consequently the expression (B.7) is equivalent to an expectation with respect to X, as if the components of X were independent $\mathcal{CN}(0,1)$. With the components having this distribution, the joint probability density of the ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ is [2]

$$p(\lambda) = \frac{e^{-\sum_{\ell=1}^{\min(N,T)}\lambda_\ell}\cdot\left(\prod_{\ell=1}^{\min(N,T)}\lambda_\ell\right)^{|T-N|}\cdot\prod_{i<j}(\lambda_i-\lambda_j)^2}{\prod_{\ell=1}^{\min(N,T)}\Gamma(T-\ell+1)\cdot\Gamma(N-\ell+1)}. \tag{B.8}$$

Therefore (B.7) becomes

$$\mathrm{E}\,\log p(X) = \int d\lambda\cdot p(\lambda)\cdot f(\lambda)\cdot\log\left[\frac{\exp\left(-\sum_{\ell=1}^{\min(N,T)}\lambda_\ell\right)}{\pi^{TN}}\cdot f(\lambda)\right]. \tag{B.9}$$

Finally, subtracting (B.9) from (B.4) gives a simplified expression for mutual information,

$$\begin{aligned} I(X;S) &= -TN\log e - N\cdot\sum_{m=1}^{M}\mathrm{E}\,\log\left(1+\frac{\rho v_m^2}{M}\right) \\ &\quad - \int d\lambda\cdot p(\lambda)\cdot f(\lambda)\cdot\left[\log f(\lambda) - (\log e)\cdot\sum_{\ell=1}^{\min(N,T)}\lambda_\ell\right], \end{aligned} \tag{B.10}$$

where $p(\lambda)$ is given by (B.8) and $f(\lambda)$ is given by (B.6). Thus, computation of the mutual information requires integrations over the M real components of V, the $\min(N,T)$ real components of λ, and $M\cdot\min(N,T)$ complex components of Φ; as shown in Appendix B.1, the integration over these components of

Φ can be performed analytically in some cases.

C Appendix: Perfect-Knowledge Upper Bound on Capacity

If the receiver somehow knew the random propagation coefficients, the capacity would be greater than for the case of interest where the receiver does not know the propagation coefficients. Telatar [7] computes the perfect-knowledge capacity for the case $T = 1$; it is straightforward to extend his analysis to $T > 1$.

To obtain the perfect-knowledge upper bound for the signal model (3), we suppose that the receiver observes the propagation matrix H through a separate noise-free channel. This perfect-knowledge fading link is completely described by the conditional probability density, $p(X, H \mid S) = p(X \mid H, S) \cdot p(H)$. The perfect-knowledge capacity is obtained by maximizing the mutual information between (X, H) and S with respect to $p(S)$. The mutual information is

$$\begin{aligned} I(X,H;S) &= \mathrm{E}\,\log \frac{p(X,H \mid S)}{p(X,H)} \\ &= \mathrm{E}\,\log \frac{p(X \mid H,S)}{p(X \mid H)} \\ &= \mathrm{E}\left\{\mathrm{E}\left\{\log\left(\frac{p(X \mid H,S)}{p(X \mid H)}\right) \mid H\right\}\right\}. \end{aligned}$$

The inner expectation, conditioned on H, is simply the mutual information for the classical additive Gaussian noise case and is maximized by making the components of S independent $\mathcal{CN}(0,1)$. (In performing the maximization, the expected power of each component of S is constrained to be equal, since the transmitter does not know H and therefore cannot allocate power among its antennas in accordance with H.) The resulting perfect-knowledge capacity is

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

This expression is the capacity associated with a block of T symbols, where T is the coherence interval. The normalized capacity, C_u/T, is independent of T.

The $N \times N$ matrix $H^\dagger H/M$ is equal to the average of M statistically independent outer products. For fixed T and N, when M grows large this matrix converges to the identity matrix. Therefore,

$$\lim_{M\to\infty} C_u = T \cdot N \cdot \log(1+\rho).$$

Although the total power that is radiated is unchanged as M increases, it appears that one ultimately achieves the equivalent of N independent nonfading subchannels, each with SNR ρ.

When the number of receiver antennas is large, we use the identity $\det\left[I_N + \frac{\rho}{M} H^\dagger H\right] = \det\left[I_M + \frac{\rho}{M} H H^\dagger\right]$. For fixed T and M, the $M \times M$ matrix, $HH^\dagger/N$, is equal to the average of N statistically independent outer products. This converges, as $N \to \infty$, to the identity matrix, and therefore

$$C_u \sim T \cdot M \cdot \log\left(1 + \frac{\rho N}{M}\right).$$

In effect, one has M independent nonfading subchannels, each having signal-to-noise ratio $\rho N/M$.

D Appendix: Asymptotic behavior of C_l as $T \to \infty$

We start with equation (12), which can be written

$$\begin{aligned} C_l &= (\log e) \cdot \Big[-T - \ln(T-1) + \mathrm{E}\,\Big\{(T-1)\ln\Big(\frac{\rho T \lambda}{1+\rho T}\Big) - \ln\gamma\Big(T-1, \frac{\rho T\lambda}{1+\rho T}\Big)\Big\} \\ &\quad + \frac{1}{1+\rho T}\mathrm{E}\,\lambda\Big], \end{aligned} \tag{D.1}$$

where the expectation is with respect to the probability density

$$q(\lambda) = \frac{(T-1)e^{-\lambda/(1+\rho T)}\gamma\Big(T-1, \frac{\rho T\lambda}{1+\rho T}\Big)}{\Gamma(T)(1+\rho T)\left[\frac{\rho T}{1+\rho T}\right]^{T-1}}, \qquad \lambda \geq 0.$$

Lemma A, which is now presented, helps simplify these expectations. It turns out that the second expectation is negligible when compared with the first.

Lemma A: For any function $g(\lambda)$,

$$\mathrm{E}\, g(\lambda) = [e^{1/\rho} + O(T^{-1})] \int_0^\infty du\, e^{-u} \frac{u^{T-2}}{(T-2)!} \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} g((1+\rho T)\lambda)$$

as $T \to \infty$.

Proof:

$$\int_0^\infty q(\lambda) g(\lambda)\, d\lambda = \left[1 + \frac{1}{\rho T}\right]^{T-1} \int_0^\infty d\lambda\, \frac{e^{-\lambda/(1+\rho T)}}{1+\rho T} g(\lambda) \int_0^{\frac{\rho T \lambda}{1+\rho T}} du\, e^{-u} \frac{u^{T-2}}{(T-2)!}$$

$$
\begin{aligned}
&= [e^{1/\rho} + O(T^{-1})] \int_0^\infty d\lambda\, e^{-\lambda} g((1+\rho T)\lambda) \int_0^{\rho T\lambda} du\, e^{-u} \frac{u^{T-2}}{(T-2)!} \\
&= [e^{1/\rho} + O(T^{-1})] \int_0^\infty du\, e^{-u} \frac{u^{T-2}}{(T-2)!} \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} g((1+\rho T)\lambda).
\end{aligned}
$$

□

Define

$$
f_T(u) \stackrel{\text{def}}{=} \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} g((1+\rho T)\lambda).
$$

Then the two expectations in (D.1) involve integrals of the form

$$
\int_0^\infty du\, e^{-u} \frac{u^{T-2}}{(T-2)!} f_T(u) \tag{D.2}
$$

for different functions $f_T(\cdot)$. For the second expectation, $g(\lambda) = \lambda/(1+\rho T)$, and hence

$$
f_T(u) = \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} \lambda = e^{-\frac{u}{\rho T}} \Big(1 + \frac{u}{\rho T}\Big).
$$

The integral (D.2) can be explicitly evaluated in this case, the result being

$$
\int_0^\infty du\, e^{-u} \frac{u^{T-2}}{(T-2)!} f_T(u) = e^{-1/\rho}(1 + 1/\rho) + O(T^{-1}),
$$

and hence, from Lemma A, we obtain

$$
\begin{aligned}
\frac{1}{1+\rho T} \mathrm{E}\, \lambda &= [e^{1/\rho} + O(T^{-1})][e^{-1/\rho}(1+1/\rho) + O(T^{-1})] \\
&= 1 + 1/\rho + O(T^{-1})
\end{aligned}
$$

The first expectation in (D.1) is not so easy to evaluate, and requires a lemma to help approximate it for large T.

Lemma B: Let $f_T(u)$, $T = 1, 2, \ldots$, be any sequence of functions that are continuous on $u \in (0, \infty)$, and where $\sup_{u \geq T} |f_T(u)|$ is bounded by some polynomial $r(T)$, as $T \to \infty$. Then, for any positive p,

$$
\begin{aligned}
\int_0^\infty du\, e^{-u} \frac{u^T}{T!} f_T(u) &= f_T(T) + O\Big(\max_{v^2 \leq \frac{(2p+1)\ln T}{T}} |f_T(T(1+v)) - f_T(T)| \Big) \\
&\quad + O\Big(\frac{r(T)}{T^p}\Big) + O\Big(\frac{1}{T^{p+1}}\Big) \int_0^T du\, |f_T(u)|
\end{aligned} \tag{D.3}
$$

as $T \to \infty$.

Proof: Since $\int_0^\infty du\, e^{-u} u^T/T! = 1$, we have that

$$\left| \int_0^\infty du\, e^{-u} \frac{u^T}{T!} f_T(u) - f_T(T) \right| \le \int_0^\infty du\, e^{-u} \frac{u^T}{T!} |f_T(u) - f_T(T)|$$
$$= \sqrt{\frac{T}{2\pi e^{\theta_T/(6T)}}} \int_{-1}^\infty dv\, g_T(v), \quad \text{(D.4)}$$

where $g_T(v) \stackrel{\text{def}}{=} e^{-Tv}(1+v)^T |f_T(T(1+v)) - f_T(T)|$, and, by Stirling's approximation, $0 < \theta_T < 1$. The function $e^{-Tv}(1+v)^T$ has its maximum value one at $v = 0$, and decreases monotonically as v increases or decreases. Define $a_T \stackrel{\text{def}}{=} \sqrt{((2p+1)\ln T)/T}$; then

$$\begin{aligned} e^{Ta_T}(1-a_T)^T &= e^{\sqrt{(2p+1)T\ln T}} e^{T\ln(1-\sqrt{((2p+1)\ln T)/T})} \\ &= e^{\sqrt{(2p+1)T\ln T}} e^{T[-\sqrt{((2p+1)\ln T)/T} - ((2p+1)\ln T)/(2T) + o(1/T)]} \\ &= e^{-(p+1/2)\ln T + o(1)} \\ &= O(T^{-p-1/2}). \end{aligned} \quad \text{(D.5)}$$

We proceed by breaking the range of integration in (D.4) into the four disjoint intervals, $[-1, -a_T]$, $(-a_T, a_T]$, $(a_T, 1]$, and $(1, \infty)$. Define $c_T \stackrel{\text{def}}{=} \sqrt{T/(2\pi e^{\theta_T/(6T)})}$; for the first interval, (D.5) implies that

$$\begin{aligned} c_T \int_{-1}^{-a_T} dv\, g_T(v) &\le c_T \max_{v\in[-1,-a_T]} \left\{ e^{-Tv}(1+v)^T \right\} \int_{-1}^{-a_T} dv\, [|f_T(T)| + |f_T(T(1+v))|] \\ &\le O\left(\frac{1}{T^p}\right) \int_{-1}^0 dv\, [r(T) + |f_T(T(1+v))|] \\ &= O\left(\frac{r(T)}{T^p}\right) + O\left(\frac{1}{T^{p+1}}\right) \int_0^T du\, |f_T(u)|. \end{aligned} \quad \text{(D.6)}$$

For the second interval, we obtain

$$\begin{aligned} c_T \int_{-a_T}^{a_T} dv\, g_T(v) &\le c_T \int_{-a_T}^{a_T} dv\, e^{-Tv}(1+v)^T \max_{|v|\le a_T} |f_T(T(1+v)) - f_T(T)| \\ &\le \max_{|v|\le a_T} |f_T(T(1+v)) - f_T(T)| \cdot c_T \int_{-1}^\infty dv\, e^{-Tv}(1+v)^T \\ &= \max_{|v|\le a_T} |f_T(T(1+v)) - f_T(T)| \end{aligned} \quad \text{(D.7)}$$

where the indicated maximum exists because of the continuity of $f_T(\cdot)$.

In the third interval, an expansion similar to (D.5) shows that $e^{-Ta_T}(1+a_T)^T = O(T^{-p-1/2})$, and, from the same reasoning that leads to (D.6),

$$\begin{aligned} c_T \int_{a_T}^{1} dv\, g_T(v) &= O\Big(\frac{r(T)}{T^p}\Big) + O\Big(\frac{1}{T^{p+1}}\Big) \int_T^{2T} du\, |f_T(u)| \\ &= O\Big(\frac{r(T)}{T^p}\Big). \end{aligned} \tag{D.8}$$

For the fourth interval, note that $(1+v)^T \le e^{T(v+1/2)/2}$ for $v \ge 1$. Therefore,

$$\begin{aligned} c_T \int_1^\infty dv\, g_T(v) &\le c_T \int_1^\infty dv\, e^{-T(v-1/2)/2} |f_T(T(1+v)) - f_T(T)| \\ &\le \frac{c_T}{T} e^{-T/4} \int_0^\infty du\, e^{-u/2} [|f_T(u+2T)| + [f_T(T)|] \\ &\le 2\frac{c_T}{T} r(T) e^{-T/4} \int_0^\infty du\, e^{-u/2} \\ &= 4\frac{c_T}{T} r(T) e^{-T/4}. \end{aligned} \tag{D.9}$$

This term is negligible compared with the $O(r(T)/T^p)$ term in (D.8) for sufficiently large T. Combining equations (D.6)–(D.9) completes the proof. □

We are now in a position to evaluate the first expectation in (D.1) for large T. Here, $g(\lambda) = (T-1)\ln(\rho T\lambda/(1+\rho T)) - \ln\gamma(T-1, \rho T\lambda/(1+\rho T))$. Integration by parts yields

$$f_T(u) = \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} \Big[(T-1)\ln(\rho T\lambda) - \ln\gamma(T-1,\rho T\lambda)\Big] \tag{D.10}$$

$$\begin{aligned} &= \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} \Big[(T-1)\ln(\rho T\lambda) - \ln\int_0^{\rho T\lambda} dv\, e^{-v} v^{T-2}\Big] \\ &= (T-1)\Big[-e^{-\lambda}\ln(\rho T\lambda)\Big|_{\frac{u}{\rho T}}^\infty + \int_{\frac{u}{\rho T}}^\infty d\lambda\, \frac{e^{-\lambda}}{\lambda}\Big] \\ &\quad + e^{-\lambda}\ln\gamma(T-1,\rho T\lambda)\Big|_{\frac{u}{\rho T}}^\infty - \int_{\frac{u}{\rho T}}^\infty d\lambda\, e^{-\lambda} \frac{1}{\gamma(T-1,\rho T\lambda)} e^{-\rho T\lambda} (\rho T\lambda)^{T-2}\rho T \\ &= (T-1)\Big[e^{-\frac{u}{\rho T}}\ln u + \mathrm{E}_1\Big(\frac{u}{\rho T}\Big)\Big] - e^{-\frac{u}{\rho T}}\ln\gamma(T-1,u) \\ &\quad - \int_u^\infty d\lambda\, e^{-(1+\frac{1}{\rho T})\lambda} \frac{\lambda^{T-2}}{\gamma(T-1,\lambda)}, \end{aligned} \tag{D.11}$$

where $\mathrm{E}_1(x)$ is the exponential integral defined in (13). We are interested in studying the behavior of $f_T(u)$ for u in the neighborhood of T. The following lemma, which is proven in [8], aids us.

Lemma C: Uniformly for α in any finite positive interval,

$$\frac{\gamma(T,\alpha T)}{\Gamma(T)} = \frac{1}{2}\mathrm{erfc}(\zeta\sqrt{T}) + o(\mathrm{erfc}(\zeta\sqrt{T}))$$

as $T \to \infty$, where $\mathrm{erfc}(y) \stackrel{\mathrm{def}}{=} (2/\sqrt{\pi})\int_y^\infty e^{-u^2}du$ is the *complementary error function*, and $\zeta \stackrel{\mathrm{def}}{=} \pm\sqrt{\alpha - 1 - \ln\alpha}$; the positive sign is taken when $\alpha < 1$, and the negative when $\alpha > 1$.

Employing Lemma C with $\alpha = T/(T-1)$, we see that $\zeta\sqrt{T} = O(1/\sqrt{T})$, and

$$\ln\gamma(T-1,T) = \ln\Gamma(T-1) + O(1) = T\ln T - T + O(\ln T), \tag{D.12}$$

where the second equality is a consequence of Stirling's approximation of $\Gamma(T-1)$. Furthermore,

$$\begin{aligned}\int_T^\infty dy\, e^{-(1+\frac{1}{\rho T})y}\frac{y^{T-2}}{\gamma(T-1,y)} &\leq \frac{1}{\gamma(T-1,T)}\int_T^\infty dy\, e^{-(1+\frac{1}{\rho T})y}y^{T-2} \\ &\leq \frac{1}{\gamma(T-1,T)}\int_0^\infty dy\, e^{-y}y^{T-2} \\ &= O(1)\end{aligned}$$

as $T \to \infty$. Therefore, for $u = T$ and large T the integral in (D.11) is negligible in comparison with the other three terms, and it is dropped from further consideration. We now apply Lemma B to $f_T(u)$ as given in (D.11). The hypotheses of the lemma are satisfied because, by (D.12),

$$\begin{aligned} f_T(T) &= (T-1)\Big[e^{-1/\rho}\ln T + \mathrm{E}_1(1/\rho)\Big] - e^{-1/\rho}\Big[T\ln T - T\Big] + O(\ln T) \\ &= T\mathrm{E}_1(1/\rho) + Te^{-1/\rho} + O(\ln T), \end{aligned} \tag{D.13}$$

and, by (D.10), $\lim_{u\to\infty} f_T(u) = 0$.

Equation (D.13) is the first term in equation (D.3) of Lemma B. The second term in (D.3) requires an estimate of $\max_{v^2 \leq (2p+1)(\ln T)/T} |f_T(T(1+v)) - f_T(T)|$, for which we use Lemma 3 and assume that $p > 0$ is arbitrary. For $v^2 = O((\ln T)/T)$, we have that $T(1+v) = T + O(\sqrt{T\ln T})$, and therefore $\alpha = 1 + O(\sqrt{(\ln T)/T})$, and $\zeta\sqrt{T} = O(\sqrt{\ln T})$. Furthermore, $(1/2)\mathrm{erfc}(O(\sqrt{\ln T}))$ either goes to one, or goes to zero with rate at most proportional to $1/(T\sqrt{\ln T})$, depending on whether the $O(\sqrt{\ln T})$ term is negative or positive. Thus, by Lemma C and (D.12),

$$\ln\gamma(T-1, T+O(\sqrt{T\ln T})) \quad = \quad \ln\Gamma(T-1) + \ln\{(1/2)\mathrm{erfc}(\zeta\sqrt{T-1})$$

$$
\begin{aligned}
&\quad + o(\mathrm{erfc}(\zeta\sqrt{T-1}))\} \\
&= T\ln T - T + O(\ln T) + \ln\{(1/2)\mathrm{erfc}(\zeta\sqrt{T-1}) \\
&\quad + o(\mathrm{erfc}(\zeta\sqrt{T-1}))\} \\
&= T\ln T - T + O(\ln T).
\end{aligned}
$$

We also have $\ln(T + O(\sqrt{T\ln T})) = \ln T + O(\sqrt{(\ln T)/T})$ and $\mathrm{E}_1((1/\rho) + O(\sqrt{(\ln T)/T})) = \mathrm{E}_1(1/\rho) + O(\sqrt{(\ln T)/T})$. Combining all these facts, we obtain

$$
\max_{v^2 \le \frac{(2p+1)\ln T}{T}} |f_T(T(1+v)) - f_T(T)| = O(\sqrt{T\ln T}) = O(\sqrt{T\log T}). \tag{D.14}
$$

The bound given in (D.14) holds for arbitrary $p > 0$. Therefore, p may be chosen large enough so that the remaining terms in (D.3) are negligible in comparison to (D.14). Lemmas A and B in conjunction with (D.13) and (D.14) consequently yield

$$
\begin{aligned}
&\mathrm{E}\left\{(T-1)\ln\frac{\rho T\lambda}{1+\rho T} - \ln\gamma\Big(T-1, \frac{\rho T\lambda}{1+\rho T}\Big)\right\} \\
&= [e^{1/\rho} + O(T^{-1})]\int_0^\infty du\, e^{-u}\frac{u^{T-2}}{(T-2)!}f_T(u) \\
&= Te^{1/\rho}\mathrm{E}_1(1/\rho) + T + O(\sqrt{T\log T}).
\end{aligned}
$$

Thus, (D.1) becomes

$$
C_l = T(\log e)e^{1/\rho}\mathrm{E}_1(1/\rho) - O(\sqrt{T\log T}). \tag{D.15}
$$

E Appendix: Signaling Scheme that Achieves Capacity as $T \to \infty$

For $M = N = 1$ and T sufficiently large, we show that the mutual information for $p(v_1) = \delta(v_1 - \sqrt{T})$, given in (12), exceeds the mutual information for any other $p(v_1)$. Suppose that $p(v_1) = p_a\delta(v_1 - \sqrt{aT}) + p_b\delta(v_1 - \sqrt{bT})$, where a and b are contained in a positive finite interval as $T \to \infty$. Since $\mathrm{E}\, v_1^2 = T$, it must hold that $ap_a + bp_b = 1$, and we assume that p_a and p_b both remain positive as $T \to \infty$. It is then a simple matter to parallel the derivation of C_l in (12) to obtain the mutual information

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

where

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

The same analysis as in Appendix D now yields

$$\begin{aligned}&\int_0^\infty q(\lambda)\ln\Big(q(\lambda)\Gamma(T)/\lambda^{T-1}\Big)d\lambda \\ &= p_a\left[1+\frac{1}{\rho aT}\right]^{T-1}\int_0^\infty du\, e^{-u}\frac{u^{T-2}}{(T-2)!}\int_{\frac{u}{\rho aT}}^\infty d\lambda\, e^{-\lambda}g((1+\rho aT)\lambda) \\ &\quad + p_b\left[1+\frac{1}{\rho bT}\right]^{T-1}\int_0^\infty du\, e^{-u}\frac{u^{T-2}}{(T-2)!}\int_{\frac{u}{\rho bT}}^\infty d\lambda\, e^{-\lambda}g((1+\rho bT)\lambda),\end{aligned}$$

where

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

Omitting tedious details that are essentially the same as in Appendix D, we obtain

$$I = (\log e)T\Big(p_a e^{\frac{1}{\rho a}}\mathrm{E}_1\Big(\frac{1}{\rho a}\Big) + p_b e^{\frac{1}{\rho b}}\mathrm{E}_1\Big(\frac{1}{\rho b}\Big)\Big) + O(\sqrt{T\log T}) \qquad \text{(E.1)}$$

as $T \to \infty$, where $\mathrm{E}_1(x)$ is the exponential integral defined in (13). But the function $e^{1/\rho}\mathrm{E}_1(1/\rho)$ is strictly concave in ρ, and therefore by Jensen's inequality,

$$p_a e^{\frac{1}{\rho a}}\mathrm{E}_1\Big(\frac{1}{\rho a}\Big) + p_b e^{\frac{1}{\rho b}}\mathrm{E}_1\Big(\frac{1}{\rho b}\Big) \le e^{\frac{1}{\rho(ap_a+bp_b)}}\mathrm{E}_1\Big(\frac{1}{\rho(ap_a+bp_b)}\Big) = e^{1/\rho}\mathrm{E}_1(1/\rho),$$

with equality if and only if $a = b = 1$. If $a = b = 1$, the two masses in $p(v_1)$ collapse into a single mass at $v_1 = \sqrt{T}$. Hence, by (D.15), $C_l \ge I$ for sufficiently large T, with equality if and only if the two masses in $p(v_1)$ approach a single mass at $\sqrt{T}$, as $T \to \infty$.

We now outline how to generalize the above argument to show that any $p(v_1)$ asymptotically generates less mutual information than $p(v_1) = \delta(v_1 - \sqrt{T})$. The expansion (E.1) can be generalized to n masses $p(v_1) = \sum_{j=1}^{n} p_j \delta(v_1 - \sqrt{a_j T})$, to obtain

$$I = (\log e)T \left(\sum_{j=1}^{n} p_j e^{\frac{1}{\rho a_j}} \mathrm{E}_1\left(\frac{1}{\rho a_j}\right) \right) + O(\sqrt{T \log T}). \tag{E.2}$$

Provided that $a_1, \ldots, a_n$ are taken from some finite positive interval, the asymptotic expansion (E.2) is uniform, and hence remains valid even if we let n become unbounded (say, for example, as a function of T). As $T \to \infty$, the mutual information (E.2) is therefore maximized by having $a_1, \ldots, a_n \to 1$, which reduces the multiple masses to a single mass at $v_1 = \sqrt{T}$. On a finite interval, we can uniformly approximate any continuous density for $p(v_1/\sqrt{T})$ with masses, and the preceding argument therefore tells us that we are asymptotically better off replacing the continuous density on this finite interval with a mass at $v_1/\sqrt{T} = 1$.

References

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

[2] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*. PhD thesis, Massachusetts Institute of Technology, May 1989.

[3] 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.

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

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

[6] G. W. Stewart, "The efficient generation of random orthogonal matrices with an application to condition estimators," *Siam J. Numer. Anal.*, vol. 17, pp. 403–409, 1980.

[7] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Tech. Rep. #BL0112170-950615-07TM, AT&T Bell Laboratories, 1995.

[8] N. M. Temme, "Uniform asymptotic expansions of the incomplete gamma functions and the incomplete beta functions," *Math. Comp.*, vol. 29, pp. 1109–1114, 1975.

[9] 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.

Figure 1: Wireless link comprising M transmitter and N receiver antennas. Every receiver antenna is connected to every transmitter antenna through an independent, random, unknown propagation coefficient having Rayleigh distributed magnitude and uniformly distributed phase. Normalization ensures that the total expected transmitted power is independent of M for a fixed ρ.

Figure 2: Propagation coefficients change randomly every T symbol periods. Channel coding is performed over multiple independent fading intervals.

Figure 3: The transmitted signals that achieve capacity are mutually orthogonal with respect to time. The constituent orthonormal unit vectors are isotropically distributed (see Appendix A), and independent of the signal magnitudes, which have mean-square value T. The solid sphere of radius $T^{1/2}$ demarcates the root mean-square. For $T \gg M$, the vectors all lie approximately on the surface of this sphere. The shell of thickness $\varepsilon T^{1/2}$ is discussed in Section 5.

Figure 4: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR= 0dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 12$. As per Theorem 3, the capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

Figure 5: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR= 6dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 4$. The capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

Figure 6: Normalized capacity, and upper and lower bounds, versus coherence interval T (SNR= 12dB, one transmitter antenna, one receiver antenna). The lower bound and capacity meet at $T = 3$. The capacity approaches the perfect-knowledge upper bound as $T \to \infty$.

Figure 7: Capacity and perfect-knowledge upper bound versus number of receiver antennas N (SNR=0dB, 6dB, 12dB, arbitrary number of transmitter antennas, coherence interval equal to one). The gap between capacity and upper bound only widens as $N \to \infty$.

Figure 8: Normalized capacity lower bounds and perfect-knowledge upper bounds versus number of transmitter antennas M (SNR=20dB, one receiver antenna, coherence interval equal to 100). The actual channel capacity lies in the shaded region. Lower bound peaks at $M = 3$; this peak is a valid lower bound for $M \geq 3$, giving us the modified lower bound.