Coding Requirements for Multiple-Antenna Channels with Unknown Rayleigh Fading

Ibrahim Abou-Faycal and Bertrand Hochwald

I. Abou-Faycal is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, Massachusetts. *E-mail: iaboufay@mit.edu.*

B. Hochwald is with the Mathematics of Communications Group, Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey. *E-mail: hochwald@research.bell-labs.com.*

March 18, 1999

Abstract

Multiple transmitter and receiver antennas can boost the reliability and capacity of wireless fading channels. When the channel is unknown to the transmitter and receiver, it is known that the fading coherence time in a piecewise-constant fading channel ultimately limits the channel capacity. In this work, we examine coding requirements for the unknown fading channel by computing the random coding error exponent. We show that the fading coherence time also plays a fundamental role in the error exponent, by proving that the error exponent is not increased by having more transmitter antennas than the number of samples in the coherence time. The signal structure that maximizes the exponent is computed and is shown to be very similar to the signal structure that achieves capacity. We calculate the minimum coding block length requirements as a function of probability of error for various fading coherence times. We conclude that coding over a certain number of independent fades is always needed to guarantee low probabilities of error, regardless of the coherence time.

Keywords

Antenna arrays, Random coding error exponent, Space-time coding

I. Introduction

Multiple antennas have recently been shown to be remarkably effective in wireless Rayleigh fading channels. The pioneering works of Foschini [1] and Telatar [2] demonstrate that channel capacities grow linearly with the smaller of the number of transmitter and receiver antennas, without boosting total transmitter power. However, these works assume that the receiver knows the fading or propagation environment perfectly. The transmitter is not assumed to know the fading.

Roughly speaking, when the receiver knows the channel fading, the capacity and coding issues become multi-dimensional versions of the capacity and coding issues for scalar known-fading channels. Codes that have large Euclidean distance generally perform well over such channels and the coding requirements to achieve low probability of error are basically understood [3]. However, when the fading is unknown to the receiver, especially in the presence of multiple antennas, coding requirements are less well understood.

When the fading is unknown to the receiver, Marzetta and Hochwald [4] study the capacity of a channel whose fading is piecewise-constant in blocks of T time samples, and changes independently between these blocks. They find that increasing the number of

transmitter antennas beyond T does not increase channel capacity, and they determine the signal structure that achieves capacity.

In this paper, we are interested in finding the coding requirements for the multiple-antenna channel with unknown fading. For this purpose we compute the random coding error exponent (RCEE), which provides us with an upper bound on the probability of error that can be achieved with a block code of given rate. Moreover, for a fixed "high" data rate and a given target probability of error, the RCEE allows us to determine roughly the minimum number of blocks one ought to use to achieve this probability error.

We show that the transmitter signal that maximizes the error exponent has properties very similar to the signal that achieves capacity. That is, it can be written as a $T \times M$ matrix of the form $S = \Phi V$, where $M \leq T$ is the number of transmitter antennas, Φ is a $T \times T$ isotropically distributed matrix, and V is a $T \times M$ independent real, diagonal, and nonnegative matrix. We show that the error exponent for $M > T$ is the same as for $M = T$. When $M = 1$, we prove that V (which can now be parameterized by a scalar) is a discrete random variable with a finite number of mass points. Moreover, we show that V is asymptotically deterministic when the signal-to-noise ratio (SNR), or when T, tends to infinity. In fact, we find numerically that V is generally deterministic for finite (but sufficiently large) values of SNR and T. This remarkable result allows us to completely characterize the optimal input that yields the best error exponent.

The outline of the paper is as follows. In Section II, we present the channel model and operating assumptions, and in Section III we define the RCEE. In Section IV, we study the structure of the optimal input, in Section V we write the optimality condition for the RCEE-achieving input random variable, and in Section VI we prove its discrete character. In Section VII, we study the optimal input when the SNR grows to infinity. Numerical results are discussed in Section VIII, and in Section IX we study the behavior of the RCEE when T goes to infinity. Section X concludes the paper, and detailed proofs are given in appendices.

II. The Channel Model

Figure 1 displays the multiple-antenna communication link in a Rayleigh flat-fading environment. The transmitter has M antennas and the receiver has N antennas. We assume

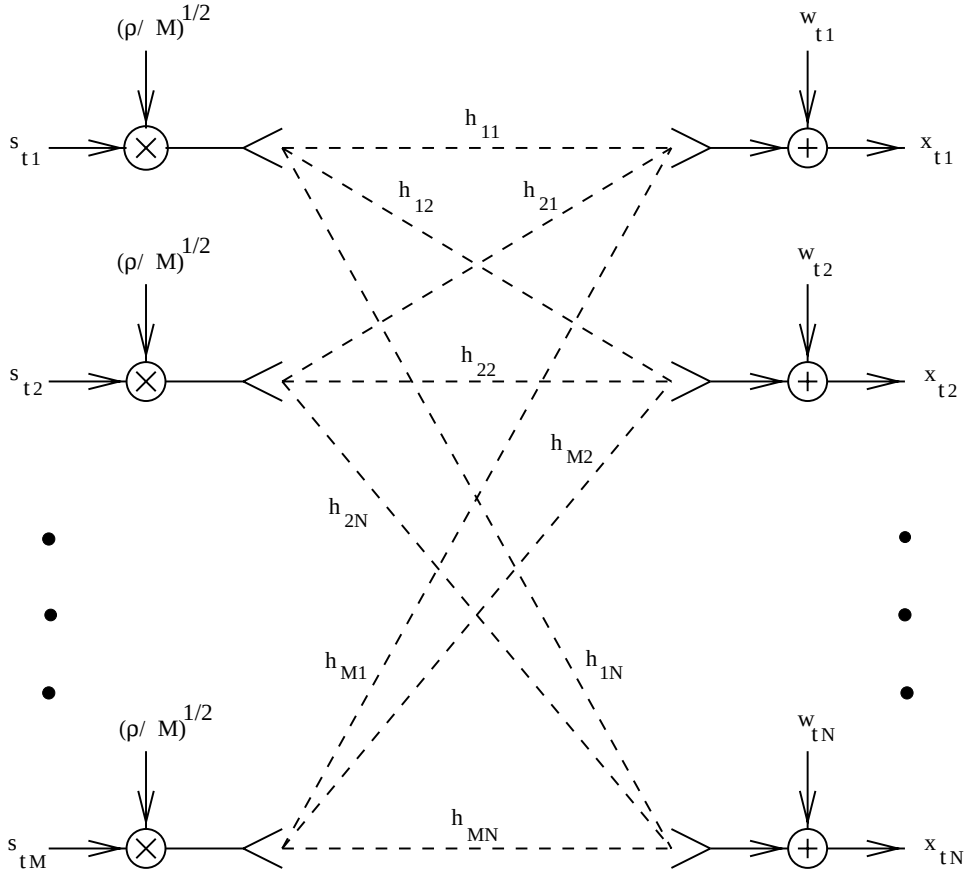


Fig. 1. Rayleigh fading channel comprising M transmitter and N receiver antennas. The transmitter power is normalized so that it does not vary with M. The fading coefficients h_{mn} and additive noise w_{tn} are zero-mean unit-variance complex Gaussian random variables.

that each receiver antenna responds to each transmitter antenna through a statistically independent fading coefficient that is constant for T symbol periods, changes independently to a new value, and is again constant for T symbol periods, and so on. Moreover, we assume the fading coefficients h_{mn} to be iid with a Rayleigh distributed amplitude and a uniform phase, that is, they are iid zero-mean complex circular Gaussian random variables with unit variance. The multiplicative factor $\sqrt{\rho/M}$ at every transmitter antenna ensures that the total transmitted power is independent of M and that ρ is the effective SNR at each receiver antenna.

Because the fading coefficients acquire new values (independent of the previous ones) every T symbols, the channel is memoryless when looking at blocks of length T. Let $S = [s_{tm}]$, $X = [x_{tn}]$, $H = [h_{mn}]$ and $W = [w_{tn}]$ for $t = 1, \ldots, T$, $m = 1, \ldots, M$, and

$n = 1, \ldots, N$. Note that S is a $T \times M$ complex matrix, H is $M \times N$, and X and W are $T \times N$. Then,

$$X = \sqrt{\frac{\rho}{M}} SH + W, \tag{1}$$

where neither H nor W is known to the receiver. The channel conditional probability density is therefore

$$p(X|S) = \frac{\exp\left(-\mathrm{tr}\,\left[[I_T + \frac{\rho}{M} SS^\dagger]^{-1} XX^\dagger\right]\right)}{\pi^{TN} \det^N [I_T + \frac{\rho}{M} SS^\dagger]}. \tag{2}$$

This is a zero-mean Gaussian density, where the transmitted signal matrix S appears in the covariance of the received signal. We assume that the space-time-average power constraint

$$\frac{1}{TM} E[\mathrm{tr}\, SS^\dagger] \leq 1 \tag{3}$$

is obeyed.

By using the channel L times in succession, we group together L signals, each a $T \times M$ matrix of the form given above for S, to form channel codewords and ask how small the block probability of error can be. The lowest probability of error is upper-bounded by the random coding error exponent defined in the next section.

III. The Random Coding Error Exponent

In this section we investigate the achievable error probability when decoding a block of L signals. The probability of error is generally a function of R, the bit rate per signal S; R is therefore the base-two logarithm of the number of signals $S_1, S_2, \ldots$, in our constellation.

An upper bound to the minimum achievable error probability is obtained by using maximum likelihood decoding of a random codebook. Under input power constraint $\int dP(S)[\mathrm{tr}\, SS^\dagger - TM] \leq 0$, Gallager [5] establishes that

$$P_e \leq B \exp\left[-L\{E_o(\mu, P, r) - \mu R\}\right], \tag{4}$$

for arbitrary $0 \leq \mu \leq 1$, $r \geq 0$ and probability distribution $P(S)$, where B is such that $\ln(B)/L \to 0$ as $L \to \infty$, and

$$E_o(\mu, P, r) = -\ln \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(X|S)^{1/(1+\mu)} e^{r[\mathrm{tr}\, SS^\dagger - TM]}\, dP(S)\right]^{1+\mu} dX. \tag{5}$$

The parameter r can be thought of as a Lagrange multiplier corresponding to the input power constraint. In general, r is strictly positive because the power constraint is always active, while μ is a free parameter that varies between 0 and 1. The function P is the joint probability distribution on the elements of the signal matrix S.

Define

$$F_o(\mu, P, r) = \exp\{-E_o(\mu, P, r)\} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(X|S)^{1/(1+\mu)} e^{r[\mathrm{tr}\, SS^\dagger - TM]} dP(S) \right]^{1+\mu} dX. \tag{6}$$

We prefer to work with $F_o(\mu, P, r)$ because it is convex in P, whereas $E_o(\mu, P, r)$ is neither convex nor concave in P.

The random coding exponent is defined to be the one that yields the tightest bound:

$$E_r(R) = \max_{0 \le \mu \le 1} \max_{r \ge 0} \max_{P} \{E_o(\mu, P, r) - \mu R\}, \tag{7}$$

where the maximization over P is subject to the input power constraints. As is suggested by equation (7), we perform the maximization over the input probability distribution first, then we maximize over the parameter r, and finally we look for the optimal μ that generates the highest exponent.

IV. Structure of the Optimal Input

Before attempting to obtain the optimal input (i.e., the one that achieves the RCEE), a few remarks can be made about its structure. It is shown in [4] that the channel capacity for $M > T$ is the same as for $M = T$; it does not pay to have more transmitter antennas M than the fading coherence time T. This result, however, does not immediately imply that we cannot gain advantages in coding by having $M > T$. However, we now show that the random coding error exponent does not improve by increasing T beyond M.

A. Error exponent for $M > T$ equals exponent for $M = T$

We observe in the expression of the conditional probability density (2) that the effect of the transmitted $T \times M$ signal S is only through the $T \times T$ matrix $SS^\dagger$. Hence, the RCEE depends on the signal S only through $SS^\dagger$. Therefore, even if $M > T$ we can expect to achieve no better error exponent than with $M = T$ transmitter antennas.

Theorem 1: The random coding error exponent obtained with $M > T$ transmitter antennas can be also obtained by using only $M = T$ antennas.

Proof: It remains to show only that for every $T \times M$ matrix S with $T > M$, there exists a $T \times T$ matrix yielding the same product $SS^\dagger$ and obeying the power constraint (3). This is done in the proof of Theorem 1 of [4]. ■

Consequently, we assume in the remaining sections that $M \leq T$. Note that when $T = 1$ there is no point in using more than $M = 1$ transmitter antenna, a case that is studied by Richters in [6]. Although the following results are valid for all values of T, we are mainly interested in $T > 1$.

B. The optimal input

Theorem 2: The optimal input that achieves the random coding error exponent can be written as

$$S = \Phi V, \tag{8}$$

where Φ is a $T \times T$ isotropically distributed matrix, and V is an independent $T \times M$ real, diagonal, non-negative matrix. Furthermore, V can be chosen such that its density is symmetric in its diagonal elements.

An *isotropically distributed* unitary matrix is defined to have a probability density that is unchanged when the matrix is multiplied by any deterministic unitary matrix. The columns of the matrix, while being orthonormal, are equally likely to point anywhere in the T-dimensional complex space. Details about the isotropic distribution may be found in [4].

Proof: The conditional probability density for our channel is:

$$p(X|S) = \frac{\exp\left(-\mathrm{tr}\,\left[[I_T + \frac{\rho}{M} SS^\dagger]^{-1} XX^\dagger\right]\right)}{\pi^{TN} \det{}^N [I_T + \frac{\rho}{M} SS^\dagger]}. \tag{9}$$

For any random unitary $M \times M$ matrix Ψ and any deterministic unitary $T \times T$ matrix Φ, we have

$$p(X|S\Psi^\dagger) = p(X|S), \tag{10}$$

$$p(\Phi X|\Phi S) = p(X|S). \tag{11}$$

Let $S = \Phi V \Psi^\dagger$ be a singular value decomposition of S. Then, using (10), we have

$$F_o(\mu, P(S), r) = F_o(\mu, P(\Phi V), r). \tag{12}$$

Therefore, we may restrict our input to have the form ΦV, where Φ is $T \times T$ unitary and V is $T \times M$ diagonal, real, and non-negative. Note, for the moment, that Φ and V are not necessarily independent.

By changing integration variables in equation (11), we see that

$$F_o(\mu, P(\Phi V), r) = F_o(\mu, P(\Theta \Phi V), r), \tag{13}$$

for any deterministic unitary $T \times T$ matrix Θ.

Lemma 1: For any given value of μ and r, $F_o(\mu, P, r)$ is convex in the probability distribution P.

Proof: See Appendix I.

Let us consider $F_o(\mu, P(\Theta \Phi V), r)$, where Θ is isotropically distributed. By Lemma 1,

$$F_o(\mu, P(\Theta \Phi V), r) \leq E\left[F_o(\mu, P(\Theta \Phi V | \Theta), r)\right]. \tag{14}$$

Since Θ is unitary, it follows from (11) that $F_o(\mu, P(\Theta \Phi V | \Theta), r) = F_o(\mu, P(\Phi V), r)$. Moreover, $\Theta\Phi$ conditioned on Φ is also isotropically distributed, and independent of Φ (and V). Consequently, the optimal input has the form ΦV where Φ is isotropically distributed and V is an independent diagonal, real, non-negative matrix. ■

The important aspect of Theorem 2 is that it does not only provides us with the density of Φ, it establishes that Φ and V are independent. Consequently, equation (6) can be significantly simplified by carrying on the integration over Φ. Our problem may now be formulated as minimizing $F_o(\mu, P_V, r)$ (which depends on S now only through V) with respect to P_V, the probability distribution of V.

Interestingly, the structure for S given above coincides with the capacity-achieving structure for S given in [4]. It is there shown that the diagonal elements $v_1, \ldots, v_M$ of V are constrained to satisfy $E v_m^2 \leq T$. The next section looks at the RCEE-achieving distribution of V.

V. The Optimality Condition

Integrating over Φ (as mentioned in the previous section) is very tedious unless $M = 1$ and $N = 1$ (one transmitter and one receiver antenna), where a closed-form expression of $F_o(\mu, P_V, r)$ can be obtained. In this case V is a $T \times 1$ nonnegative diagonal matrix; V therefore may be parameterized by a scalar, which we call v. We omit the details of the integration, but the channel probability density can be simplified as in [4] because it depends on X through only one non-zero eigenvalue λ. The result is

$$F_o(\mu, P_v, r) = \int_0^\infty d\lambda p(\lambda) \left[\int dP_v(v) f(\mu, v, r) \right]^{1+\mu}, \tag{15}$$

where,

$$f(\mu, v, r) = \frac{(T-1)e^a e^{r(v^2-T)} \gamma(T-1, a)}{a^{T-1}(1+\rho v^2)^{\frac{1}{1+\mu}}}, \quad p(\lambda) = \frac{e^{-\lambda}\lambda^{T-1}}{, (T)} \tag{16}$$

$\gamma(T, z)$ is the incomplete gamma function

$$\gamma(T, z) = \int_0^z q^{T-1} e^{-q} dq, \text{ and } a = \frac{\lambda \rho v^2}{(1+\mu)(1+\rho v^2)}. \tag{17}$$

Let the RCEE-achieving optimal input v be denoted v^*. Unfortunately, we are not able to solve explicitly for the statistics of v^*. We do, however, determine a necessary and sufficient condition that v^* must satisfy. We establish that v^* is discrete with a finite number of mass points. Our results closely parallel those of Abou-Faycal, Trott and Shamai [7], and Smith [8].

Lemma 2: For a given μ, P_v^* and r are optimal if and only if the following Kuhn-Tucker condition is satisfied:

$$\begin{cases} (1+\mu) \int_0^\infty d\lambda p(\lambda) \left[\int dP_v^*(v) f(\mu, v, r) \right]^\mu f(\mu, v, r) \geq K, \ \forall v \\ \text{and} \\ (1+\mu) \int_0^\infty d\lambda p(\lambda) \left[\int dP_v^*(v) f(\mu, v, r) \right]^\mu f(\mu, v, r) = K, \ \forall v \in E_v^* \end{cases} \tag{18}$$

where, E_v^* is the set of points of increase of P_v^*, and

$$K = (1+\mu) \int_0^\infty d\lambda p(\lambda) \left[\int dP_v^*(v) f(\mu, v, r) \right]^{1+\mu},$$

and $p(\lambda)$ and $f(\mu, v, r)$ are defined in (16).

Proof: The proof is identical to the proof for $T = 1$ given in [7], and is not duplicated here. ■

Since v appears in $F_o(\mu, P_v, r)$ only through its square, we find it easier to perform the optimization as a function of $\alpha = v^2$, a real non-negative random variable. Equation (18) is combined with (16) and (17) to yield

$$\int_0^\infty d\lambda [h(\lambda)]^\mu e^{-\lambda/(1+\rho\alpha)} \gamma(T-1, \frac{\lambda\rho\alpha}{(1+\rho\alpha)}) \geq K' \frac{(1+\rho\alpha)^{1/(1+\mu)} (\rho\alpha)^{T-1} e^{-r\alpha}}{(1+\rho\alpha)^{T-1}}, \qquad (19)$$

where,

$$h(\lambda) = \int dP_\alpha^* e^{\frac{-\lambda}{(1+\mu)}} f(\mu, \alpha, r), \qquad (20)$$

and K' is a constant that depends on T and μ. The inequality (19) holds for every $\alpha \geq 0$, with equality on the support of the optimal input α^*.

Using ideas and techniques developed in [7] and explained in the next section, we prove that the optimum input v^* is discrete with a finite number of mass points, and then find its distribution numerically. Since the Kuhn-Tucker condition (19) is necessary and sufficient for optimality, any numerical results can be checked for true optimality by verifying that (19) is satisfied.

VI. The Discrete Character of v^*

When $M = N = 1$, the optimal input v^* is a scalar and must have exactly one of the following properties:

1. Its support contains an interval,
2. It is discrete, with an infinite number of mass points on *some* bounded interval,
3. It is discrete and infinite, but with only a finite number of mass points on *any* bounded interval, or
4. It is discrete with a finite number of mass points.

Following [7], [8] and [9], we use the Kuhn-Tucker condition to prove that the first three cases are impossible and that therefore Case 4 prevails.

A. A positive accumulation point

Assume that Case 1 or 2 holds. Then the support of v^* includes a bounded infinite set of distinct points $\mathcal{S} \subseteq [0, A]$. The interval $[0, A]$ is compact; hence by the Bolzano-Weierstrass

theorem the set $\mathcal{S}$ has an accumulation point in $[0, A]$.

Define $g(z)$ as a function of the complex variable z to be

$$g(z) = -,\ (T)e^{-z}\left[\frac{z^{T-1}}{(T-1)!} + \frac{z^{T-2}}{(T-2)!} + \cdots + z + 1\right] + ,\ (T). \tag{21}$$

Clearly, $g(z)$ is analytic over the whole complex domain. Moreover, on the real axis $g(z)$ coincides with $\gamma(T-1, z)$ defined in (17).

Using equation (19), we define the function of the complex variable z

$$c(z) = \int_0^\infty d\lambda [h(\lambda)]^\mu e^{-\lambda/(1+\rho z)} g(\frac{\lambda\rho z}{1+\rho z}) - K' \frac{(1+\rho z)^{1/(1+\mu)}(\rho z)^{T-1}e^{-rz}}{(1+\rho z)^{T-1}}, \tag{22}$$

which is analytic over the whole complex domain except for real $z \in (-\infty, -1/\rho]$ (using the principal branch of the logarithm). The Kuhn-Tucker condition says that $c(z) = 0$ for z in the support of v^*. Under Cases 1 or 2, $c(z)$ is therefore zero over a set with an accumulation point inside the domain. The identity theorem [10] therefore implies that the function is zero on the whole domain, particularly on the real non-negative axis. We thus have equality in (19) for all $\alpha \geq 0$. We examine the consequences of this.

The left hand term (LHT) of (19) has terms that obey

$$\gamma(T-1, \lambda\rho\alpha/(1+\rho\alpha)) \ \rightarrow \ \gamma(T-1, \lambda), \tag{23}$$

$$e^{-\lambda/(1+\rho\alpha)} \ \rightarrow \ 1, \tag{24}$$

as $\alpha \rightarrow \infty$. Hence, using Lebesgue's dominated convergence theorem

$$\text{LHT} \rightarrow \int_0^\infty d\lambda [h(\lambda)]^\mu \gamma(T-1, \lambda), \tag{25}$$

clearly a positive number because $\gamma(T-1, \lambda)$ and $h(\lambda)$ are both positive for $\lambda > 0$. On the other hand, because r is positive, the right hand term (RHT) goes to zero as $\alpha \rightarrow \infty$. But because the LHT is positive, the conclusion that $c(z)$ is identically zero is contradicted; hence Cases 1 and 2 are excluded.

B. No accumulation points

Assume now that Case 3 holds; that is, the optimal α^* is discrete with an infinite number of mass points, but with only a finite number of them on any bounded interval.

Let $\{\alpha_i\}$ be an ordered sequence of points in the support of α^*, with the property $\alpha_i \to \infty$ as $i \to \infty$. As in the previous argument, equation (19) implies that we have equality for all α_i. But it has been proven in the previous section that while the LHT goes to a positive number for large values of α_i, the RHT goes to zero, so equality cannot be achieved for arbitrarily large α_i; this again is a contradiction.

C. Finite number of mass points

Cases 1, 2, and 3 are ruled out, and we therefore conclude that α^* is discrete with finitely many mass points.

We believe that for more than one transmitter antenna ($M > 1$) or receiver antenna ($N > 1$), the RCEE-achieving $v_1, \ldots, v_M$ (diagonal elements of V) are jointly discrete random variables. We know that their joint distribution is symmetric, but little else has been proven.

VII. One Impulse Distribution

In this section we study the behavior of the optimal input distribution P_v^* when the SNR ρ grows to infinity. We show that $v^* \to \sqrt{T}$ as $\rho \to \infty$.

Theorem 3: As $\rho \to \infty$, v^* converges in distribution to a deterministic random variable with one mass point at $\sqrt{T}$.

Proof: See Appendix II.

Numerical results presented in the next section show that v^* converges to $\sqrt{T}$ for even relatively small values of ρ. We note a similar result holds when $T \to \infty$. In fact numerical results show also that $v^* = \sqrt{T}$ is optimal also for finite values of T. The proof of this is difficult and is omitted. (It turns out that $v^* = \sqrt{T}$ is asymptotically optimal also for capacity calculations when either $\rho \to \infty$ or $T \to \infty$; proofs may be found in [4] and [12]).

With Theorem 3, we have characterized the optimal input. Indeed, in the appropriate ranges of T and ρ, we know that for $M = N = 1$, the optimal input S is $\sqrt{T}$ times an isotropically distributed vector. Remarkably, we do not need to use large values of SNR or T for this S to be optimal. For example, when $T = 2$, we find that for $\rho > 17$ dB, $v^* = \sqrt{T}$ is optimal. For $T = 5$, we find that $\rho > 8$ dB. Similarly, fixing $\rho = 10$ dB, we find that $T > 4$ yields $v^* = \sqrt{T}$. These numerical results are explained in more detail in

the next section.

VIII. Numerical Results

Having characterized the optimal input distribution for $M = N = 1$, we can now compute numerically the random coding error exponent of the channel. We use the following algorithm:

- Choose a value of μ and r.
- Find the locations and the probabilities of the masses using a conjugate gradient method.
- Confirm that the necessary and sufficient optimality condition is satisfied.
- Compute the corresponding power.

For a fixed value of μ, we can vary the parameter r in order to obtain a specified power level. For this power, the random coding exponent is then obtained by collecting the values for different μ, and then maximizing $E_o(\mu, P_v, r) - \mu R$ for different values of R, where $E_o(\mu, P_v, r) = -\ln F_o(\mu, P_v, r)$ and F_o is given in (15).

Although the optimization problem is convex over the space of all input probability distributions , there is no reason to believe it remains convex when parameterized by the mass point probabilities and locations. Furthermore, we have no bounds on the number of mass points needed as a function of the power constraint.

As a practical matter these potential difficulties do not seem to arise. Because $F_o(\mu, P, r)$ is continuous and strictly convex in the input distribution, the optimal input distribution changes continuously (in the weak* topology) with the SNR ρ. Also, we have found empirically that two mass points are optimal for low SNR and that for values of SNR high enough, one mass point is optimal. These conclusions are supported by the Kuhn-Tucker condition (19), which, being necessary and sufficient for optimality, allows us to establish (up to the resolution of the numerical algorithms) that a local maximum found by a descent method is also a global maximum. To apply the Kuhn-Tucker test to a postulated α^*, we plot the LHT of (19) minus its RHT as a function of α. The resulting graph must be nonnegative and must touch zero at the atoms of α^*, as in Fig. 2.

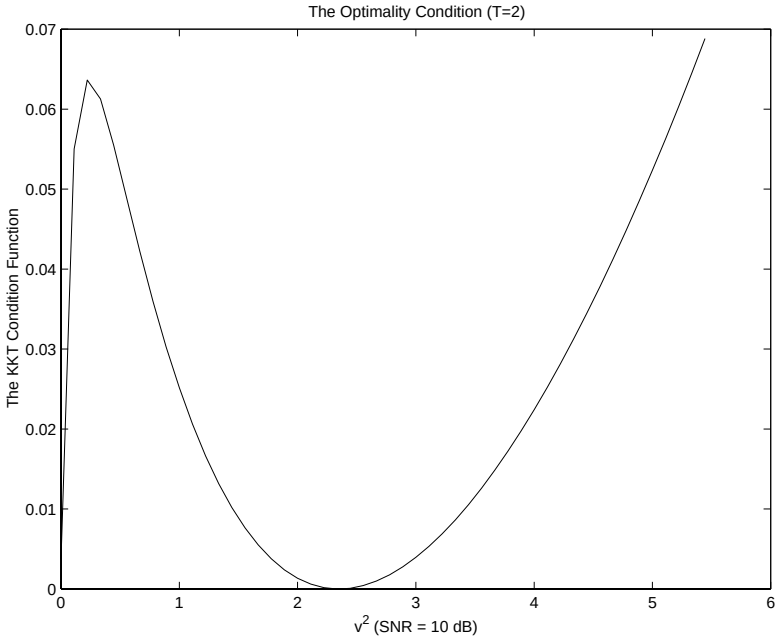


Fig. 2. This figure displays the LHT minus the RHT of equation (19) for SNR = 10 dB, $T = 2$, and $\mu = 1$. The optimal input consists of two mass points, one at zero with mass 0.21 and another at $v^2 = 2.33$ with mass 0.79.

A. Random coding exponents

Using the algorithm described above, it is possible to obtain the random coding error exponent for any given value of T and SNR as is shown in Figures 3, 4, and 5, where E_r is plotted as a function of the normalized rate R/T for different values of coherence time T, and SNR ρ.

For a given block length L, these plots enable us to upper-bound the probability of error that can be achieved for any rate. Conversely, given a target probability of error, we can obtain an estimate of the minimum block length one should use to achieve it. We discuss this relationship more thoroughly in Section IX.

B. Random coding exponent as function of T

Another aspect of interest is the dependence of the RCEE E_r on T. We examine the behavior of E_r as a function of T when the rate per channel use $R_o = R/T$ is held fixed; that is, the code rate R increases linearly with T. We ask, does E_r increase to infinity or remain bounded as $T \to \infty$?

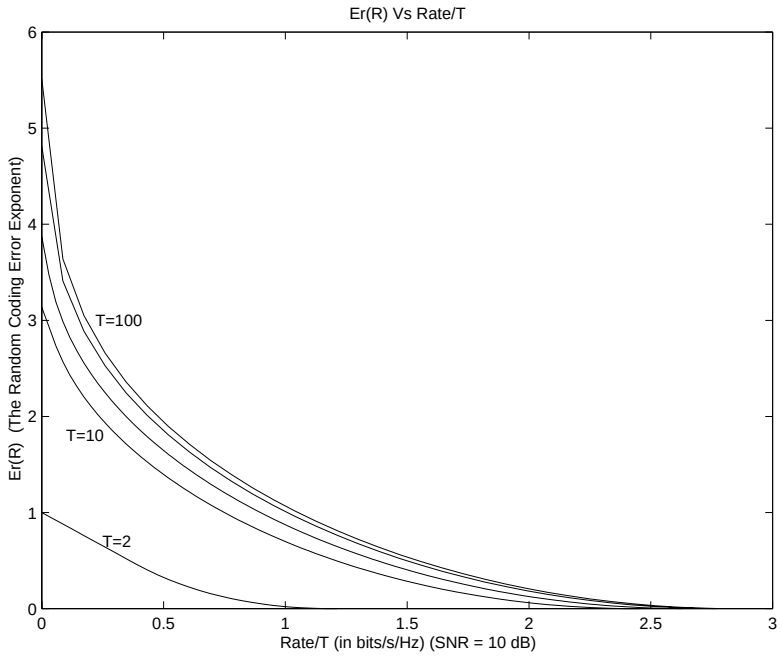


Fig. 3. Random coding error exponent as a function of rate for $M = N = 1$, SNR = 10 dB, and $T = 2$, 10, 20, 50, 100

For any fixed R/T and ρ, Figures 3, 4, and 5 suggest that E_r converges to a finite value for any R_o as T increases. In the next section we will formally establish this result and analyze its significance.

IX. Minimum Number of Independent Fades

It is established in [4] that the capacity of our unknown-fading channel converges to the capacity of the equivalent known-fading channel as T increases (or the fading rate slows). We would like to establish what effect increasing T has on the error exponent.

As T increases, the transmitted $T \times 1$ complex signals get longer, allowing us to construct longer codewords, but also increasing the chances that we will be stuck for a long time in a deep fade, at which point decoding errors are easily made. Therefore, it is reasonable to expect that a minimum number of independent realizations of the fading are always needed (independently of T) to be able to decode to a certain given probability of error. Hence, we expect the exponent to remain bounded. The numerical results presented in the previous section suggest this; Figures 3, 4, and 5 show that E_r appears to converge as $T \to \infty$.

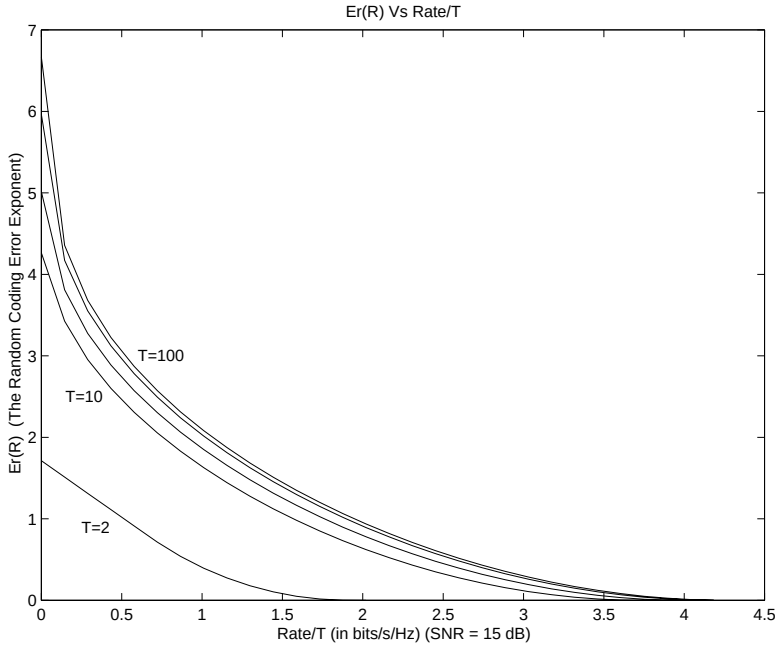


Fig. 4. Random coding error exponent as a function of rate for $M = N = 1$, SNR = 15 dB, and $T = 2$, 10, 20, 50, 100

It turns out that the behavior of the optimal μ as a function of T is critical to showing that E_r converges. We know that when $T \to \infty$ the optimal v is numerically found to be deterministic, $v^* = \sqrt{T}$ (see Section VII). Hence, we look at $E_o(\mu, P_v, r)$ when v is a unit mass at $\sqrt{T}$. Since, in this case, E_o is independent of r, we write $E_o(\mu)$ to emphasize the dependence on μ.

Recall that

$$E_r(R) = E_o(\mu^*) - \mu^* R. \tag{26}$$

We wish to study the behavior of E_r when the rate per signal (each signals comprises T symbols) is linearly proportional to T. Hence, let $R = R_o T$ for some fixed R_o. Equation (26) can now be written $E_r(R) = E_o(\mu^*) - \mu^* T R_o \geq 0 = E_o(0)$.

Theorem 4: $E_r(R)$ converges to a constant as $T \to \infty$.

Proof: See Appendix III.

Theorem 4 implies that, even for large T, we need to code over a minimum number of signals to achieve a given probability of error. To help see this, consider channel code rates close to capacity. For these rates, the random coding bound is generally tight because a

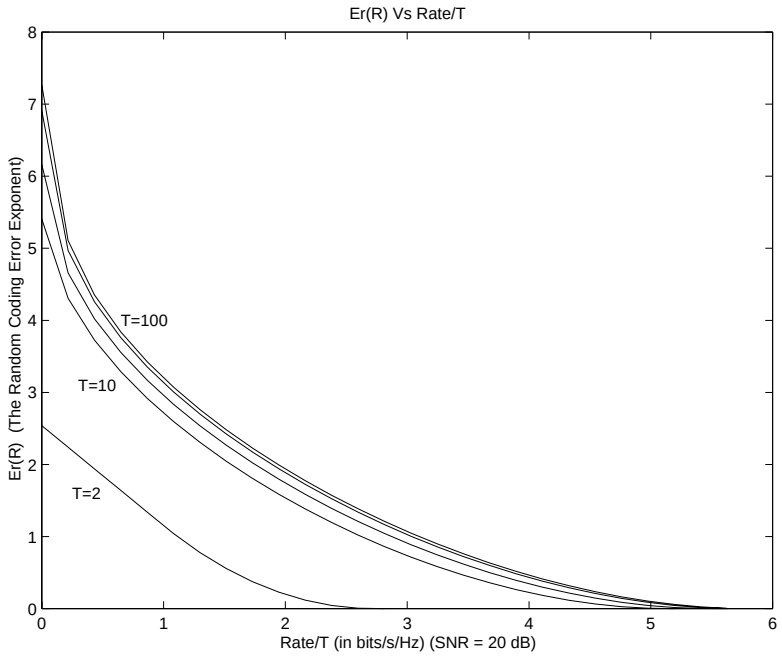


Fig. 5. Random coding error exponent as a function of rate for $M = N = 1$, SNR = 20 dB, and $T = 2$, 10, 20, 50, 100

random codebook generally performs well. We therefore also expect the behavior of the best code to have the same exponential dependence on the block length L as the random coding bound, and Theorem 4 says that this exponential dependence approaches a limit as $T \to \infty$.

Assuming that the error exponent is tight allows us to obtain a rough estimate on the number of signals over which we should code to achieve a given probability of error, for a fixed data rate and a fixed SNR. Figure 6 shows the minimum block length L required to achieve a target probability of error when the SNR = 10 dB and the rate is equal to half capacity ($R/T = 1.45$ bits/channel use). The curves show that, even for very large values of T, we need to code over relatively many signals to achieve any given probability of error.

In practice, it is well known that channel codes generally need to observe a certain number of independent realizations of the fading coefficients to obtain low probability of error. When the fading rate is slow (T is large), symbols are often interleaved so that the effective fading rate is increased and successive symbols have approximately independent

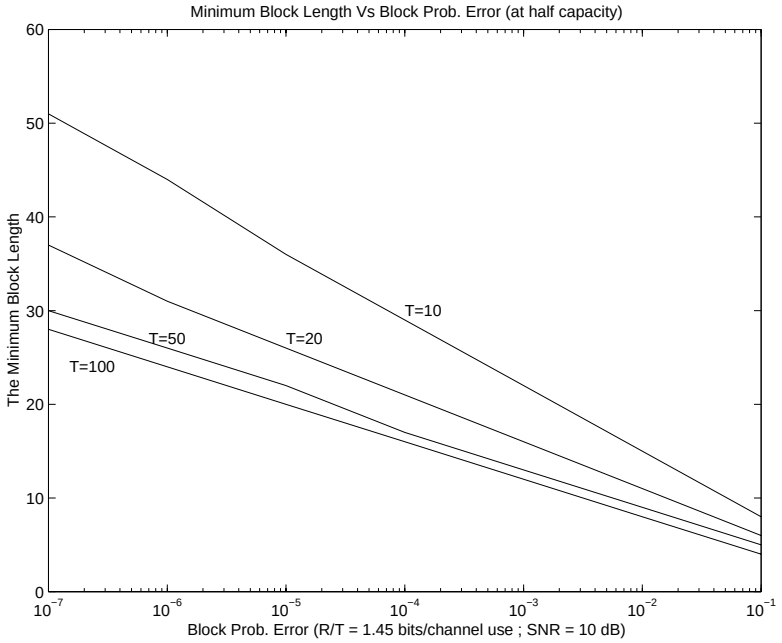


Fig. 6. Minimum block length L needed to get prescribed error probability for $M = N = 1$, SNR = 10 dB, R/T=1.45 bits/channel use, and $T = 10, 20, 50, 100$

fades. The number of successive symbols that are needed to form an effective channel code is then given approximately by the minimum L we obtain in Figure 6.

X. Conclusion

In this study we looked at the Rayleigh fading multiple-antenna channel when the receiver does not know the fading and does not attempt to estimate it. We computed the channel's Random Coding Error Exponent (RCEE) which provided us with an upper bound on the achievable error probability. We showed that many of the characteristics of the capacity-achieving input are found in the RCEE-achieving input.

Indeed, we showed that, as with channel capacity, there is no advantage in having more transmitter antennas M than the length of the coherence time T. We also established that the optimal input consists of vectors orthogonal with respect to time, with directions that are isotropically distributed and that are independent of their amplitudes.

When there is one transmitter and receiver antenna, we showed that the amplitudes are discrete, and for high values of SNR they are, in fact, deterministic. Numerical evidence suggested that the amplitudes are deterministic for even relatively low values of T and

SNR.

Finally, we studied the behavior of the RCEE when $T \to \infty$. We found that the exponent is upper-bounded for any rate $R = R_o T$, which implies that, even when $T \to \infty$ we still need to code over multiple T-symbol blocks to obtain a low probability of error.

Because of the complexity of some of the computations, some of our results were restricted to one transmitter and receiver antenna. While it is intuitively reasonable to expect some of these results to extend to more than one antenna, a rigorous study is a possible direction for future work.

Acknowledgments

We thank Thomas Marzetta of Bell Laboratories for numerous helpful discussions.

Appendices

I. Convexity of F_o

Lemma 1: For any given value of μ and r, $F_o(\mu, P, r)$ is convex in P.

Proof: Let $g(S) = E[\mathrm{tr}\, SS^\dagger] - TM$. Recall that

$$F_o(\mu, P, r) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} p(X|S)^{1/(1+\mu)} e^{rg(S)} dP(S) \right]^{1+\mu} dX.$$

Note that the function $t^{1+\mu}$ is convex since $\mu \geq 0$. Hence,

$$\begin{aligned} F_o(\mu, \lambda P + (1-\lambda)G, r) &= \int \left[\int p(X|S)^{1/(1+\mu)} e^{rg(S)} d(\lambda P(S) + (1-\lambda)G(S)) \right]^{1+\mu} dX \\ &\leq \lambda \int \left[\int p(X|S)^{1/(1+\mu)} e^{rg(S)} dP(S) \right]^{1+\mu} dX + \\ &\quad (1-\lambda) \int \left[\int p(X|S)^{1/(1+\mu)} e^{rg(S)} dG(S) \right]^{1+\mu} dX \\ &= \lambda F_o(\mu, P, r) + (1-\lambda) F_o(\mu, G, r), \end{aligned}$$

and $F_o(\mu, P, r)$ is convex in P. ■

II. Asymptotic Behavior when SNR Goes to Infinity

Theorem 3: As $\rho \to \infty$, v^* converges in distribution to a deterministic random variable with one mass point at $\sqrt{T}$.

Proof: When $v = \sqrt{T}$ with probability one,

$$\int dP_v(v) f(\mu, v, r) = \frac{(T-1)e^a \gamma(T-1, a)}{a^{T-1}(1+\rho T)^{\frac{1}{1+\mu}}}, \tag{27}$$

where,

$$a = \frac{\lambda \rho T}{(1+\mu)(1+\rho T)}. \tag{28}$$

For $v = \sqrt{T}$ to be optimal, it should satisfy the KKT optimality condition (18), which can be written in the form

$$\int d\lambda p(\lambda) \left[\int dP_v(v) f(\mu, v, r)\right]^{\mu} \left[\frac{e^b e^{r(\alpha-T)} \gamma(T-1, b)}{b^{T-1}(1+\rho\alpha)^{1/1+\mu}} - \frac{e^a \gamma(T-1, a)}{a^{T-1}(1+\rho T)^{1/1+\mu}}\right] \geq 0, \tag{29}$$

with the inequality becoming an equality when $\alpha = T$ (recall that $\alpha = v^2$), and where

$$b = \frac{\lambda \rho \alpha}{(1+\mu)(1+\rho\alpha)}.$$

We would like to show that equation (29) is satisfied asymptotically in ρ.

As $\rho \to \infty$, $a \to \lambda/(1+\mu)$. Similarly, $b \to \lambda/(1+\mu)$. Hence, equation (29) is to the first order in $1/T$ equivalent to

$$\int d\lambda p(\lambda) \left[\int dP_v(v) f(\mu, v, r)\right]^{1+\mu} \left[\frac{e^{r(\alpha-T)}}{(1+\rho\alpha)^{1/1+\mu}} - \frac{1}{(1+\rho T)^{1/1+\mu}}\right] \geq 0, \tag{30}$$

for every α, with equality for $\alpha = T$.

We now look more closely at the difference term in the integrand. Since, when $\alpha = T$, any r satisfies the optimality condition, we choose

$$r = \frac{\rho}{(1+\mu)(1+\rho T)}. \tag{31}$$

Consequently,

$$\frac{d}{d\alpha}\left[\frac{e^{r(\alpha-T)}}{(1+\rho\alpha)^{1/1+\mu}} - \frac{1}{(1+\rho T)^{1/1+\mu}}\right]$$

has the same sign as

$$\frac{\rho}{1+\mu}\left[\frac{1+\rho\alpha}{1+\rho T} - 1\right],$$

which is negative for $\alpha < T$ and positive for $\alpha > T$. Since the function is zero at $\alpha = T$, the integrand is non-negative for all α, zero for $\alpha = T$, and hence so is its integral. We conclude that $v = \sqrt{T}$ satisfies the KKT conditions asymptotically in ρ, and is therefore asymptotically optimal. ■

III. Behavior of E_r as $T \to \infty$

In this appendix we look at the behavior of the exponent when $T \to \infty$. Since the capacity results [4] and numerical simulations strongly suggest that a deterministic $v = \sqrt{T}$ is asymptotically optimal in T (in fact numerical results show that this v is optimal even for small finite values of T), we look at $E_o(\mu, P_v, r)$ when $v = \sqrt{T}$. With this v, the power constraint is automatically satisfied and the parameter r is irrelevant. Consequently, the simpler notation $E_o(\mu)$ is used.

Theorem 4: $E_r(R)$ converges to a constant as $T \to \infty$.

Proof: We first look at the asymptotic behavior of $F_o = \exp\{-E_o(\mu)\}$ as $T \to \infty$. When $v = \sqrt{T}$, equations (15), (16) and (17) yield

$$\begin{aligned}
F_o(\mu) &= \frac{(T-1)^{1+\mu}(1+\mu)^{(1+\mu)(T-1)}(1+\rho T)^{(T-1)}}{,\,(T)(\rho T)^{(1+\mu)(T-1)}} \int_0^\infty \frac{e^{-\lambda}}{\lambda^{(T-1)\mu}} \gamma^{1+\mu}(T-1, \frac{\lambda\rho T}{1+\mu}) d\lambda \\
&\geq \frac{(T-1)^{1+\mu}(1+\mu)^{(1+\mu)(T-1)}(1+\rho T)^{(T-1)}}{,\,(T)(\rho T)^{(1+\mu)(T-1)}} \int_{1/\rho}^\infty \frac{e^{-\lambda}}{\lambda^{(T-1)\mu}} \gamma^{1+\mu}(T-1, \frac{\lambda\rho T}{1+\mu}) d\lambda \\
&\geq \frac{(T-1)^{1+\mu}(1+\mu)^{(1+\mu)(T-1)}(1+\rho T)^{(T-1)}}{,\,(T)(\rho T)^{(1+\mu)(T-1)}} \gamma^{1+\mu}(T-1, \frac{T}{1+\mu}) \int_{1/\rho}^\infty \frac{e^{-\lambda}}{\lambda^{(T-1)\mu}} d\lambda \\
&\geq \frac{(T-1)^{1+\mu}(1+\mu)^{(1+\mu)(T-1)}(1+\rho T)^{(T-1)}}{,\,(T)(\rho T)^{(1+\mu)(T-1)}} \gamma^{1+\mu}(T-1, \frac{T-1}{1+\mu}) \int_{1/\rho}^\infty \frac{e^{-\lambda}}{\lambda^{(T-1)\mu}} d\lambda.
\end{aligned} \tag{32}$$

We study the behavior of the entire right hand term (RHT) in equation (32) when T goes to infinity. Note that

$$\begin{aligned}
\frac{(1+\rho T)^{(T-1)}}{(\rho T)^{(T-1)}} &= O(1) \\
\frac{1}{\rho^{\mu(T-1)}} \int_{1/\rho}^\infty \frac{e^{-\lambda}}{\lambda^{(T-1)\mu}} d\lambda &= O\left(\frac{1}{1/\rho + (T-1)\mu}\right)
\end{aligned}$$

as $T \to \infty$. Moreover, it has been established by Temme [11] that

$$\gamma(T-1, \alpha(T-1)) \sim \frac{1}{2}\mathrm{erfc}\,(\xi\sqrt{T-1}),\ (T-1),$$

uniformly in α, where $\xi = \sqrt{\alpha - 1 - \ln\alpha}$, and erfc (x) is the complementary error function defined by

$$\mathrm{erfc}\,(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \tag{33}$$

Let us look at all the different possibilities for μ:

1. If μ decreases strictly slower than $1/\sqrt{T}$ (meaning that $\mu\sqrt{T} \to \infty$ as $T \to \infty$), then $\xi\sqrt{T-1} \to \infty$, which implies that

$$\gamma(T-1, \alpha(T-1)) \;=\; O\left(\frac{e^{-\xi^2(T-1)}}{2\sqrt{\pi}\xi\sqrt{T-1}}, (T-1)\right).$$

The RHT of equation (32) is then

$$O\left(\frac{, (T)^{\mu} e^{(T-1)\mu}}{T^{\mu(T-1)}(2\sqrt{\pi})^{1+\mu}\left[(T-1)\left(\frac{-\mu}{1+\mu} + \ln(1+\mu)\right)\right]^{(1+\mu)/2}(T-1)\mu}\right)$$

$$= \; O\left(\frac{1}{e^{\mu}2^{\mu/2}\sqrt{T-1}\left(\frac{-\mu}{1+\mu} + \ln(1+\mu)\right)^{(1+\mu)/2}(T-1)\mu}\right).$$

Therefore, $E_o = O\left(c_1 \ln \mu + c_2 \ln T\right)$ for some positive constants c_1 and c_2.

2. If $\mu = O(1/\sqrt{T})$, then

$$\gamma(T-1, \alpha(T-1)) \;=\; O\left(, (T-1)\right),$$

and the RHT of equation (32) is

$$O\left(\frac{, (T)^{\mu}(1+\mu)^{(1+\mu)(T-1)}}{T^{\mu(T-1)}\left[1/\rho + (T-1)\mu\right]}\right) = O\left(\frac{1}{1/\rho + (T-1)\mu}\right).$$

Therefore,

(a) if μ decreases strictly slower than $1/T$, then $E_o = O(\ln \mu T)$.

(b) if $\mu = O(1/T)$, then $E_o = O(1)$.

Recall that $R = R_o T$, and (26) therefore becomes

$$E_r(R) = E_o(\mu^*) - \mu^* T R_o \geq 0 = E_o(0). \tag{34}$$

If μ^* satisfies Cases 1 or 2(a), then it is straightforward to see that the exponent $E_r(R)$ becomes negative for T sufficiently large, which contradicts the inequality in equation (34). Consequently, Case 2(b) prevails and $\mu^* = O(1/T)$, and $E_o(\mu^*) = O(1)$. Combining this conclusion with equation (34) yields that $E_r(R)$ converges to a constant as $T \to \infty$. ■

References

[1] 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.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Laboratories Tech. Rep., 1995.

[3] 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, no. 2, pp. 744–765, Mar. 1998.

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

[5] Robert G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Info. Theory*, vol. 11, pp. 3-17, January 1965.

[6] J. S. Richters, "Communication over fading dispersive channels," *MIT Research Laboratory of Electronics*, Technical Report 464, Nov. 30, 1967.

[7] I. Abou-Faycal, M. Trott and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh fading channels," in *Proc. Int. Symp. Info. Th.*, Ulm, Germany, June-July 1997, p. 473.

[8] J.G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inform. Contr.*, vol. 18, pp. 203-219, 1971. See also "On the Information Capacity of Peak and Average Power Constrained Gaussian Channels," Ph.D. dissertation, Dep. Elec. Eng., Univ. of California, Berkeley, CA, 1969.

[9] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Info. Theory*, vol. 41, no. 4, pp. 1060–1071, July 1995.

[10] H. Silverman, *Complex Variables*, Boston: Houghton Mifflin Company, 1975.

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

[12] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," Lucent Technologies Bell Laboratories Tech. Rep., 1998.