Capacity and Reliability Function for Small Peak Signal Constraints

Bruce Hajek, *Fellow, IEEE,* and Vijay G. Subramanian, *Member, IEEE*

***Abstract*—The capacity and the reliability function as the peak constraint tends to zero are considered for a discrete-time memoryless channel with peak constrained inputs. Prelov and van der Meulen showed that under mild conditions the ratio of the capacity to the squared peak constraint converges to one-half the maximum eigenvalue of the Fisher information matrix and if the Fisher information matrix is nonzero, the asymptotically optimal input distribution is symmetric antipodal signaling. Under similar conditions, it is shown in the first part of the paper that the reliability function has the same asymptotic shape as the reliability function for the power-constrained infinite bandwidth white Gaussian noise channel. The second part of the paper deals with Rayleigh-fading channels. For such channels, the Fisher information matrix is zero, indicating the difficulty of transmission over such channels with small peak constrained signals. Asymptotics for the Rayleigh channel are derived and applied to obtain the asymptotics of the capacity of the Marzetta and Hochwald fading channel model for small peak constraints, and to obtain a result of the type of Médard and Gallager for wide-band fading channels.**

***Index Terms*—Fisher information, peak constraints, Rayleigh fading, reliability function, Shannon capacity.**

I. Introduction

CONSIDER a discrete-time memoryless channel with input alphabet equal to the n-dimensional Euclidean space $\mathcal{R}^n$ for some n and output space an arbitrary measurable space Ω. Assume that given a symbol x is transmitted, the output has density $q(y|x)$, relative to some fixed reference measure on Ω. Given $\epsilon > 0$, the peak constrained channel is obtained by restricting the input alphabet to the ball of radius ϵ in $\mathcal{R}^n$. For many channels this means that the energy of each transmitted symbol is constrained to ϵ^2. The peak constrained channel is itself a discrete memoryless channel, so that the channel capacity and reliability function are well defined. The focus of this paper is to study the asymptotic behavior of the capacity and reliability function as $\epsilon \to 0$. Prelov and Van der Meulen[18] showed that under mild regularity conditions the ratio of the capacity to the squared peak constraint converges to one-half the maximum eigenvalue of the Fisher information matrix. They also showed that if the Fisher information matrix is nonzero, the asymptotically optimal input distribution is equiprobable symmetric antipodal signaling. We prove, under a set of technical conditions somewhat different from those of [18], that the asymptotic behavior of both the capacity and channel reliability function can be identified. Two examples of the capacity result are given: application to a Rician channel and to a channel composed of parallel subchannels. We then examine the asymptotics of the capacity for the block Rayleigh-fading channel for which the Fisher information matrix is zero, and relate this example to the poor performance of nonbursty signaling schemes for certain broad-band fading channels.

The paper is organized into two parts, as follows. The first part of the paper consists of Sections II–VI. Section II presents the theorems giving the limiting normalized capacity and limiting normalized reliability function and two examples are considered. Preliminary implications of the regularity assumptions are given in Section III, and the theorem regarding normalized capacity is proved in Section IV. Upper and lower bounds on the optimal probability of error are given in Sections V and VI, respectively, yielding the proof of the theorem regarding normalized reliability. The upper bounds follow by random coding, while the lower bounds follow by sphere-packing, a low-rate bound, and straight-line bound. The expression for the limiting normalized capacity is the same as obtained in [18].

The second part of the paper consists of two sections dealing with Rayleigh-fading channel models. Section VII describes the asymptotic capacity of a multiple-antenna block Rayleigh-fading channel of the type considered in [12], and Section VIII applies the results of Section VII to provide additional understanding of the limitations of nonbursty spread-spectrum signaling over wide-band fading channels with sufficiently fast Rayleigh fading. Section VIII complements the work of [7], which constrains burstiness by constraining the fourth moments of the signal coordinates arising in a time–frequency decomposition of the transmitted signal. Here the peak signal energy in each time–frequency bin is constrained.

II. Capacity and Reliability Function for Small Peak Signal Constraints

Let $V(y, x)$ denote the gradient of $\log q(y|x)$ with respect to x and let $E_x[\cdot]$ denote expectation with respect to the probability measure on Ω with density $q(y|x)$. Let $||v||$ denote the Euclidean norm of a vector v. Let $x_0 \in \mathcal{R}^n$, $\bar{\epsilon}_0 > 0$, and $D_0 > 0$.

Manuscript received June 17, 2000; revised August 1, 2001. This work was supported by a Motorola Fellowship, by the U.S. Army Research Office under Grant DAAH05-95-1-0246, and by the National Science Foundation under Contracts NCR 93-14253, CCR 99-79381, and ITR 00-85929. The material in this paper was presented in part at the 2000 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 2000.

B. Hajek is with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: b-hajek@uiuc.edu).

V. G. Subramanian was with the University of Illinois at Urbana-Champaign. He is now with Motorola Corporation, Arlington Heights, IL 60004 USA (e-mail: vijay.subramanaian@motorola.com).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(02)01997-1.

Regularity Assumption $RA(x_0, \overline{\epsilon}_0, D_0)$: The function $\log q(y|x)$ is continuously differentiable in x with

$$||V(y, x) - V(y, \tilde{x})|| \leq L(y)u(||x - \tilde{x}||) \tag{1}$$

for some nondecreasing function $u(\cdot)$ such that $u(r) \to 0$ as $r \to 0$ and some measurable function L on Ω. Let $B(y, x) = L(y) + ||V(y, x)||$. Then

$$E_{x_0}[e^{\overline{\epsilon}_0 B(Y, x_0)}] \leq D_0.$$

Some remarks about the regularity assumption are in order. Condition (1) implies that the gradient $V(y, x)$ is uniformly continuous in x for each y. On the other hand, if $\log q(y|x)$ is twice continuously differentiable in x, then (1) is satisfied by taking $u(r) = r$ and $L(y)$ an upper bound on the spectral radius of the Hessian of $\log q(y|x)$ with respect to x. For the purpose of this paper, condition (1) need only be satisfied for x and $\tilde{x}$ in some neighborhood of x_0, but for ease of exposition it is assumed the condition holds for all x and $\tilde{x}$. Often in this paper the constant x_0 is taken to be zero.

Define C_ϵ to be the capacity of the channel with peak constraint ϵ, described above, and define $C_{\rm ss} = \lim_{\epsilon \to 0} \frac{C_\epsilon}{\epsilon^2}$ if the limit exists. Here "ss" stands for "small signal."

Theorem II.1: Suppose the regularity assumption RA $(0, \overline{\epsilon_0}, D_o)$ holds for some $\overline{\epsilon_0} > 0$ and some finite D_o. Then $C_{\rm ss}$ exists and $C_{\rm ss} = \frac{1}{2} I_0$, where I_0 is the maximum eigenvalue of the Fisher information matrix K for q evaluated at $x = 0$, given [2] by $K = E_0[V(Y, 0)V(Y, 0)^T]$.

Remarks: The investigation of channel capacity and mutual information in the limit of small signals has been of interest for a long time. The most closely related to this paper is that of [18]. See [18] for comments on early papers including [10], [11], and [17]. There is also a variety of more recent work involving information for certain random processes with small input signals [13]–[16].

The basic setting of Theorem II.1 is the same as that of [18]. Both assume that the density $q(x|y)$ is continuously differentiable in x. On the other hand, the exact technical conditions are rather difficult to compare, and in practice one or the other may be easier to verify. The conclusion of Theorem II.1 is the same as the conclusion of the corollary in [18] (with the parameter K of [18] set equal to 1). The same technical conditions as in Theorem II.1 are used in Theorem II.2, giving the asymptotics of the reliability function. The proof of Theorem II.1 helps prepare the reader for the similar proof of Theorem II.2.

Example (Rician Channel): Say that $Z = Z_{\rm re} + jZ_{\rm im}$ has the complex normal distribution with mean μ and variance var, and write $Z \sim \mathcal{CN}(\mu, {\rm var})$, if $Z_{\rm re}$ and $Z_{\rm im}$ are independent, Gaussian random variables with means ${\rm Re}(\mu)$ and ${\rm Im}(\mu)$, respectively, and with variance ${\rm var}/2$ each. Consider a discrete-time memoryless channel such that for input X in one channel use, the output Y is given by $Y = (\alpha + H)X + N$, where $H \sim \mathcal{CN}(0, \gamma^2)$, $N \sim \mathcal{CN}(0, \sigma^2)$, and $\alpha^2 + \gamma^2 = 1$. Without loss of generality it is assumed that $\alpha \geq 0$. This channel is the Rician-fading channel if $\alpha > 0$ and the Rayleigh-fading channel if $\alpha = 0$. This channel satisfies the conditions of Theorem II.1; in fact, $V(y, x)$ is given by

$$V(y, x) = 2\begin{pmatrix} \frac{\alpha y_{\rm re} - (\alpha^2+\gamma^2)x_{\rm re}}{\sigma^2+\gamma^2||x||^2} + \frac{\gamma^2 x_{\rm re}||y-\alpha x||^2}{(\sigma^2+\gamma^2||x||^2)^2} \\ \frac{\alpha y_{\rm im} - (\alpha^2+\gamma^2)x_{\rm im}}{\sigma^2+\gamma^2||x||^2} + \frac{\gamma^2 x_{\rm im}||y-\alpha x||^2}{(\sigma^2+\gamma^2||x||^2)^2} \end{pmatrix}$$

and satisfies (1) with $u(r) = r$. Therefore, $C_{\rm ss} = \frac{\alpha^2}{\sigma^2}$. If $\alpha = 0$ then $C_{\rm ss} = 0$. A higher order expansion for C_ϵ in this case is given in Section VII.

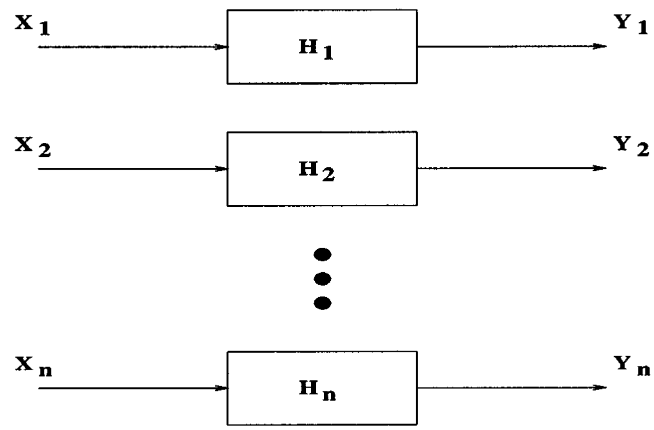

Fig. 1. A discrete-time memoryless channel.

Example (Parallel Subchannels): Consider the channel depicted in Fig. 1. The transmitter chooses an input $X = (X_1, X_2, \ldots, X_n)$ and each coordinate X_i is transmitted through a subchannel H_i to yield the output Y_i. The subchannels are statistically independent but are tied together through the peak constraint on the input vector X, namely, the requirement that $||X|| \leq \epsilon$ for some $\epsilon > 0$. This can model a remote-sensing scenario where a low-power measuring device sends measurements to a set of collection centers. Assume that each of the subchannels satisfies the conditions of Theorem II.1. The Fisher information matrix K for the overall channel is block diagonal, with the blocks being the Fisher information matrices for the subchannels. The maximum eigenvalue of K is thus the maximum of the eigenvalues of the blocks. Therefore, $C_{\rm ss}$ for the overall channel is the maximum of $C_{\rm ss}$ over the subchannels. Moreover, if $C_{\rm ss} > 0$ then an asymptotically optimal signaling scheme is to use only one of the subchannels (one with maximum value of $C_{\rm ss}$), and to use antipodal signaling on that subchannel.

Remark: Closely related to the capacity for small peak signal constraints is the notion of capacity per unit cost studied by Verdú [22] with the cost of an input symbol x being the energy $||x||^2$. The capacity per unit energy C_E is the supremum over ϵ of $\overline{C}_\epsilon/\epsilon^2$, where $\overline{C}_\epsilon$ is the capacity subject to the constraint that the average energy per symbol of each transmitted codeword be at most ϵ^2. Moreover, the supremum over ϵ is achieved as $\epsilon \to 0$ [22]. Every valid codeword in the definition of C_ϵ has peak energy per channel use at most ϵ, and therefore average energy per channel use at most ϵ^2, so $C_\epsilon \leq \overline{C}_\epsilon$ for all $\epsilon > 0$. Therefore, $C_{\rm ss} \leq C_E$. The inequality can be strict. For example, for the Rician channel $C_E = (\alpha^2+\gamma^2)/\sigma^2$ whereas $C_{\rm ss} = \alpha^2/\sigma^2$. Verdú also noted that C_E is lower-bounded by $I_0/2$, and he discussed an interesting connection between the channel capacity per unit energy and the significance of the Fisher information in signal estimation.

Next considered is the first-order asymptotics of the reliability function for channels with small peak constraints under the same regularity assumptions regarding the channel. For $\epsilon > 0$, let $P_\epsilon(N, R)$ be the minimum average probability of error for any block code with peak constraint ϵ, block length N, and rate at least R. The reliability function $E^\epsilon(R)$ is then defined as [5, p. 160]

$$E^\epsilon(R) = \limsup_{N\to\infty} \frac{-\log P_\epsilon(N, R)}{N}.$$

Define

$$E^{\rm ss}(R) = \lim_{\epsilon\to 0} \frac{E^\epsilon(R\epsilon^2)}{\epsilon^2}$$

if the limit exists. The main result is presented in the following theorem.

Theorem II.2: Suppose the regularity assumption RA $(0, \overline{\epsilon_0}, D_o)$ holds for some $\overline{\epsilon_0} > 0$ and some finite D_o. Then $E^{\rm ss}(R)$ is well defined and

$$E^{\rm ss}(R) = \begin{cases} \frac{C_{\rm ss}}{2} - R, & 0 < R \le \frac{C_{\rm ss}}{4} \\ (\sqrt{C_{\rm ss}} - \sqrt{R})^2, & \frac{C_{\rm ss}}{4} \le R \le C_{\rm ss} \\ 0, & R \ge C_{\rm ss}. \end{cases} \tag{2}$$

Remarks: The function $E^{\rm ss}(R)$ has the same shape as the reliability function of the power-constrained infinite-bandwidth additive white Gaussian noise channel [5, p. 381]. Reference [5, Example 3, pp. 147–149, and Exercise 5.31, p. 541] discussed a similar case where instead of the inputs being small the channel transition probabilities were almost independent of the input. This channel, Reiffen's very noisy channel, also has a reliability function with the same shape. Finally, [23], [24] showed that the reliability function for the Poisson channel in the limit of large noise also has the same shape. All these channels can be viewed as infinite-bandwidth channels or as very noisy channels, so perhaps it is not surprising that they all have the same reliability function. A somewhat more technical reason for why they have the same reliability function is that, as shown in the proof, the relevant log-likelihood ratios under the relevant measures asymptotically have the same exponential moments as if they were Gaussian. Intuitively this makes sense because the log-likelihood ratios are sums of a large number of random variables that tend to be small.

The limit $E^{\rm ss}(R)$ is related to the reliability function per unit energy $\tilde{E}^E(R)$ defined by [6], just as $C_{\rm ss}$ is related to the capacity per unit energy C_E. Therefore, $E^{\rm ss}(R) \le \tilde{E}^E(R)$ by the same reasoning we used to deduce that $C_{\rm ss} \le C_E$.

III. Preliminary Implications of the Regularity Assumptions

Let $\Delta q(y|x) = \frac{q(y|x)}{q(y|0)} - 1$. Let $\mathcal{D}$ denote the unit ball centered at the origin in $\mathcal{R}^n$ and let $\epsilon\mathcal{D}$ denote the ball of radius ϵ. Let $\mathcal{P}$ denote the set of all probability measures on $\mathcal{D}$ and let $\mathcal{P}_\epsilon$ denote the set of all probability measures on $\epsilon\mathcal{D}$. Given a measure $\mu_\epsilon \in \mathcal{P}_\epsilon$ let

$$q(y|\mu_\epsilon) = \int q(y|x)\mu_\epsilon(dx).$$

There is a one-to-one correspondence between $\mathcal{P}$ and $\mathcal{P}_\epsilon$. To each $\mu_\epsilon \in \mathcal{P}_\epsilon$ corresponds $\mu \in \mathcal{P}$ given by $\mu(A) = \mu_\epsilon(\epsilon A)$ where A is a Borel subset of $\mathcal{D}$. Equivalently, a random variable Z has distribution μ if and only if ϵZ has distribution μ_ϵ.

Lemma III.1: Given finite positive constants $\overline{\epsilon}_0, D_0, \eta_1 < 1, \eta_2, \eta_3, c_1$, and $n \ge 3$, suppose $RA(0, \overline{\epsilon}_0, D_0)$ holds, and suppose $f: [0, +\infty) \to \Re$ satisfies the following three conditions:

a) f is three times continuously differentiable over $[1-\eta_1, 1+\eta_1]$ with $|f'''(r)| \le 6c_1$, for $r \in [1-\eta_1, 1+\eta_1]$.

b) $|f(r)| \le \eta_2(1+r)^n$, for $r \in [0, +\infty)$.

c) $|f(1)| + |f'(1)| + |f''(1)| \le \eta_3$.

Then, for all $x \in \epsilon\mathcal{D}$

$$E_0\left[f\left(\frac{q(Y|x)}{q(Y|0)}\right)\right] = f(1) + \frac{1}{2}f''(1)x^TKx + o(\epsilon^2) \tag{3}$$

where $o(\epsilon^2)/\epsilon^2 \to 0$ as $\epsilon \to 0$, uniformly over $x \in \epsilon\mathcal{D}$ and over all functions f satisfying the given assumptions. In addition

$$E_0\left[f\left(\frac{q(Y|\mu_\epsilon)}{q(Y|0)}\right)\right] = f(1) + \frac{\epsilon^2 f''(1)E[Z]^TKE[Z]}{2} + o(\epsilon^2) \tag{4}$$

where Z is a random variable in $\mathcal{D}$ such that ϵZ has distribution μ_ϵ, and $o(\epsilon^2)/\epsilon^2 \to 0$ as $\epsilon \to 0$, uniformly over $\mu_\epsilon \in \mathcal{P}_\epsilon$ and over all functions f satisfying the given assumptions.

Proof: The idea of the proof is to apply Taylor's theorem, but first some moments of $\Delta q(Y|x)$ under E_0 are examined. The continuous differentiability of $\log q(y|x)$ in x yields that

$$\frac{q(y|x)}{q(y|0)} = \exp\left(\int_0^1 x^TV(y, tx)\,dt\right). \tag{5}$$

Suppose ϵ is so small that $u(\epsilon) \le 1$. In view of (1), if $||x|| \le \epsilon$ and $0 \le t \le 1$, then

$$\begin{aligned}|x^TV(y, tx)| &\le \epsilon||V(y, tx)|| \\ &\le \epsilon(||V(y, 0)|| + L(y)u(\epsilon)) \le \epsilon B(y, 0).\end{aligned}$$

Inserting this into (5) yields that for $||x|| \le \epsilon$

$$e^{-\epsilon B(y,0)} \le \frac{q(y|x)}{q(y|0)} \le e^{\epsilon B(y,0)}. \tag{6}$$

Using (5) again yields that for $||x|| \le \epsilon$

$$\frac{q(y|x)}{q(y|0)} = \exp(x^TV(y, 0) + \delta(x, y)) \tag{7}$$

where, also using (1),

$$\begin{aligned}|\delta(x, y)| &= \left|\int_0^1 x^T(V(y, tx) - V(y, 0))\,dt\right| \\ &\le \epsilon\int_0^1 L(y)u(t\epsilon)\,dt \le L(y)\epsilon u(\epsilon).\end{aligned}$$

Note that the first moment of $\Delta q(Y|x)$ under E_0 is simply given by

$$E_0[\Delta q(Y|x)] = 0. \tag{8}$$

Next, we investigate higher order moments of $\Delta q(Y|x)$. Expand $\exp(\cdot)$ in a power series, assume that ϵ is small enough that $u(\epsilon) \leq 1$, and use (7) to obtain

$$\Delta q(y|x) = x^T V(y, 0) + \alpha(x, y) \tag{9}$$

where the error term α is bounded by

$$|\alpha(x, y)| \leq L(y)\epsilon u(\epsilon) + \sum_{i=2}^{+\infty} \frac{\epsilon^i (B(y, 0))^i}{i!}.$$

Therefore, given any $n \geq 1$, if ϵ is so small that $u(\epsilon) \leq 1$ and $\epsilon \leq \overline{\epsilon_0}/n$ then

$$|\alpha(x, y)| \leq L(y)\epsilon u(\epsilon) + \epsilon^2 (n/\epsilon_0)^2 e^{(\epsilon_0/n)B(y, 0)}$$

which means that

$$|\alpha(x, y)| \leq \epsilon \max(\epsilon, u(\epsilon))A(y)$$

for some function A on Ω with $E_{x_0}[A^n]$ bounded by a finite constant depending only on $\overline{\epsilon}_0$, D_0, and n. Using (9) and the observation just made with $n = 2$ yields that

$$E_0[\Delta q(y|x)^2] = x^T K x + o(\epsilon^2) \tag{10}$$

where the $o(\epsilon^2)$ is uniform over all $x \in \epsilon\mathcal{D}$. Similarly, if $k > 1$

$$E_0[|\Delta q(Y|x)|^k] = E_0[|x^T V(Y, 0) + \alpha(x, Y)|^k]$$

so that if $\epsilon \leq \frac{\overline{\epsilon}_0}{k}$ and $u(\epsilon) \leq 1$ then

$$\begin{aligned} E_0[|\Delta q(Y, x)|^k] &\leq \epsilon^k E_0[(\|V(Y, 0)\| + A(Y)\max(\epsilon, u(\epsilon)))^n] \\ &\leq \epsilon^k (\text{const}) \end{aligned} \tag{11}$$

for all $x \in \epsilon\mathcal{D}$, where the constant in (11) depends only on $\overline{\epsilon}$, D_o, and k.

Taylor's theorem can now be applied to prove (3). Write

$$f(r) = f(1) + f'(1)(r-1) + \tfrac{1}{2} f''(1)(r-1)^2 + e(r).$$

Replacing r by $1 + \Delta q(Y|x)$ and applying (8) and (10) yields (3), except it remains to show that the contribution of the error term $e(r)$ is covered by the $o(\epsilon^2)$ term in (3). If $r \in [1-\eta_1, 1+\eta_1]$, then Taylor's theorem and the Mean Value theorem imply that $e(r) = \frac{1}{6} f'''(\xi)(r-1)^3$ for some ξ in the closed interval with endpoints 1 and r, Therefore, $|e(r)| \leq c_1|r-1|^3$ if $r \in [1-\eta_1, 1+\eta_1]$. If $r \in [0, 1-\eta_1) \cup (1+\eta_1, +\infty)$ then

$$\begin{aligned} |e(r)| &= |f(r) - f(1) - f'(1)(r-1) - \frac{1}{2} f''(1)(r-1)^2| \\ &\leq \eta_2 (1+r)^n + \eta_3 \max(1, |r-1|, |r-1|^2) \\ &\leq c_2 |1-r|^n \end{aligned}$$

for $c_2 = \eta_2(\frac{2}{\eta_1} - 1)^n + \eta_3(\frac{1}{\eta_1})^n$. Thus,

$$|e(r)| \leq c_1|r-1|^3 + c_2|r-1|^n$$

for all $r \in [0, +\infty)$. Therefore, the bound (11) applied for $k = 3$ and $k = n$ yields that

$$E_0[|e(\frac{q(y|x)}{q(y|0)})|] \leq (\text{const})\,\epsilon^3$$

whenever $x \in \epsilon\mathcal{D}$, if ϵ is so small that $u(\epsilon) \leq 1$ and $\epsilon \leq \frac{\overline{\epsilon}_0}{n}$. This completes the proof of (3).

The proof of (4) is similar. To begin, integrate each side of (9) against μ_ϵ to yield the similar equation

$$\frac{q(y|\mu_\epsilon)}{q(y|x_0)} - 1 = \epsilon E[Z]^T V(y, 0) + \beta(\mu_\epsilon, y) \tag{12}$$

where ϵZ has distribution μ_ϵ, and

$$\beta(\mu_\epsilon, y) = \int \alpha(x, y)\mu_\epsilon(dx).$$

Therefore, it follows that $|\beta(\mu_\epsilon, y)| \leq \epsilon \max(\epsilon, u(\epsilon))A(y)$. Using (12) we can establish (8), (10), and (11) with x replaced by μ_ϵ, and apply Taylor's theorem to obtain (4). The proof of Lemma III.1 is complete. □

Lemma III.1 and its proof can be generalized to functions of several variables. The following is a version that applies to one function, rather than a family of functions, since that is the only generalization that is needed in this paper. The proof is a straightforward modification of the proof of Lemma III.2 and is omitted.

Lemma III.2: Given finite positive constants $\overline{\epsilon}_0$, D_0, η, and $n \geq 3$. Suppose $RA(0, \overline{\epsilon}_0, D_0)$ holds, suppose $J \geq 1$ and suppose F: $[0, +\infty)^K \to \Re$, such that all derivatives of F up to order three exist and are continuous in a neighborhood of $(1, \ldots, 1)^T$, and

$$|F(r_1, \ldots, r_J)| \leq \eta \sum_{j=1}^{J} (1 + r_j)^n.$$

Then

$$\begin{aligned} &E_0\left[F\left(\frac{q(Y|x_1)}{q(Y|0)}, \ldots, \frac{q(Y|x_J)}{q(Y|0)}\right)\right] \\ &\qquad = \sum_i \sum_j \left[\frac{\partial^2 F}{\partial r_i \partial r_j}(1, \ldots, 1)\right] x_i^T K x_j + o(\epsilon^2) \end{aligned}$$

where $o(\epsilon^2)/\epsilon^2 \to 0$ as $\epsilon \to 0$, uniformly over $x_1, x_2, \ldots, x_J \in \epsilon\mathcal{D}$.

IV. Proof of Theorem II.1

Given $\mu_\epsilon \in \mathcal{P}_\epsilon$, let X be a random variable with probability measure μ_ϵ and let Y denote the corresponding channel output. Then Y has the probability density $q(y|\mu_\epsilon)$. By well-known results in information theory [5], $C_\epsilon = \sup_{\mu_\epsilon \in \mathcal{P}_\epsilon} I(X;\ Y)$. The first step of the proof is to establish that $I(X;\ Y) = \tilde{h}(Y) - \tilde{h}(Y|X)$ where $h(\tilde{Y})$ is the relative entropy of Y and $\tilde{h}(Y|X)$ is the conditional relative entropy of Y given X, both relative to $q(y|0)$

$$\tilde{h}(Y) = -E_0\left[\frac{q(Y|\mu_\epsilon)}{q(Y|0)} \log \frac{q(Y|\mu_\epsilon)}{q(Y|0)}\right] \tag{13}$$

$$\tilde{h}(Y|X) = -\int E_0\left[\frac{q(Y|\epsilon x)}{q(Y|0)} \log \frac{q(Y|\epsilon x)}{q(Y|0)}\right] \mu(dx). \tag{14}$$

To establish that $I(X;\,Y) = \tilde{h}(Y) - \tilde{h}(Y|X)$ we need only show that $\tilde{h}(Y) > -\infty$.

Application of Lemma III.1 with $f(r) = r\log(r)$ shows that not only are $\tilde{h}(Y)$ and $\tilde{h}(Y|X)$ finite for sufficiently small ϵ, but also

$$\tilde{h}(Y) = -\frac{\epsilon^2}{2}\,E[Z]^T K E[Z] + o(\epsilon^2)$$
$$\tilde{h}(Y|X) = -\frac{\epsilon^2}{2}\,E[Z^T K Z] + o(\epsilon^2)$$

where Z has distribution μ, and each $o(\epsilon^2)$ term is uniform in μ. Using these approximations of $\tilde{h}(Y)$ and $\tilde{h}(Y|X)$ yields

$$I(X;\,Y) = \frac{\epsilon^2}{2}\,E[(Z - E[Z])^T K (Z - E[Z])] + o(\epsilon^2)$$

where again the $o(\epsilon^2)$ term is uniform in μ. Thus, taking limits yields that

$$\lim_{\epsilon\to 0}\frac{C_\epsilon}{\epsilon^2} = \frac{1}{2}\sup_{\mu\in\mathcal{P}} E[(Z - E[Z])^T K (Z - E[Z])]. \tag{15}$$

Since K is positive semidefinite, it can be diagonalized by a unitary transformation. For a diagonal K, and therefore in general, it is clear that the right-hand side of (15) cannot exceed half the largest eigenvalue. To attain equality, it is necessary and sufficient that $E[Z] = 0$, $P[\|Z\| = 1] = 1$, and Z be distributed within the eigenspace of the largest eigenvalue. For example, if $\tilde{x}$ is a unit eigenvector corresponding to the largest eigenvalue I_0 of K, then choosing Z to be $+\tilde{x}$ and $-\tilde{x}$ equiprobably achieves the supremum in (15), and the theorem follows.

V. Random-Coding Upper Bound on Error Probability

The following lemma is established in this section.

Lemma V.1 (Random-Coding Bound): For $R > 0$

$$\liminf_{\epsilon\to 0}\frac{E^\epsilon[R\epsilon^2]}{\epsilon^2} \geq b(R)$$

where $b(R)$ denotes the right-hand side of (2).

Proof: Let P be a distribution on $\mathcal{D}$ with finite support $\{x_1, x_2, \ldots, x_J\}$ and respective probability masses $p_1, \ldots, p_J$, let $0 \leq \rho \leq 1$, and let $R > 0$. The well-known random coding bound for finite input channels [5, pp. 138–145] yields that

$$E^\epsilon(R) \geq -\rho R + E_0(\rho,\,P) \tag{16}$$

where

$$E_0(\rho,\,P) = -\log E_0\left[\left\{\sum_{j=1}^{J} p_k\left(\frac{q(Y|\epsilon x_k)}{q(Y|0)}\right)^{\frac{1}{1+\rho}}\right\}^{1+\rho}\right].$$

The function $E_0[\rho,\,P]$ can be expressed as

$$E_0(\rho,\,P) = -\log E_0\left[F\left(\frac{q(Y|\epsilon x_1)}{q(Y|0)}, \ldots, \frac{q(Y|\epsilon x_J)}{q(Y|0)}\right)\right]$$

where F is defined on $[0,\,\infty)^J$ by

$$F(r_1, \ldots, r_J) = \left\{\sum_{j=1}^{J} p_j r_j^{\frac{1}{1+\rho}}\right\}^{1+\rho}.$$

Calculation yields that $F(1, \ldots, 1) = 1$ and

$$\frac{\partial^2 F}{\partial r_i \partial r_j}(1, \ldots, 1) = \frac{\rho}{1+\rho}\left\{p_i p_j - I_{\{i=j\}} p_i\right\}.$$

Lemma III.2 and the fact $-\log(1 - s) = s + o(s)$ thus yields that

$$E_0(\rho,\,P) = \epsilon^2\frac{\rho E[(Z - E[Z])^T K (Z - E[Z])]}{2(1+\rho)} + o(\epsilon^2) \tag{17}$$

where Z has distribution P. Select $J = 2, p_1 = p_2 = 0.5$, let x_1 be the eigenvector corresponding to the maximum eigenvalue of K (which is I_0), and let $x_2 = -x_1$. Then

$$E[(Z - E[Z])^T K (Z - E[Z])] = I_0.$$

For this choice of P, combining (16) and (17) yields that

$$\frac{E^\epsilon(R\epsilon^2)}{\epsilon^2} \geq -\rho R + \frac{\rho I_0}{2(1+\rho)} + o(1). \tag{18}$$

Therefore,

$$\liminf_{\epsilon\to 0}\frac{E^\epsilon[R\epsilon^2]}{\epsilon^2} \geq -\rho R + \frac{\rho I_0}{2(1+\rho)}. \tag{19}$$

Taking $\rho \in [0,\,1]$ to maximize the right-hand side of (19) and using $I_0/2 = C_{\rm ss}$ completes the proof of Lemma V.1. □

VI. Lower Bounds on Error Probability

The proof of Theorem II.2 is completed in this section by providing a complement to Lemma V.1. First, a subsection with some further implications of the regularity assumption is given. Next a sphere-packing lower bound on error probability is given which matches the random coding bound for rates greater than or equal to $\frac{C_{\rm ss}}{4}$, and a low-rate lower bound on error probability is given which matches the random coding bound at $R = 0+$. The sphere-packing bound and the low-rate bound then combine by well-known arguments [2], [4], [5], [20], [21] to provide a straight-line bound. The straight-line bound for rates below $\frac{C_{\rm ss}}{4}$ and the sphere-packing bound for rates above $\frac{C_{\rm ss}}{4}$ exactly match the random coding bound.

A. Further Implications of the Regularity Assumptions

The regularity assumption at a point implies that the regularity assumption (with a change of constant) holds uniformly in a neighborhood of the point, as shown in the next lemma.

Lemma VI.1 (Local Uniformity of the Regularity Assumption): Suppose RA $(0,\,\overline{\epsilon}_0,\,D_0)$ holds. Then, for some $r > 0$ and $\epsilon_0 > 0$, RA $(x_0,\,\epsilon_0,\,D_0)$ holds for all $\|x_0\| \leq r$.

Proof: Concentrate on the second part of the regularity assumption, since the first part does not depend on x_0. By (1)

$$\begin{aligned} B(y,\,x_0) &= L(y) + \|V(y,\,x_0)\| \\ &\leq L(y) + \|V(y,\,0)\| + L(y)u(r) \\ &\leq (1 + u(r))B(y,\,0). \end{aligned}$$

By (6), if r is so small that $u(r) \leq 1$ then

$$\frac{q(y|x_0)}{q(y|0)} \leq \exp(rB(y,\,0)).$$

Select r so small that $u(r) \le 1$ and $r < \bar{\epsilon}_0$, and select $\epsilon_0 > 0$ so that $r + \epsilon_0(1 + u(r)) = \bar{\epsilon}_0$. Then

$$\begin{aligned} E_{x_0}\left[e^{\epsilon_0 B(Y, x_0)}\right] &= E_0\left[\frac{q(y|x_0)}{q(y|0)} e^{\epsilon_0 B(Y, x_0)}\right] \\ &\le E_0\left[e^{[r+\epsilon_0(1+u(r))]B(Y,0)}\right] \\ &= E_0\left[e^{\bar{\epsilon}_0 B(Y,0)}\right] \le D_0. \qquad \square \end{aligned}$$

Let $x, \tilde{x} \in \epsilon\mathcal{D}$ and $\lambda \in \Re$. Define the density w^λ on Ω by

$$w^\lambda_{x,\tilde{x}}(y) = \frac{q(y|x)^\lambda q(y|\tilde{x})^{1-\lambda}}{\int q(\overline{y}|x)^\lambda q(\overline{y}|\tilde{x})^{1-\lambda} m(d\overline{y})}$$

where m denotes the reference measure on Ω, and let $E^\lambda_{x,\tilde{x}}$ denote expectation for the probability measure $P^\lambda_{x,\tilde{x}}$ on Ω with density $w^\lambda_{x,\tilde{x}}$. Let $\Lambda_{1,x,\tilde{x}} = \log\left(\frac{q(Y|x)}{q(Y|\tilde{x})}\right)$.

Lemma VI.2: Let $RA(0, \bar{\epsilon}_0, D_0)$ hold. Then

$$\begin{aligned} &-\log\left(\int q(\overline{y}|x)^\lambda q(\overline{y}|\tilde{x})^{1-\lambda} m(d\overline{y})\right) \\ &\qquad = \frac{\lambda(1-\lambda)}{2}(x-\tilde{x})^T K(x-\tilde{x}) + o(\epsilon^2) \\ &E^\lambda_{x,\tilde{x}}[\Lambda_{1,x,\tilde{x}}] = \left(\lambda - \frac{1}{2}\right)(x-\tilde{x})^T K(x-\tilde{x}) + o(\epsilon^2) \\ &E^\lambda_{x,\tilde{x}}[(\Lambda_{1,x,\tilde{x}})^2] = (x-\tilde{x})^T K(x-\tilde{x}) + o(\epsilon^2) \end{aligned}$$

where in each equation the $o(\epsilon^2)$ term is uniform over $x, \tilde{x} \in \epsilon\mathcal{D}$ and over λ in bounded subsets of the real line.

Proof: It suffices to prove the lemma for λ restricted to $(-\infty, 0.5]$ and for λ restricted to $[0.5, +\infty)$. By symmetry, it suffices to consider only one of these cases, so we establish the lemma under the added assumption that $\lambda \ge 0.5$. By Lemma III.1 and the continuity of $K_x = E_x[V(Y, x)V(Y, x)^T]$ at $x = 0$ it suffices to prove the theorem for $\tilde{x} = 0$. In summary, it must be shown that

$$-\log E_0\left[\left(\frac{q(Y|x)}{q(Y|0)}\right)^\lambda\right] = \frac{\lambda(1-\lambda)}{2} x^T K x + o(\epsilon^2) \quad (20)$$

$$E^\lambda_{x,0}[\Lambda_{1,x,0}] = \left(\lambda - \frac{1}{2}\right) x^T K x + o(\epsilon^2) \quad (21)$$

$$E^\lambda_{x,0}[(\Lambda_{1,x,0})^2] = x^T K x + o(\epsilon^2) \quad (22)$$

where the $o(\epsilon^2)$ terms are uniform in $x \in \epsilon\mathcal{D}$ and in λ over bounded subsets of $[0.5, +\infty)$.

Take $\phi_0(s) = s^\lambda$. Note that for λ in a bounded subset of $[0.5, \infty)$ the constants in Lemma III.1 can be selected so that ϕ_0 satisfies the hypothesis of the lemma for all λ in the bounded set. Thus,

$$E_0\left[\left(\frac{q(Y|x)}{q(Y|0)}\right)^\lambda\right] = 1 + \frac{\lambda(\lambda-1)}{2} x^T K x + o(\epsilon^2)$$

where the $o(\epsilon^2)$ term is uniform in $x \in \epsilon\mathcal{D}$ and λ in the bounded subset of $[0.5, +\infty)$. Since $-\log(1+s) = -s + O(s^2)$, (20) is proved.

The left-hand side of (21) is

$$\frac{E_0\left[\phi_1\left(\frac{q(Y|x)}{q(Y|0)}\right)\right]}{E_0\left[\phi_2\left(\frac{q(Y|x)}{q(Y|0)}\right)\right]}$$

for $\phi_1(r) = r^\lambda \log(r)$ and $\phi_2(r) = r^\lambda$. The left-hand side of (22) is the same, but with $\phi_3(r) = r^\lambda(\log(r))^2$ in place of $\phi_1(r)$. Equations (21) and (22) thus follow by applying Lemma III.1. $\square$

Based upon the result of Lemma VI.2, define the distance $d(x, \tilde{x})$ between x and $\tilde{x}$ by $d(x, \tilde{x}) = (x-\tilde{x})^T K(x-\tilde{x})$. Let $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \epsilon\mathcal{D}^N$ and $\lambda > 0$. Let $E^{N,\lambda}_{\boldsymbol{x},\tilde{\boldsymbol{x}}}$ denote expectation for $(Y_1, Y_2, \ldots, Y_N)$ where $Y_1, \ldots, Y_N$ are mutually independent, and Y_n has density $w^\lambda_{x_n, \tilde{x}_n}(y_n)$. Denote the corresponding measure on Ω^N by $P^{N,\lambda}_{\boldsymbol{x},\tilde{\boldsymbol{x}}}$, and let

$$\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} = \Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}(Y) = \sum_{n=1}^N \log\left(\frac{P(Y_n|x_n)}{P(Y_n|\tilde{x}_n)}\right).$$

Also define

$$d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \sum_{n=1}^N d(x_n, \tilde{x}_n).$$

Similarly, let $E^N_{\boldsymbol{x}}$ denote expectation for independent $Y_1, \ldots, Y_N$ with Y_n having density $q(y_n|x_n)$, and let $P^N_{\boldsymbol{x}}$ denote the corresponding probability measure on Ω^N. The following lemma holds.

Lemma VI.3: Let $a = N\epsilon^2$. The following hold:

$$E_{\tilde{\boldsymbol{x}}}[e^{\lambda\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}}] = \exp\left(-\frac{\lambda(1-\lambda)}{2} d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) + o(a, \epsilon)\right)$$

$$E^{N,\lambda}_{\boldsymbol{x},\tilde{\boldsymbol{x}}}[\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}] = \left(\lambda - \frac{1}{2}\right) d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) + o(a, \epsilon) \quad (23)$$

$$\mathrm{Var}^{N,\lambda}_{\boldsymbol{x},\tilde{\boldsymbol{x}}}[\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}] = d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) + o(a, \epsilon) \quad (24)$$

$$D(P^N_{\boldsymbol{x}} \| P^N_{\tilde{\boldsymbol{x}}}) = \frac{1}{2} d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) + o(a, \epsilon) \quad (25)$$

where in each equation

$$\lim_{\epsilon\to 0} \lim_{a\to\infty} \frac{|o(a, \epsilon)|}{a} = 0$$

uniformly over $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \epsilon\mathcal{D}^N$ and over λ in bounded subsets of $\Re$.

Proof: The first three equations are immediate from Lemma VI.2. The fourth is the same as the second with $\lambda = 1$. $\square$

Remark: Take $\lambda = 0$ in (23) and (24) to see that under $P_{\tilde{x}}$, $\Lambda_{N,x,\tilde{x}}$ has, up to small-order terms, mean $-\frac{1}{2} d_N(x, \tilde{x})$ and variance $d_N(x, \tilde{x})$. The first equation of Lemma VI.3 thus shows that $\Lambda_{N,x,\tilde{x}}$ asymptotically has the same exponential moments under $P_{\tilde{x}}$ as if it had a Gaussian distribution.

B. Sphere-Packing Lower Bound on Error Probability

The following lemma is established in this subsection.

Lemma VI.4 (Sphere-Packing Bound): For $0 < R < C_{\mathrm{ss}}$

$$\limsup_{\epsilon\to 0} \frac{E^\epsilon[R\epsilon^2]}{\epsilon^2} \le \left(\sqrt{C_{\mathrm{ss}}} - \sqrt{R}\right)^2.$$

Proof: Let $E > 0$ satisfy

$$E < \limsup_{\epsilon \to 0} \frac{E^{\epsilon}[R\epsilon^2]}{\epsilon^2}.$$

Then it suffices to show that $E \leq (\sqrt{C_{ss}} - \sqrt{R})^2$. By the choice of E there exists $\epsilon > 0$ arbitrarily small such that for some sequence with $a_m \to +\infty$ as $m \to +\infty$, there exist $(N_m = \frac{a_m}{\epsilon^2}, M_m = e^{a_m R})$ codes with the maximum probability of error $p_{\max} \leq e^{-a_m E}$. Here we can use the maximum error probability because a code with $2M$ codewords and average error probability $\overline{p}$ can be thinned to a produce a code with M codewords and a maximum error probability $2\overline{p}$. For brevity, the subscripts m will henceforth be dropped from a, N, and M.

The decoding sets $(F_1, F_2, \ldots, F_M)$ partition the output space Ω^N. Let $\mathbf{0}$ denote the N vector of all zeros. The idea of the sphere-packing bound (and the reason for the name) is that the decoding sets F_i cannot be too small, or else the probability of error will be too large. This limits the number of the sets, and hence the rate of the code. In the setting of this paper, there is a natural measure of the size of these sets, namely, the probability measure $P_{\mathbf{0}}^N$.

Since $P_{\mathbf{0}}^N(\Omega^N) = 1$, there exists a codeword index i such that α defined by $\alpha = P_0^N(F_i)$ satisfies $\alpha \leq \frac{1}{M}$. On the other hand, write $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ for the ith codeword, define

$$\rho = \frac{d_N(\boldsymbol{x}, \mathbf{0})}{a} = \frac{1}{a} \sum_{n=1}^{N} x_n^T K x_n$$

and define β by $\beta = P_{\boldsymbol{x}}^N(F_i^c)$. Then $\beta \leq e^{-aE}$. It is useful to view the numbers α and β as the Type I and Type II error probabilities for the test of hypotheses H_0: $(Y_1, \ldots, Y_N)$ has measure $P_{\mathbf{0}}^N$ versus H_1: $(Y_1, \ldots, Y_N)$ has measure $P_{\boldsymbol{x}}^N$. Since refinement cannot decrease divergence distance

$$\begin{aligned} D(P_{\boldsymbol{x}}^N \| P_{\mathbf{0}}^N) &\geq \beta \log\left(\frac{\beta}{1-\alpha}\right) + (1-\beta)\log\left(\frac{1-\beta}{\alpha}\right) \\ &\geq (1-\beta)\log(M) - h(\beta) \\ &\geq (1-\beta)aR - \log(2). \end{aligned}$$

Thus, $D(P_{\boldsymbol{x}}^N \| P_{\mathbf{0}}^N) \geq aR + o(a)$. Combining this with (25) of Lemma VI.3 (with $\tilde{\boldsymbol{x}} = \mathbf{0}$) shows that

$$\rho \geq 2R + o(a, \epsilon)/a. \tag{26}$$

This inequality is used later when Chebychev's inequality is applied to get a large deviations lower bound.

By the Neyman–Pearson lemma, there is a threshold τ so that

$$P_{\boldsymbol{x}}\{\Lambda_{N, \boldsymbol{x}, \mathbf{0}} < a\tau\} \leq P_{\boldsymbol{x}}(F_i^c) \leq e^{-aE} \tag{27}$$

$$P_{\mathbf{0}}\{\Lambda_{N, \boldsymbol{x}, \mathbf{0}} > a\tau\} \leq P_{\mathbf{0}}(F_i) \leq \frac{1}{M} = e^{-aR}. \tag{28}$$

Next, large deviations type lower bounds are applied to (27) and (28). The standard method of changing the measure and applying the law of large numbers under the new measure is applied. This is done uniformly in $\boldsymbol{x}$ (subject to (26)).

Let us examine (27) first. By (23) and (24) of Lemma VI.3 with $\lambda = 1$ and $\tilde{\boldsymbol{x}} = \mathbf{0}$, (26), and Chebychev's inequality, it follows that the threshold τ must satisfy $\tau \leq \frac{\rho}{2} + o(a, \epsilon)$. Equation (27) will be used to show that $\tau \leq \frac{\rho}{2} - \sqrt{2\rho E} + o(a, \epsilon)$. The argument is by contradiction. If false, then for any $\overline{\delta} > 0$ the threshold τ can be taken to lie in the bounded interval

$$\tau \in \left[\frac{\rho}{2} - \sqrt{2\rho E} + \overline{\delta},\, \frac{\rho}{2} + \overline{\delta}\right] \tag{29}$$

for some subsequence of $\epsilon \to 0$ and arbitrarily large a. By (27), for any $\delta > 0$ and $\theta \geq 0$

$$\begin{aligned} e^{-aE} &\geq P_{\boldsymbol{x}}^N\{a(\tau - 2\delta) \leq \Lambda_{N, \boldsymbol{x}, \mathbf{0}} < a\tau\} \\ &\geq E_{\boldsymbol{x}}^N\left[e^{\theta[a(\tau-2\delta) - \Lambda_{N, \boldsymbol{x}, \mathbf{0}}]} I_{\{a(\tau-2\delta) \leq \Lambda_{N, \boldsymbol{x}, \mathbf{0}} < a\tau\}}\right] \\ &\geq P_{\boldsymbol{x}, \mathbf{0}}^{N, 1-\theta}\{a(\tau - 2\delta) \leq \Lambda_{N, \boldsymbol{x}, \mathbf{0}} < a\tau\} \\ &\quad \times e^{\theta a(\tau - 2\delta)} E_{\mathbf{0}}^N\left[e^{(1-\theta)\Lambda_{N, \boldsymbol{x}, \mathbf{0}}}\right] \end{aligned} \tag{30}$$

where $P_{\boldsymbol{x}, \mathbf{0}}^{N, 1-\theta}$ is defined as in Lemma VI.3 (with $\lambda = 1 - \theta$ and $\tilde{\boldsymbol{x}} = \mathbf{0}$).

The next step is to specify θ so that the $P_{\boldsymbol{x}}^{N, 1-\theta}$ term in (30) is at least $\frac{1}{2}$. To that end, let θ be defined by

$$\left(\frac{1}{2} - \theta\right)\rho = \tau - \delta. \tag{31}$$

It must be checked that $\theta \geq 0$ as required, and for application of Lemma VI.3 it must also be checked that θ is bounded for all sufficiently small ϵ and large enough a. But since $\tau \leq \frac{\rho}{2} + o(a, \epsilon)$, it follows that $\theta \geq \delta + o(a, \epsilon)$, so that $\theta > 0$ for ϵ sufficiently small and a sufficiently large (depending on δ). Also, θ is bounded from above since τ is bounded from above and ρ is bounded from below. Thus, Lemma VI.3 can be applied with $\lambda = 1 - \theta$ and $\tilde{\boldsymbol{x}} = \mathbf{0}$. Chebychev's inequality, (26), (23) and (24) yield that the $P_{\boldsymbol{x}, \mathbf{0}}^{N, 1-\theta}$ term in (30) is at least $\frac{1}{2}$. The $E_{\mathbf{0}}^N$ term in (30) is, by the first equation of Lemma VI.2 with $\lambda = 1 - \theta$ and $\tilde{\boldsymbol{x}} = \mathbf{0}$

$$\exp\left(-\frac{(1-\theta)\theta}{2}\rho a + o(a, \epsilon)\right).$$

Thus,

$$e^{-aE} \geq \frac{1}{2} e^{\theta a(\tau - 2\delta)} e^{-\frac{(1-\theta)\theta}{2}\rho a + o(a, \epsilon)}$$

where θ is given by (31). This gives

$$E \leq \frac{\rho}{2}\left[\frac{\tau}{\rho} - \frac{1}{2}\right]^2 + O(\delta) + o(a, \epsilon).$$

Recall that $\tau \leq \frac{\rho}{2} + o(a, \epsilon)$, so $\frac{\tau}{\rho} - \frac{1}{2} \leq o(a, \epsilon)$. Thus, for ϵ sufficiently small and a large (depending on δ)

$$\sqrt{E + O(\delta)} \leq \sqrt{\frac{\rho}{2}}\left[\frac{1}{2} - \frac{\tau}{\rho}\right]$$

so that

$$\tau \leq \frac{\rho}{2} - \sqrt{2(E + O(\delta))\rho}$$

which is a contradiction to (29) for δ sufficiently small. Thus, the inequality

$$\tau \leq \frac{\rho}{2} - \sqrt{2\rho E} + o(a, \epsilon)$$

is established. Since (28) remains true if τ is increased, we can and do assume

$$\tau = \max\left\{\frac{\rho}{2} - \sqrt{2\rho E} + o(a, \epsilon),\, -\frac{\rho}{2}\right\}. \tag{32}$$

By (28), for any $\delta > 0$ and $\lambda > 0$

$$\begin{aligned} e^{-aR} &\geq P_{\mathbf{0}}^N\{a(\tau + 2\delta) \geq \Lambda_{N,\boldsymbol{x},\mathbf{0}} > a\tau\} \\ &\geq E_{\mathbf{0}}^N \left[e^{\lambda\{\Lambda_{n,\boldsymbol{x},\mathbf{0}} - a(\tau+2\delta)\}} I_{\{a(\tau+2\delta) \geq \Lambda_{N,\boldsymbol{x},\mathbf{0}} > a\tau\}} \right] \\ &\geq P_{\boldsymbol{x},\mathbf{0}}^{N,\lambda}\{a(\tau + 2\delta) \geq \Lambda_{N,\boldsymbol{x},\mathbf{0}} > a\tau\} \\ &\quad \times e^{-\lambda a(\tau+2\delta)} E_{\mathbf{0}} \left[e^{\lambda \Lambda_{N,\boldsymbol{x},\mathbf{0}}} \right]. \end{aligned} \tag{33}$$

By Lemma VI.3, (26), and Chebychev's inequality, if we select λ so that

$$\tau + \delta = \left(\lambda - \tfrac{1}{2}\right) \rho \tag{34}$$

then the $P_{\boldsymbol{x}}^{N,\lambda}$ term in (33) is at least $\frac{1}{2}$ for ϵ small enough and all sufficiently large a. Applying this and Lemma VI.3 yields

$$e^{-aR} \geq \tfrac{1}{2}\, e^{-\lambda a(\tau+2\delta)} e^{\frac{\lambda(\lambda-1)}{2}\rho a + o(a,\epsilon)}$$

where λ is given by (34). Thus, for ϵ sufficiently small and all a sufficiently large

$$R \leq \frac{\rho}{2}\left(\frac{\tau}{\rho} + \frac{1}{2}\right)^2 + O(\delta).$$

Using (32) to substitute for τ yields

$$R \leq \frac{\rho}{2}\left(1 - \sqrt{\frac{2E}{\rho}}\right)_+^2 + O(\delta).$$

Since $R > 0$ it must be that $\sqrt{\frac{2E}{\rho}} < 1$ for δ sufficiently small. Since $\delta > 0$ is arbitrary and $\rho \leq I_0$

$$\sqrt{R} + \sqrt{E} \leq \sqrt{\frac{I_0}{2}} = \sqrt{C_{\text{ss}}}$$

which establishes the sphere-packing bound. □

C. Lower Bound on Error Probability for Low Rate Codes

The Bhattacharyya single letter distance $d_B(x, \tilde{x})$ for x, $\tilde{x} \in \Re^d$ is defined by

$$d_B(x, \tilde{x}) = -\log \int \sqrt{q(y|x)q(y|\tilde{x})}\, m\,(dy).$$

Using Lemma VI.2 with $\lambda = \frac{1}{2}$, for $x, \tilde{x} \in \epsilon\mathcal{D}$ we get

$$d_B(x, \tilde{x}) = \frac{1}{8} d(x, \tilde{x}) + o(\epsilon^2).$$

It follows that for $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \epsilon\mathcal{D}^N$, the Bhattacharyya distance between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$, is given by $\frac{1}{8}\, d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) + o(a, \epsilon)$. Methods similar to those used to prove the sphere-packing bound in the previous subsection are applied to prove the following lemma regarding the maximum $p_{\max}(\boldsymbol{x}, \tilde{\boldsymbol{x}})$, of the Type I and Type II errors for the hypothesis testing problem $P_{\boldsymbol{x}}^N$ versus $P_{\tilde{\boldsymbol{x}}}^N$.

Lemma VI.5: Let $a = \frac{N}{\epsilon^2}$. Then

$$p_{\max}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \geq e^{-\frac{1}{8} d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) - o(a,\epsilon)}$$

where

$$\lim_{\epsilon \to 0} \lim_{a \to \infty} \frac{|o(a, \epsilon)|}{a} = 0$$

uniformly over $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \epsilon\mathcal{D}^N$.

Proof: If no subsequence of $p_{\max}(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ converges to 0 at a rate exponential in a there is nothing to prove. Passing to a subsequence if necessary, it can be assumed that $p_{\max}(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ does indeed tend to zero exponentially in a. It follows from the reasoning used to derive (26) that $d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ grows linearly in a.

By the Neyman–Pearson lemma

$$p_{\max}(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \geq \min\{P_{\tilde{\boldsymbol{x}}}^N\{\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} > 0\},\, P_{\boldsymbol{x}}^N\{\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} < 0\}\}. \tag{35}$$

Both terms on the right-hand side of (35) can be bounded below as follows. For any $\delta > 0$ and $\lambda > 0$

$$\begin{aligned} P_{\tilde{\boldsymbol{x}}}^N\{\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} > 0\} &\geq E_{\tilde{\boldsymbol{x}}}^N \left[I_{\{2\delta a \geq \Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} > 0\}} e^{\lambda(\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} - 2\delta a)} \right] \\ &= e^{-2\lambda\delta a} P_{\boldsymbol{x},\tilde{\boldsymbol{x}}}^{N,\lambda}[2\delta a \geq \Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}} > 0] \\ &\quad \times E_{\boldsymbol{x},\tilde{\boldsymbol{x}}}^{N,\lambda} \left[e^{\lambda\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}} \right] \\ &\geq e^{-2\lambda\delta a} \tfrac{1}{2}\, E_{\boldsymbol{x},\tilde{\boldsymbol{x}}}^{N,\lambda} \left[e^{\lambda\Lambda_{N,\boldsymbol{x},\tilde{\boldsymbol{x}}}} \right] \\ &= \frac{1}{2}\, e^{-2\lambda\delta a} e^{\frac{-\lambda(1-\lambda)}{2} d_N(\boldsymbol{x},\tilde{\boldsymbol{x}}) + o(a,\epsilon)} \end{aligned} \tag{36}$$

where by Chebychev's inequality and (23) and (24) of Lemma VI.3, the second inequality holds for ϵ small enough, a large enough, and λ selected such that

$$\left(\lambda - \frac{1}{2}\right) d_N(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = \delta \boldsymbol{a}. \tag{37}$$

By symmetry, the second term in (35) is also bounded by the right-hand side of (36). Equation (37) implies that $\lambda = \frac{1}{2} + O(\delta)$. Substituting this into the lower bound of (36) yields the lemma. □

The main result of this section is given by the following lemma.

Lemma VI.6 (Low-Rate Bound): For any $R > 0$

$$\limsup_{\epsilon \to 0} \frac{E^\epsilon(R\epsilon^2)}{\epsilon^2} \leq \frac{C_{\text{ss}}}{2}.$$

Proof: Select any $E > 0$ such that

$$E < \limsup_{\epsilon \to 0} \frac{E^\epsilon(R\epsilon^2)}{\epsilon^2}.$$

It suffices to show that $E \leq \frac{C_{\text{ss}}}{2}$. There exist $\epsilon > 0$ arbitrarily small such that for some sequence $a \to \infty$, there exist codes with block length $N = \frac{a}{\epsilon^2}$ and with e^{aR} codewords with $p_{\max} \leq e^{-aE}$. Note that for any codeword $\boldsymbol{x}$

$$d_N(\boldsymbol{x}, \mathbf{0}) = \sum_{n=1}^N x_n^T K x_n \leq I_0 a$$

because $0 \leq x_n^T K x_n \leq \epsilon^2 I_0$ for each n.

Fix an integer M. Then for large enough a there are at least M words in the code, denoted by $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^M\}$. Let $\boldsymbol{x}^*$ denote the sum of these M codewords. The minimum pairwise distance d^* for these M codewords satisfies (Plotkin bound)

$$\begin{aligned} d^* &\leq \frac{1}{M(M-1)} \sum_{m=1}^M \sum_{l=1}^M d_N(\boldsymbol{x}^{\boldsymbol{m}}, \boldsymbol{x}^{\boldsymbol{l}}) \\ &= \frac{1}{M(M-1)} \sum_{m=1}^M \sum_{l=1}^M \sum_{n=1}^N (x_n^m - x_n^l)^T K (x_n^m - x_n^l) \\ &= \frac{2}{M(M-1)} \left[M \sum_{m=1}^M d_N(\boldsymbol{x}^{\boldsymbol{m}}, 0) - d_N(\boldsymbol{x}^*, 0) \right] \\ &\leq \frac{2}{M-1} \sum_{m=1}^M d_N(\boldsymbol{x}^{\boldsymbol{m}}, 0) \leq \frac{2MI_0 a}{(M-1)}. \end{aligned}$$

Now $\frac{M}{M-1}$ can be made arbitrarily close to 1 by selecting M large. Thus, by Lemma (VI.5)

$$p_{\max} \geq \exp\left(-a\frac{I_0}{4} + o(\epsilon, a)\right). \tag{38}$$

Thus, $E \leq \frac{I_0}{4} = \frac{C_{ss}}{2}$ as required. □

D. Straight-Line Bound and Completion of Proof of Theorem II.2

The straight line traced by $\frac{C_{ss}}{2} - R$ as R ranges over the interval $[0, \frac{C_{ss}}{4}]$ starts from the low rate bound at $R = 0+$ and meets the sphere-packing bound at $R = \frac{C_{ss}}{2}$. As explained in the beginning of this section, the bounds of Lemmas VI.4 and VI.6 can be combined by a well-known argument [2], [4], [5], [20], [21] based on a simple extension of the sphere-packing bound for list decoding. The result is that

$$\limsup_{\epsilon \to 0} \frac{E^{\epsilon}[R\epsilon^2]}{\epsilon^2} \leq \frac{C_{ss}}{2} - R, \qquad \text{for } 0 < R \leq \frac{C_{ss}}{4}.$$

This and the sphere-packing bound imply that for all $R > 0$

$$\limsup_{\epsilon \to 0} \frac{E^{\epsilon}[R\epsilon^2]}{\epsilon^2} \leq b(R)$$

where $b(R)$ denotes the right-hand side of (2). Combined with Lemma V.1, this proves Theorem II.2.

VII. Low SNR Asymptotics of Capacity for Block Rayleigh Fading

Consider the model of [12]. There are M transmit antennas, N receive antennas, and T symbol periods during which the $N \times M$ matrix H of fading coefficients is constant. The model for each channel use of this discrete-time memoryless channel is given by

$$Y_{in} = \sqrt{\rho/M} \sum_{m=1}^{M} Z_{\text{im}} H_{mn} + W_{in}, \qquad i = 1, \ldots, T, \; n = 1, \ldots, N$$

where Z is the channel input taking values in $\mathcal{C}^{MT}$ and Y is the channel output taking values in $\mathcal{C}^{NT}$. The fading coefficients H_{mn} and additive noise variables W_{in} are all mutually independent, $\mathcal{CN}(0, 1)$ random variables. Assume the transmitted signal satisfies the constraint $\text{Tr}(ZZ^{\dagger}) \leq MT$, so that the signal-to-noise ratio (SNR) at each receive antenna, averaged over the T output samples, is at most ρ. The notation for this section adheres to that in [12]. To map to our notation think of $X = \sqrt{\frac{\rho}{M}} Z$ as the input signal and set $\epsilon = \sqrt{\rho T}$ so the peak constraint becomes $\text{Tr}(XX^{\dagger}) \leq \epsilon^2$.

The conditional probability density for the channel is given by

$$q(Y|Z) = \frac{\exp\left(-\text{Tr}\left\{\left(I_T + \frac{\rho}{M} ZZ^{\dagger}\right)^{-1} YY^{\dagger}\right\}\right)}{\pi^{TN} \det^N\left(I_T + \frac{\rho}{M} ZZ^{\dagger}\right)}$$

where for a matrix V, $V^{\dagger}$ denotes the transpose of the complex conjugate of V. Therefore,

$$\frac{q(Y|Z)}{q(Y|0)} = \frac{\exp \text{Tr}\left\{\left(I_T + \frac{\rho}{M} ZZ^{\dagger}\right)^{-1} \frac{\rho}{M} ZZ^{\dagger} YY^{\dagger}\right\}}{\det^N\left(I_T + \frac{\rho}{M} ZZ^{\dagger}\right)}.$$

Thus, as ρ tends to zero

$$\Delta q(Y|Z) = \frac{\rho}{M}[\text{Tr}(ZZ^{\dagger}YY^{\dagger}) - NTr(ZZ^{\dagger})] + o(\rho). \tag{39}$$

Since $\Delta q(Y|Z)$ is linear in ρ as $\rho \to 0$, it is quadratic in ϵ as $\epsilon^2 \to 0$. Thus, $V(y, 0)$ is identically zero, so the Fisher information I_0 is zero, and, therefore, as in the special case $M = N = T = 1$ discussed in Section II, $C_{\epsilon}/\epsilon^2 \to 0$ as $\epsilon \to 0$. However, (39) is similar to (9) with ϵ replaced by ρ, so we can proceed as in Sections III and IV. Using (13), (14), and the fact $r \log r = (r-1)^2/2 + o((r-1)^2)$ yields

$$\begin{aligned}
\tilde{h}(Y|Z) &= \left(\frac{\rho^2}{2M^2}\right) E[E_0[(\text{Tr}(ZZ^{\dagger}YY^{\dagger}) - N\text{Tr}(ZZ^{\dagger}))^2]] + o(\rho^2) \\
&= \left(\frac{\rho^2}{2M^2}\right) E[E_0[\text{Tr}(ZZ^{\dagger}YY^{\dagger})^2 + N^2\text{Tr}(ZZ^{\dagger})^2 \\
&\qquad - 2N\text{Tr}(ZZ^{\dagger})\text{Tr}(ZZ^{\dagger}YY^{\dagger})]] + o(\rho^2),
\end{aligned}$$

while

$$E_0[\text{Tr}(ZZ^{\dagger}YY^{\dagger})] = N\text{Tr}(ZZ^{\dagger}) + o(\rho)$$

and

$$E_0[\text{Tr}(ZZ^{\dagger}YY^{\dagger})^2] = N^2\text{Tr}(ZZ^{\dagger})^2 + N\text{Tr}\{(ZZ^{\dagger})^2\} + o(\rho).$$

Also,

$$\begin{aligned}
\tilde{h}(Y) &= \left(\frac{\rho^2}{2M^2}\right) E_0[E_Z[(\text{Tr}(ZZ^{\dagger}YY^{\dagger}) - N\text{Tr}(ZZ^{\dagger}))]^2] + o(\rho^2) \\
&= \left(\frac{\rho^2}{2M^2}\right) E_0[E_Z[\text{Tr}(ZZ^{\dagger}YY^{\dagger})]^2 + N^2 E_Z[\text{Tr}(ZZ^{\dagger})]^2 \\
&\qquad - 2N E_Z[\text{Tr}(ZZ^{\dagger})] E_Z[\text{Tr}(ZZ^{\dagger}YY^{\dagger})]] + o(\rho^2)
\end{aligned}$$

while

$$E_0[E_Z[\text{Tr}(ZZ^{\dagger}YY^{\dagger})]^2] = N^2 E[\text{Tr}(ZZ^{\dagger})]^2 + N\text{Tr}\{E[ZZ^{\dagger}]^2\} + o(\rho).$$

Combining the above calculations yields

$$\tilde{h}(Y|Z) = -\frac{\rho^2 N}{2M^2} E[\text{Tr}\{(ZZ^{\dagger})^2\}] + o(\rho^2)$$

and

$$\tilde{h}(Y) = -\frac{\rho^2 N}{2M^2} \text{Tr}\{[E(ZZ^{\dagger})]^2\} + o(\rho^2).$$

By Marzetta and Hochwald [12, Theorem 2] it can be assumed that the input Z has the form $Z = \phi V$, where ϕ is a $T \times T$ isotropically distributed unitary matrix independent of

V, and V is a $T \times M$ random matrix that has nonnegative entries on the main diagonal and zeros off the main diagonal. The input constraint becomes $||\tilde{V}||^2 \le MT$, where $\tilde{V}$ is the main diagonal of V, which has dimension $\min(M, T)$. Let

$$d_i = \begin{cases} E\left[\tilde{V}_i^2\right], & \text{if } i \le \min(M, T) \\ 0, & \text{else} \end{cases}$$

and use properties of the distribution on ϕ to get

$$E[ZZ^\dagger]_{ij} = E\left[\sum_{k=1}^{T} \phi_{ki} d_k \phi_{kj}^\dagger\right] = \delta_{ij} \frac{\sum_k d_k}{T}.$$

Therefore,

$$\mathrm{Tr}[E[ZZ^\dagger]^2] = \frac{\left(\sum_k d_k\right)^2}{T} = \frac{E\left[\left\|\tilde{V}\right\|^2\right]^2}{T}.$$

Also,

$$E[\mathrm{Tr}\{(ZZ^\dagger)^2\}] = \sum_i \sum_j \sum_k \sum_{k'} E\left[\left|\tilde{V}_k\right|^2 \left|\tilde{V}_{k'}\right|^2\right] \cdot E[\phi_{ik}\phi_{ik'}^* \phi_{jk}^* \phi_{jk'}]. \quad (40)$$

The coefficient of $E[|\tilde{V}_k|^2 |\tilde{V}_{k'}|^2]$ in (40) is given by

$$\sum_i \sum_j E[\phi_{ik}\phi_{ik'}^* \phi_{jk}^* \phi_{jk'}] = E\left[\left|\sum_i \phi_{ik}\phi_{ik'}^*\right|^2\right] = \begin{cases} 0, & \text{if } k \ne k' \\ 1, & \text{if } k = k' \end{cases}$$

so that

$$E[\mathrm{Tr}\{(ZZ^\dagger)^2\}] = \sum_k E\left[\left|\tilde{V}_k\right|^4\right].$$

Hence,

$$I(Z;\, Y) = \frac{N}{2M^2}\rho^2 \left(\left(\sum_k E\left[|\tilde{V}_k|^4\right]\right) - \frac{E\left[\left\|\tilde{V}\right\|^2\right]^2}{T}\right) + o(\rho^2).$$

Next, find the distribution on $\tilde{V}$ to maximize the coefficient of ρ^2 in this expression for $I(Z; \; Y)$. For a given value of $E[||\tilde{V}||^2]$ and given the constraint $||\tilde{V}||^2 \le MT$, the quantity $\sum_k E[|\tilde{V}_k|^4]$ is maximized over distributions on $\tilde{V}$ by taking $\tilde{V}$ to be distributed over the two point set $\{(0, \ldots, 0)^\dagger, (\sqrt{MT}, 0, 0, \ldots, 0)^\dagger\}$. That is, either the zero signal is sent, or the peak energy is all put into the first transmit antenna. Maximizing over the one-dimensional family of such distributions for $\tilde{V}$ yields that the optimal "on" probability is given by

$$P\left(\tilde{V} = (\sqrt{MT}, 0, 0, \ldots, 0)^\dagger\right) = \begin{cases} 1/2, & \text{if } T = 1 \\ 1, & \text{if } T \ge 2. \end{cases}$$

More generally, the single antenna used could be randomly chosen, but the distribution of $\tilde{V}$ is not optimal if the event that more than one antenna is used has positive probability.

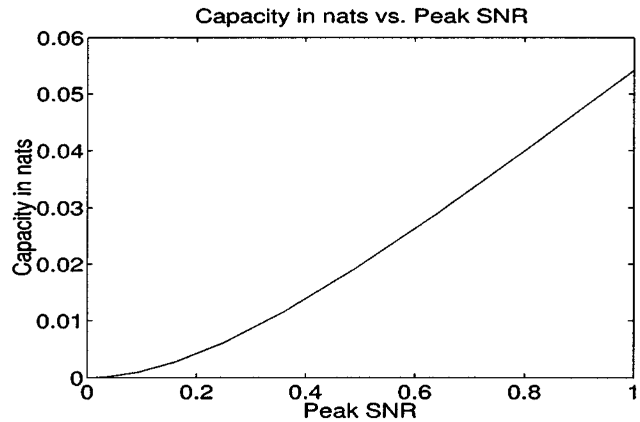

Fig. 2. Capacity in nats as a function of ρ for the peak-constrained Rayleigh-fading channel.

Therefore, if $\hat{C}(M, N, T, \rho)$ denotes the capacity per unit time for the Marzetta and Hochwald model (obtained by dividing the mutual information by T), then

$$\lim_{\rho \to 0} \frac{\hat{C}(M, N, T, \rho)}{\rho^2} = \begin{cases} \frac{N}{8}, & \text{if } T = 1 \\ \frac{(T-1)N}{2}, & \text{if } T \ge 2. \end{cases} \quad (41)$$

Note that the limiting normalized capacity is proportional to the number of receive antennas N, while it does not depend at all on the number of transmit antennas. Also, the capacity per unit time increases linearly with $T-1$. Such a large rate of increase with T reflects the fact that if very little energy is transmitted, every increase in T is very valuable in helping the channel to be estimated.

Hassibi and Hochwald [8] consider the use of training-based strategies. Comparing a tight bound they found for the capacity of training-based schemes to (41), it follows that for small ρ, the training-based strategy using one transmit antenna achieves about half of the peak-constrained channel capacity. This is consistent with the fact that the strategies of [8] do not use a highly peaked input distribution. Hassibi and Hochwald point out in [8] that this is far from the capacity for average, rather than peak, energy constraints, for which the capacity tends to zero as ρ rather than as ρ^2 [19].

Although this paper focuses on the asymptotics of capacity as the peak energy tends to zero, we briefly discuss computing the capacity numerically for finite values of the peak constraint for $M = N = T = 1$. Following the proof of [1], after removing the average energy constraint, it is not hard to show that a capacity achieving distribution for the peak-constrained discrete memoryless Rayleigh channel is discrete with a nonzero mass at 0. The capacity calculation problem is then equivalent to finding the locations and masses of the components of the distribution in order to maximize the mutual information. The conditional gradient algorithm given in [1] can be used to numerically compute the capacity. Fig. 2 displays $\hat{C}(1, 1, 1, \rho)$ as a function of the peak SNR ρ (not in decibels), and Fig. 3 displays $\hat{C}(1, 1, 1, \rho)/\rho$. The limiting slope as $\rho \to 0$ in Fig. 3 is $1/8$, as required by (41). To appreciate how much smaller this capacity is than the capacity of an additive Gaussian noise channel, or more generally a Rician channel, recall the first example of Section II. As a function of ρ, the capacity of the Rician

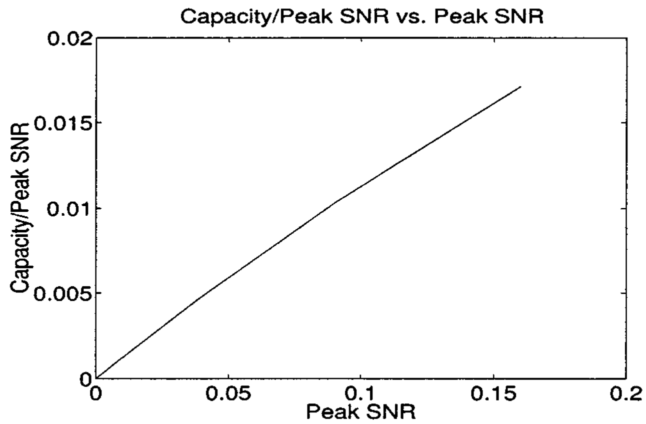

Fig. 3. Capacity *divided* by ρ as a function of ρ for the peak-constrained Rayleigh-fading channel.

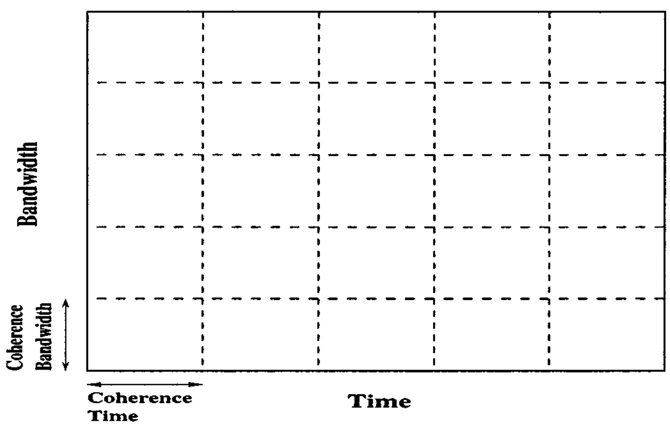

Fig. 4. A visualization of the WSSUS channel in a time-frequency band.

channel has slope α^2 at $\rho = 0$, so slope one for the Gaussian channel.

VIII. Application to Broad-Band Fading Channels

In this section, the result of the previous section is applied to find the asymptotic capacity of a wide-band time-varying multipath channel with Rayleigh fading and a burstiness constraint on the transmitted signal. A reasonable picture of a wide-sense-stationary and uncorrelated scattering (WSSUS) channel is shown in Fig. 4. Let W be the signal bandwidth, P be the power of the input signal, and N_0 be the noise power spectral density. The time–frequency plane is divided into blocks of duration T_{coh} and bandwidth W_{coh}, where T_{coh} is the coherence timewidth and W_{coh} is the coherence bandwidth. Ignore the effects of intersymbol interference and edge effects between blocks, to arrive at the following model. Assume that during each block, $T = T_{\text{coh}} W_{\text{coh}}$ symbols can be transmitted and the channel gain for a block is Rayleigh distributed and identical for all symbols of the block. The gains for different blocks are assumed independent. Assume that the transmitted signal energy in each coherence block is peak constrained to $\frac{PT}{W}$. Such constraint is satisfied by modulation schemes that are not bursty in the time–frequency plane, such as direct-sequence spread spectrum. Since there are W coherence blocks per unit time, the capacity per unit time is given by $W\hat{C}(1, 1, T, \rho)$, where $\rho = \frac{P}{N_o W}$. It is reasonable to take $\rightarrow\infty$ with T fixed with value $T \geq 2$. For example, if $T_{\text{coh}} = 10$ ms for 100-Hz doppler spread and $W_{\text{coh}} = 1$ MHz for 1-μs delay spread, then $T = 10^4$. The asymptotic result (41) yields (for $T \geq 2$ fixed) that the capacity per unit time of the wide-band fading channel with peak energy constraint per coherence block is given by

$$\frac{P^2(T_{\text{coh}} W_{\text{coh}} - 1)}{2N_o^2 W} + o\left(\frac{P^2}{N_o^2 W}\right).$$

Thus, as $W \rightarrow \infty$ for fixed P/N_o, T_{coh} and W_{coh}, the capacity per unit time tends to zero. This result was first derived by [7]. The model of Médard and Gallager is more extensive in that it allows for continuous dependence among blocks in both space and time. While Médard and Gallager constrain peakiness by imposing a fourth moment constraint on the transmitted symbols, we impose a peak constraint on the energy transmitted per block.

IX. Conclusion

Under mild regularity conditions, both the capacity and reliability function of channels with small peak constraints are closely related to the Fisher information in a straightforward way. The asymptotic shape of the reliability function is the same as observed earlier for the power-constrained infinite-bandwidth white noise channel [5, p. 381], Gallager's very noisy channel with finite inputs [5, Example 3, pp. 147–149], and the very noisy Poisson channel studied [23], [24]. These channels are similar in that they can be viewed as very large bandwidth or very large noise channels, and the relevant log-likelihood ratios asymptotically have the same exponential moments as Gaussian random variables. These channels are among the very few channels for which the reliability function is completely known.

Two extensions of the first part of the paper may be possible, but are left for future work. Recently, [3] extended Wyner's results to identify the reliability function region for the two-user multiple-access problem. Such an extension might hold in the setting of this paper. Another idea is to find a single result that includes the setting of this paper, Gallager's finite input very noisy channel, and discretized versions of the Poission and infinite bandwidth white Gaussian noise channels. The set $\mathcal{D}$ in the theorems of this paper would be replaced by any closed bounded subset of $\mathcal{R}^n$. Gallager's model correspondes to taking $\mathcal{D}$ to be the set of n unit vectors pointing along the n coordinate axes in positive directions, and taking $q(y|x)$ linear in x: $q(y|x) = q(y|0) + x^T V(y)$ for small $||x||$. In addition, an average constraint for each codeword could be imposed. The capacity C_{ss} would be given by maximizing a quadratic form involving K subject to constraints. Such generalization in not pursued in this paper, for apparently it would require a different proof technique for the reliability function results. It might involve showing that a large but finite input alphabet suffices to approach capacity and error exponents to within ϵ.

As shown in the second part of the paper, small peak signal asymptotics are informative in the case of Rayleigh fading, even though the Fisher information matrix is zero. In particular, an expression for the asymptotic capacity in the case of block fading with multiple transmit and multiple receive antennas, shows that within the class of transmission strategies with constant transmit energy per fading block, the training sequence based scheme comes within a factor of two of optimality, in the range of low SNR. Also, a simple result of the type of [7] is obtained.

REFERENCES

[1] I. C. Abou-Faycal, M. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.

[2] R. E. Blahut, *Principles and Practice of Information Theory*. Owego, NY: Addison-Wesley, 1988.

[3] S. I. Bross, M. V. Burnashev, and S. Shamai (Shitz), "Error exponents for the two-user Poisson multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1999–2016, July 2001.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Akadémiai Kiadó, 1981.

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

[6] ——, "Energy limited channels: Coding, multiaccess, and spread spectrum," LIDS, MIT, Cambridge, MA, Tech. Rep. LIDS-P-1714, Nov. 1987.

[7] M. Médard and Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 840–852, Apr. 2002.

[8] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inform. Theory*, submitted for publication.

[9] B. Hochwald and T. Marzetta, "Unitary space–time modulation for multi-antenna communications in Rayleigh falt-fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[10] I. A. Ibragimov and R. Z. Khas'minskii, "Asymptotic behavior of certain statistical estimates in the smooth case (I)," *Theory Probab. Appl.*, vol. 17, pp. 445–462, 1972.

[11] ——, "Weak signal transmission in a memoryless channel," *Probl. Pered.. Inform.*, vol. 8, pp. 28–39, Oct.–Dec. 1972. English translation in *Probl. Inform. Transm*.

[12] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–158, Jan. 1999.

[13] M. S. Pinsker and V. V. Prelov, "Information rates in stationary Gaussian channels under transmission with weak signals," *Probl. Pered. Inform.*, vol. 30, pp. 3–11, Oct.–Dec. 1994. English translation in *Probl. Inform. Transm*.

[14] M. S. Pinsker, V. V. Prelov, and E. van der Meulen, "Information rates in certain non-Gaussian channels with weak signal transmissions," *Probl. Pered. Inform.*, vol. 34, pp. 3–17, Jan.– Apr. 1998. English translation in *Probl. Inform. Transm*.

[15] ——, "Stationary channels with a totally singular process as a random parameter," *Probl. Pered. Inform.*, vol. 35, pp. 3–12, Jan.–Apr. 1999. English translation in *Probl. Inform. Transm*.

[16] M. S. Pinsker, V. V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1877–1888, Nov. 1995.

[17] V. V. Prelov, "Asymptotic behavior of the capacity of a continuous channel with a large amount of noise," *Probl. Pered. Inform.*, vol. 6, pp. 40–57, Apr.–June 1970. English translation in *Probl. Inform. Transm*.

[18] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signal," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1728–1735, Sept. 1993.

[19] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2619–2692, Oct. 1999.

[20] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I," *Inform. Contr.*, vol. 10, pp. 65–103, 1967.

[21] ——, "Lower bounds to error probability for coding on discrete memoryless channels II," *Inform. Contr.*, vol. 10, pp. 522–552, 1967.

[22] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[23] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel: I," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1449–1461, Nov. 1988.

[24] ——, "Capacity and error exponent for the direct detection photon channel: II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.