An Induced Additive-Noise Model for Memoryless Rayleigh-Fading Channels

Wenyi Zhang, *Student Member, IEEE,* and
J. Nicholas Laneman, *Member, IEEE*

***Abstract*—This correspondence investigates a noncoherent discrete-time memoryless Rayleigh-fading channel. A logarithmic transform converts it into an induced channel with additive noise that is independent of the channel input. From this perspective, it is natural and convenient for us to revisit several known results and gain new insights. In particular, we specify a class of simple, log-scale uniform channel input distributions that performs well for moderate to high signal-to-noise ratio (SNR). Furthermore, a continuous-amplitude log-scale uniform distribution is asymptotically capacity-achieving in the sense that the difference between the resulting mutual information and the channel capacity approaches zero as SNR becomes large. We also extend the induced additive-noise channel approach to memoryless multiple-antenna channels possibly with transmit, but without receive, spatial correlation.**

***Index Terms*—Additive noise, capacity, high signal-to-noise ratio (SNR), memoryless fading, mutual information, noncoherent, Rayleigh fading.**

I. Introduction

In contrast to the thorough understanding of the additive white Gaussian noise (AWGN) channel, much less is understood about the noncoherent discrete-time memoryless Rayleigh-fading channel. The simplest case is that for which the channel is scalar, i.e.

$$X = S \cdot H + Z \tag{1}$$

where $S \in \mathbb{C}$ is the channel input, and $X \in \mathbb{C}$ is the channel output. The fading coefficient $H \in \mathbb{C}$ and the additive noise $Z \in \mathbb{C}$ are both zero-mean circular complex Gaussian random variables, and independent for different channel uses. Hence in (1) we suppress all of the time indexes. Throughout the correspondence we consider the case in which the realization of H, the fading coefficient, is available to neither the transmitter nor to the receiver. For sake of simplicity and without loss of generality, we let H and Z both be of unit variance, and S be average-power-limited so that

$$\mathcal{E}\left[|S|^2\right] = P.$$

Consequently, the average signal-to-noise ratio (SNR) of the channel becomes

$$\rho = \frac{\mathcal{E}\left[|H|^2\right] \cdot \mathcal{E}\left[|S|^2\right]}{\mathcal{E}\left[|Z|^2\right]} = P.$$

Neither the capacity nor the capacity-achieving input distribution of the channel (1) is fully known. The authors of [1] prove that the capacity is achieved by a discrete input distribution with a finite number of mass points, including a mass point at $S = 0$. However, they do not provide any specific signaling design, partly because the proof technique employed is nonconstructive in nature. Furthermore, the capacity-achieving input distribution changes for different SNR; the number of mass points, their locations, and their probabilities, all depend on the operating SNR, and can only be determined through numerical nonlinear optimization. In [2], the authors obtain upper and lower bounds on the channel capacity, constructing an explicit discrete input distribution to give the lower bound. They show that, at high SNR, the channel capacity grows double-logarithmically, that is, $C = \mathcal{O}(\log\log\rho)$ as $\rho \to \infty$. This double-logarithmic growth rate is refined in [3], [4] for more general models such as Ricean fading, nonmemoryless fading processes, and multiple-antenna channels. It is shown that for very general channel models, the channel capacity behaves as $C = \log\log\rho + \chi + o(1)$ as $\rho \to \infty$, where χ is called the *fading number* of the channel. Another relevant work is [5], in which the authors derive necessary and sufficient conditions for continuous input distributions that lead to unbounded mutual information as SNR goes to infinity.

In addition to the high-SNR analyzes mentioned above, for sufficiently low SNR, it is known that the channel capacity is achieved by on–off keying (OOK), for which the nonzero input has amplitude that becomes unbounded as SNR approaches zero [1], [6]. Numerical results [1] indicate that, for SNR less than roughly −3.5 dB, optimized OOK is capacity-achieving, and for SNR less than 10 dB, there is at most a 4% gap to capacity from the mutual information achieved by optimized OOK.

The main idea of this correspondence stems from the following observation: As SNR approaches infinity, most of the channel randomness results from the fading coefficient H rather than the additive noise Z. Intuitively we may approximate the limiting channel behavior at high SNR as $\log|X| = \log|S| + \log|H|$, i.e., a channel with additive noise only. This limiting approximation is employed as a proof technique in [5]. In this correspondence, we further observe that such an additive-noise channel interpretation holds for *all* SNR.[1] Specifically, by taking the logarithm of the channel output's magnitude $|X|$, we can decompose the transformed channel output into the sum of two terms. One term is merely a deterministic function of the channel input S, and the other term is independent of the channel input and can therefore be viewed as an additive noise. This equivalent additive-noise channel model is conceptually more intuitive to communication engineers, and yields a natural geometric interpretation for many earlier results. The effects of both the multiplicative fading and the additive noise in the original channel (1) are then described by a single induced additive noise. The inefficiency of communication over the original fading channel (1) can be intuitively understood as the log-scale transform rescales the original SNR ρ to roughly $\log\rho$. Consequently, the double-logarithmic capacity behavior immediately follows.

Remark: For memoryless Rayleigh-fading channels, the additive-noise channel interpretation is not an entirely new idea. In addition to information-theoretic analyzes, practical coding design has been investigated in [7] and [8] for the special case that the input S is symmetric OOK and all codewords have constant weight. In particular, [7] proposes another induced additive-noise channel model to simplify the receiver design, but this model only applies to the special case of constant-weight OOK inputs.

We briefly summarize the main results of the correspondence as follows.

Manuscript received February 18, 2005; revised December 12, 2005. This work was supported in part by the State of Indiana through the 21st Century Research and Technology Fund, and by the National Science Foundation under Contract ECS03-29766. The material in this correspondence was presented in part at the 39th Conference on Information Sciences and Systems (CISS), Baltimore, MD, 2005.

The authors are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: wzhang1@nd.edu; jnl@nd.edu).

Communicated by M. Médard, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2006.871593

[1]The authors of [1] apparently observed the same additive-noise channel interpretation. However, they only used it implicitly in a proof [1, Appendix I, Lemma 2], and did not publish their other developments in this direction.

1) In Section II, we derive the induced additive-noise channel model from the original fading channel model (1). The two channel models are equivalent, thus the problem is converted to communication over an additive-noise channel with non-Gaussian noise.
2) Based upon the induced additive-noise channel perspective, in Section III, we establish new necessary and sufficient conditions for channel input distributions to achieve unbounded mutual information as SNR $\rho \to \infty$, and to asymptotically achieve the channel capacity. These conditions are qualitatively similar to those established in [5]. However, the induced channel model involves much less mathematical machinery, and appears to be more convenient to apply.
3) In Section III, we further show that a simple input distribution called the log-scale (continuous) uniform distribution is asymptotically capacity-achieving, i.e., it achieves both the double-logarithmic growth with SNR and the fading number. To provide practical signaling design, we re-interpret a class of discrete input distributions, which, in the induced channel perspective, turns out to be nothing but pulse amplitude modulation (PAM) with uniformly spaced pulses. For moderate SNR we demonstrate that these discrete input constellations with appropriate size achieve mutual information higher than that achieved by the log-scale (continuous) uniform input.
4) In Section IV, we further extend the induced channel modeling approach to memoryless multiple-antenna channels, possibly with transmit, but without receive, correlation. It is shown that essentially all the developments for the scalar channel case can be paralleled for the multiple antenna case.

We use capital letters to represent random variables and small letters to represent their sample values. Bold font distinguishes vectors and matrices from scalars. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{T}}$ ($\boldsymbol{A}^{\dagger}$) represents its transpose (conjugate transpose). $\mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma})$ ($\mathcal{CN}(\boldsymbol{m}, \boldsymbol{\Sigma})$) represents a Gaussian (circular complex Gaussian) distribution with mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$. We frequently use Euler's constant $\gamma = 0.5772\ldots$. Throughout the correspondence, all logarithms are taken with natural base e.

II. Log-Scale Transform: From Fading Channel to Additive-Noise Channel

For the scalar channel (1), because the fading coefficient H totally distorts the phase information contained in the channel input S, we can restrict S to be real and nonnegative. The channel output X has a sufficient statistic $|X|^2 = XX^{\dagger}$, i.e., its energy, or equivalently, $|X|$, i.e., its magnitude. Conditioned on $S = s$, we have $X \sim \mathcal{CN}(0, s^2+1)$, and its energy $R = |X|^2$ is exponentially distributed with probability density function (pdf)

$$f_{R|S}(r|s) = \begin{cases} \frac{1}{s^2+1}\exp\left(-\frac{r}{s^2+1}\right), & \text{if } r \geq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Hence the channel $S \to R$ can be viewed in multiplicative form as

$$R = (S^2+1)V \quad (3)$$

where V is independent of S and has pdf $f_V(v) = \exp(-v)$ for $v \geq 0$.

After taking the logarithm of $|X| = \sqrt{R} = \sqrt{S^2+1}\cdot\sqrt{V}$, we obtain an induced additive-noise channel equivalent to the original fading channel (1) as

$$T = U + W \quad (4)$$

where we have the following.

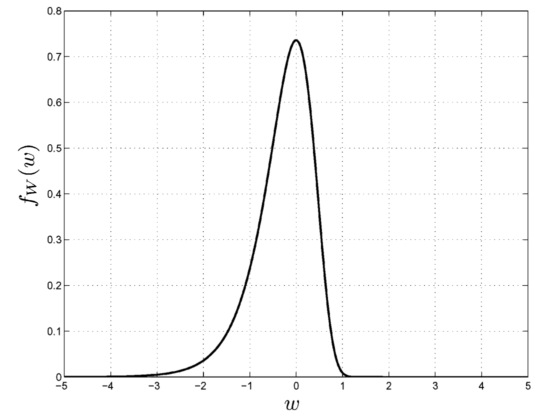

Fig. 1. The pdf of the induced additive noise W for the scalar channel case.

- The channel input is

$$U = \frac{1}{2}\log(S^2+1). \quad (5)$$

From the average power constraint for S, U should satisfy

$$\mathcal{E}\left[\exp(2U)\right] = \rho + 1 \quad (6)$$

and have support set $U \in [0, \infty)$.
- The channel output is

$$T = \log|X|. \quad (7)$$

- The additive noise $W = (\log V)/2$ is independent of U, and has pdf

$$f_W(w) = 2\exp\left[2w - \exp(2w)\right], \quad \text{for } w \in (-\infty, \infty). \quad (8)$$

Remark: The induced channel (4) is equivalent to the normalized channel (1), in which the fading coefficient H and the noise Z are both $\mathcal{CN}(0,1)$ random variables. For general parameterizations, i.e., $H \sim \mathcal{CN}(0, \sigma_H^2)$ and $Z \sim \mathcal{CN}(0, \sigma_Z^2)$, the induced channel (4) still holds, except that the induced channel input U becomes

$$U = \frac{1}{2}\log(\frac{\sigma_H^2}{\sigma_Z^2}S^2 + 1)$$

with constraint

$$\mathcal{E}\left[\exp(2U)\right] = \frac{\sigma_H^2}{\sigma_Z^2}\cdot P + 1$$

and support set $U \in [0, \infty)$. The definitions of T and W do not change under different parameterizations.

Fig. 1 illustrates the pdf of the additive noise W. We observe that the noise density is not symmetric, and it decreases much slower for $w < 0$ than for $w > 0$. These qualitative observations are more precisely quantified by the following lemma.

Lemma 2.1: (Properties of the induced additive noise W, the scalar channel case) The additive noise W in (8) has the following properties.

1) The cumulative distribution function (cdf) of W is

$$F_W(w) = 1 - \exp\left(-\exp(2w)\right). \quad (9)$$

2) The characteristic function of W is

$$\varphi_W(j\theta) = \Gamma(1 + j\theta/2) \quad (10)$$

where $\Gamma(z) = \int_0^{\infty} x^{z-1}\exp(-x)dx$ is the gamma function [9].
3) The mean of W is

$$\mathcal{E}[W] = -\frac{\gamma}{2} \quad (11)$$

and the maximum of $f_W(w)$ is attained at $w = 0$.

4) The variance of W is

$$\sigma_W^2 = \frac{\pi^2}{24}. \tag{12}$$

5) The differential entropy of W is

$$\begin{aligned} h(W) &= -\int f_W(w) \log f_W(w) dw \\ &= 1 - \log 2 + \gamma. \end{aligned} \tag{13}$$

Proof: By specializing to $N = 1$ in Appendix C.

III. Channel Mutual Information

The induced additive-noise channel (4) is convenient for studying the channel behavior. In this section we investigate the channel mutual information, focusing on the high SNR regime.

An asymptotic characterization of the channel capacity is given by [3]

$$\lim_{\rho \to \infty} (C - \log \log \rho) = \chi \tag{14}$$

where χ is called the *fading number*, and $\chi = -1 - \gamma$ for the scalar memoryless Rayleigh case.

From (4) we can write the channel mutual information as

$$\begin{aligned} I(S; X) &= I(U; T) \\ &= h(T) - h(T|U) = h(T) - h(W). \end{aligned} \tag{15}$$

From Lemma 2.1, $h(W) = 1 - \log 2 + \gamma$, hence finding channel capacity corresponds to maximizing the differential entropy of T. However, maximization of $h(T)$ is nontrivial because one of the constraints is that $T = U + W$, with the support set of U being $[0, \infty)$. In Appendix A we relax this constraint to revisit a capacity upper bound previously obtained in [2].

A. Conditions for Good Input Distributions

We have the following upper/lower bounds for any given distribution of the input U:

$$I(U; T) \le \frac{1}{2} \log \sigma_U^2 + \log \sqrt{2\pi e \left(1 + \frac{\sigma_W^2}{\sigma_U^2}\right)} - h(W) \tag{16}$$

$$I(U; T) \ge h(U) + \log \sqrt{1 + e^{-2[h(U) - h(W)]}} - h(W). \tag{17}$$

The upper bound (16) is a straightforward application of the property that the Gaussian distribution maximizes the entropy over all distributions with the same variance [10, Theorem 9.6.5], and the lower bound (17) can be obtained using the entropy power inequality [10, Theorem 16.7.1]. According to the constraint (6), the channel input U implicitly depends on the SNR ρ. From (17) we observe that if $h(U) \to \infty$ as $\rho \to \infty$ then $I(U; T) \to \infty$; on the other hand, from (16) we observe that if $I(U; T) \to \infty$ as $\rho \to \infty$ then $\sigma_U^2 \to \infty$. We thus obtain necessary and sufficient conditions for unboundedness of channel mutual information as SNR scales.

Proposition 3.1: For the channel (4), the channel input U in (5) achieves $I(U; T) \to \infty$ as $\rho \to \infty$ if

$$h(U) \to \infty, \quad \text{as } \rho \to \infty. \tag{18}$$

On the other hand, if U achieves $I(U; T) \to \infty$ as $\rho \to \infty$, then

$$\sigma_U^2 \to \infty, \quad \text{as } \rho \to \infty. \tag{19}$$

Neither of the above sufficient and necessary conditions can be simultaneously necessary and sufficient. Two simple counterexamples using discrete channel inputs are given as follows. On one hand, since the channel capacity is always achieved by discrete inputs whose differential entropies are $-\infty$, $h(U) \to \infty$ is not necessary. On the other hand, a symmetric OOK input leads to unbounded variance as $\rho \to \infty$, but the resulting channel mutual information is always upper bounded by $\log 2$ nats.

It is useful to compare the conditions in Proposition 3.1 to those obtained in [5]. A necessary and sufficient condition is that $h(T) \to \infty$ as $\rho \to \infty$ [5, Theorem 4.3, eq. (2)]. This condition is obvious from (15). On the other hand, a sufficient condition is that $h(\log |S|) \to \infty$ as $\rho \to \infty$ [5, Theorem 4.3, eq. (3)]. This condition is clearly similar to the sufficient condition in Proposition 3.1, except for slightly different transforms of the channel input. Finally, there is no necessary condition in [5] that parallels the necessary condition in Proposition 3.1.

Proposition 3.1 immediately explains why circular complex Gaussian inputs perform poorly for the channel (1). In fact we have the following corollary.

Corollary 3.2: If $S \sim \mathcal{CN}(0, \rho)$ in (1), then its induced input U in (5) has

$$h(U) \to h(W) = 1 - \log 2 + \gamma \tag{20}$$

$$\sigma_U^2 \to \sigma_W^2 = \frac{\pi^2}{24} \tag{21}$$

as $\rho \to \infty$.

Proof: Noting that for $S \sim \mathcal{CN}(0, \rho)$, $f_U(u) = \exp(1/\rho) \cdot f_W(u - \log \rho/2)$, the result then follows from straightforward manipulations using Lemma 2.1. Q.E.D.

Comparing the mutual information bounds (16), (17) and the capacity asymptote (14), we further obtain necessary and sufficient conditions for input distributions that asymptotically achieve capacity, i.e., achieve both the double-logarithmic growth with SNR and the fading number.

Proposition 3.3: In order to asymptotically achieve the channel capacity for the channel (4), i.e.

$$\lim_{\rho \to \infty} (C - I(U; T)) = 0$$

a sufficient condition on the input U in (5) is

$$\lim_{\rho \to \infty} (h(U) - \log \log \rho) \ge -\log 2 \tag{22}$$

and a necessary condition is

$$\lim_{\rho \to \infty} \left(\sigma_U^2 - \frac{(\log \rho)^2}{8\pi e}\right) \ge -\sigma_W^2. \tag{23}$$

B. Log-Scale Continuous Uniform Inputs

Now we introduce the log-scale continuous uniform input distribution as follows.

Definition 3.4: A log-scale continuous uniform (LCU) input U for the induced additive-noise channel (4) has pdf

$$f_U(u) = \begin{cases} \frac{1}{A}, & \text{if } 0 \le u \le A \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

where $A > 0$ is determined by

$$\frac{\exp(2A) - 1}{2A} - 1 = \rho. \tag{25}$$

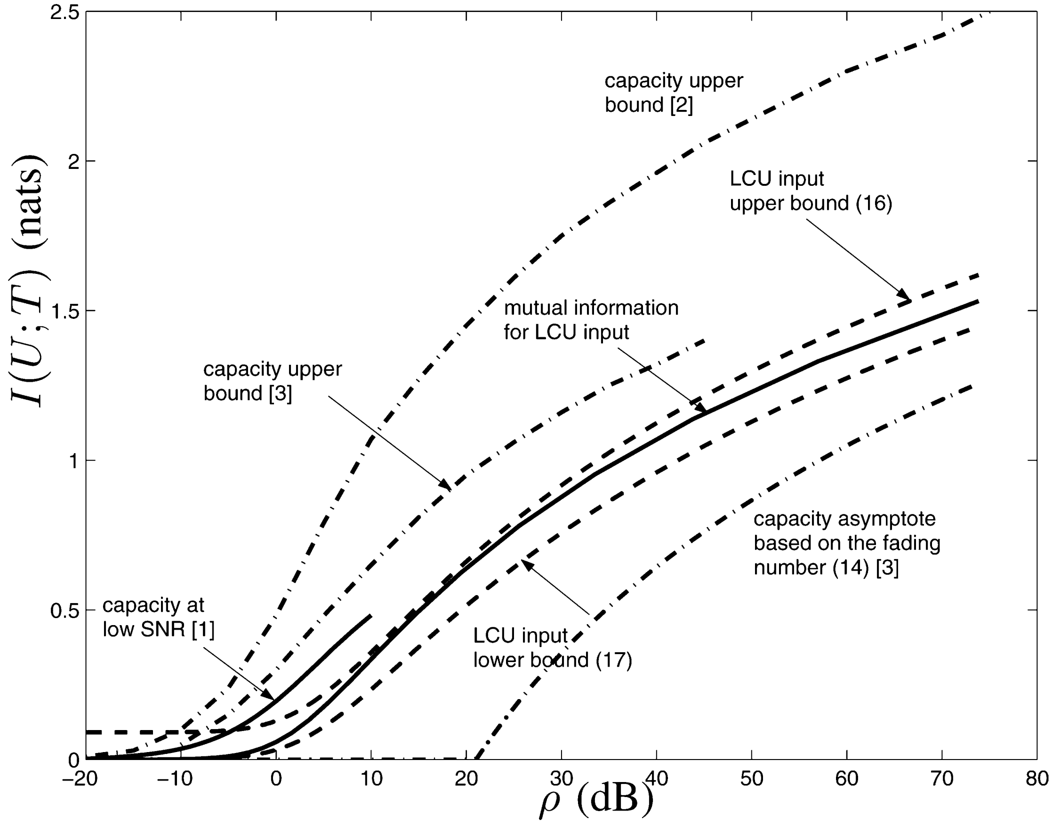

Fig. 2. Comparison of different channel mutual information and capacity bounds versus SNR.

For LCU inputs, the channel output T has pdf

$$\begin{aligned} f_T(t) &= \int f_U(u) f_W(t-u)du \\ &= \frac{(F_W(t) - F_W(t-A))}{A}. \end{aligned}$$

We can then numerically evaluate the corresponding $I(U;T)$. Furthermore, the LCU distribution is asymptotically capacity-achieving.

Corollary 3.5: For the induced additive-noise channel (4), LCU inputs achieve

$$\lim_{\rho \to \infty} (C - I(U;T)) = 0. \tag{26}$$

Proof: The parameter A in the LCU distribution is related to the SNR through (25), so that

$$\rho = \frac{\exp(2A) - 1}{2A} - 1 < \frac{\exp(2A)}{2A} < \exp(2A)$$

where the last inequality holds if SNR is sufficiently large. Applying Proposition 3.3, we have

$$\begin{aligned} \lim_{\rho \to \infty} (h(U) - \log\log\rho) &= \lim_{\rho \to \infty} (\log A - \log\log\rho) \\ &\geq \lim_{\rho \to \infty} \left(\log \left(\frac{1}{2} \log \rho \right) - \log\log\rho \right) \\ &= -\log 2. \end{aligned}$$

Hence, LCU inputs are sufficient for asymptotically achieving channel capacity. Q.E.D.

A related class of asymptotically capacity-achieving inputs is employed in [3]. Specifically, in [3] the input distribution is uniform for $\log|S|$, over an interval $[\log s_{\min}, (1/2)\log\rho]$, where $s_{\min}$ escapes to infinity, while $\log s_{\min}/\log\rho$ vanishes, as SNR increases. In fact, we can show that the LCU distribution is still asymptotically capacity-achieving even if its lower limit is chosen the same as $\log s_{\min}$. For finite SNR, however, there is a rate loss in increasing the lower limit from zero.

For LCU inputs in the high SNR limit, the lower bound (17) attains the channel capacity, and the upper bound (16) cannot be met, with a gap of $\log\sqrt{\pi e/6} \approx 0.1765$ nats. For finite SNR, both these bounds are useful in characterizing the achievable rate. In Fig. 2, we plot the lower/upper bounds and actual mutual information for LCU inputs. For comparison, we also plot the numerically computed channel capacity at low SNR [1], the capacity upper bound from [2], the capacity upper bound from [3] based upon duality methods, and the capacity asymptote based upon the fading number (14) [3].

Several observations can be made based upon Fig. 2. First, the actual mutual information for LCU inputs is close to its lower bound for low and high SNR, and its upper bound is a tight estimate for SNR between 10 and 30 dB. Second, at high SNR the gap between the capacity asymptote based upon the fading number (14) and the mutual information lower bound (17) is approximately 0.2 nats, even at $\rho \approx$ 80 dB. This indicates that the approximation of capacity based on the fading number may under-estimate the actual capacity for a wide range of SNR. As a further example, these two curves remain separated by about 0.05 nats at $\rho \approx 400$ dB. Finally, at low SNR, there exists a significant rate loss from LCU inputs compared to optimized OOK, which is capacity-achieving in that regime. This loss results because LCU inputs only yield $A = \rho + o(\rho)$ in the low SNR limit, which leads to a vanishing fourthegy [11] for the original fading channel (1).

C. Log-Scale Discrete Uniform Inputs

In this section, we consider discrete input distributions. First, practical systems usually employ discrete signal sets as modulation constellations. Second, as shown in [1], the channel capacity is achieved by a

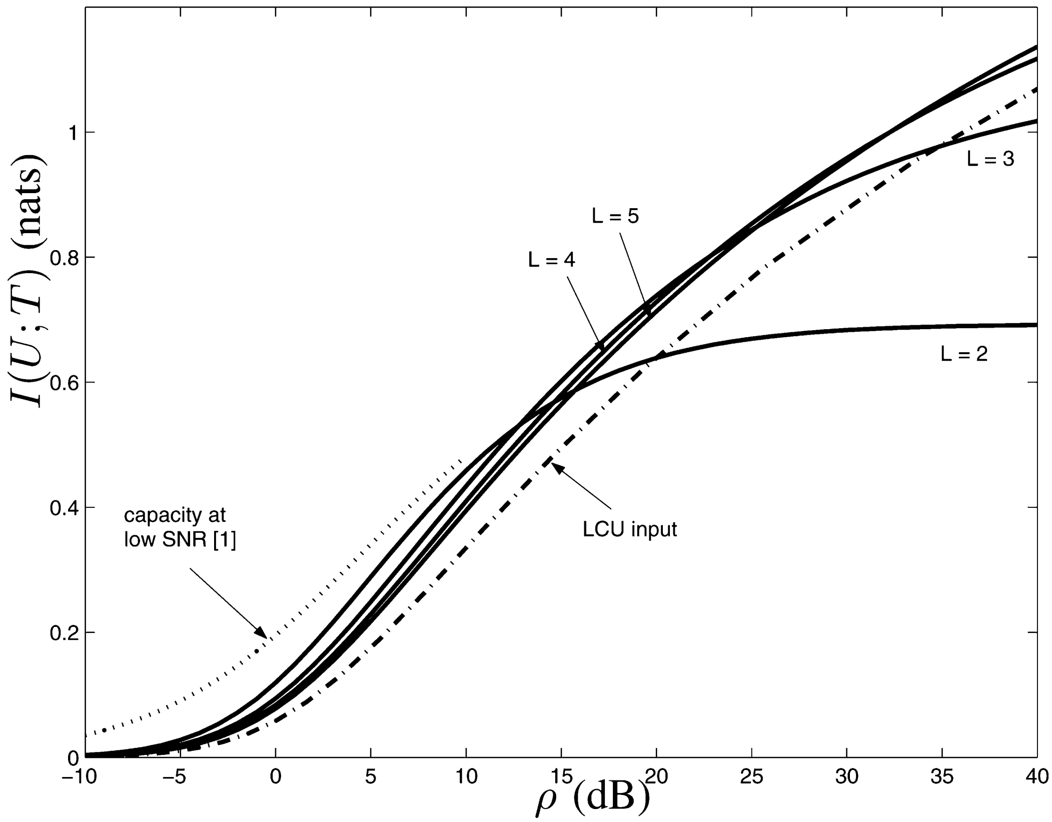

Fig. 3. Mutual information versus SNR for L-ary LDU ($L = 2, 3, 4, 5$), LCU, and optimized OOK inputs.

discrete input distribution. Motivated by the LCU distribution, and the idea of maximizing minimum distance in the induced additive-noise channel, we introduce the following log-scale discrete uniform input distribution.

Definition 3.6: An L-ary log-scale discrete uniform (LDU) input U for the induced additive-noise channel (4) has probability mass function (PMF)

$$p_l \stackrel{\text{def}}{=} \text{Prob}\,(U = l\Delta) = \frac{1}{L}, \quad \text{for } l = 0, \dots, L-1 \tag{27}$$

where the spacing Δ is determined by SNR through the equation

$$\frac{\exp(2L\Delta) - 1}{\exp(2\Delta) - 1} = L(\rho + 1). \tag{28}$$

We note that LDU inputs are precisely those proposed in [12, Theorem 3], where it is shown that these inputs maximize the Kullback–Leibler (KL) distance between the conditional output pdfs. From the induced additive-noise channel perspective, an LDU input is simply log-scale PAM with uniformly spaced signals. For $L = 2$, an LDU input reduces to symmetric OOK, and for $L \to \infty$ it approximates an LCU input in the sense of convergence in distribution.

Fig. 3 displays the results of numerically computing $I(U;T)$ for LDU inputs with different L. We observe that, at low SNR, binary inputs ($L = 2$) outperform inputs with larger L. As SNR increases beyond roughly 12.5 dB, inputs with $L = 3, 4, 5, \dots$ successively dominate. Although an LCU input is asymptotically capacity-achieving, for moderate SNR it is outperformed by LDU inputs. At low SNR, we also plot the channel capacity, achieved by optimized OOK [1]. We observe that the rate gain of optimized OOK over symmetric OOK can be significant at low SNR. Finally we note that the capacity lower bound obtained in [2] numerically approximates the upper envelope of all the curves in Fig. 3.

For LDU inputs, we can also examine the cutoff rates for different constellation sizes, as given by the following proposition.

Proposition 3.7: For the induced additive-noise channel (4) with L-ary LDU inputs, the cutoff rate is

$$R_0 = \log L - \log \left[1 + \frac{4}{L} \sum_{l=0}^{L-1} \sum_{m=0}^{l-1} \frac{e^{-(l+m)\Delta}}{e^{-2l\Delta} + e^{-2m\Delta}} \right]. \tag{29}$$

Proof: In Appendix B.

Fig. 4 shows plots of R_0 versus SNR for different L. We observe behavior similar to the plot of $I(U;T)$ versus SNR in Fig. 3. Here in terms of cutoff rates, the threshold SNR beyond which inputs with $L > 2$ outperform binary inputs is roughly 18.5 dB. For moderate SNR the curve for the $L \to \infty$ limit is again dominated by the curves corresponding to small L. At low SNR, we also plot the cutoff rate for optimized OOK. An interesting observation is that, in terms of cutoff rates, the gain of using asymmetric optimized OOK turns out to be rather limited, never exceeding 0.03 nats per channel use.

IV. Multiple-Input Multiple-Output (MIMO) Channels

In this section, we demonstrate how the induced additive-noise channel model can be extended to certain MIMO channels. Specifically, we assume that the channel is memoryless and the receive antennas are spatially uncorrelated. For an M-transmit N-receive antenna system, we adopt the channel model

$$\boldsymbol{X} = \boldsymbol{S} \cdot \boldsymbol{H} + \boldsymbol{Z} \tag{30}$$

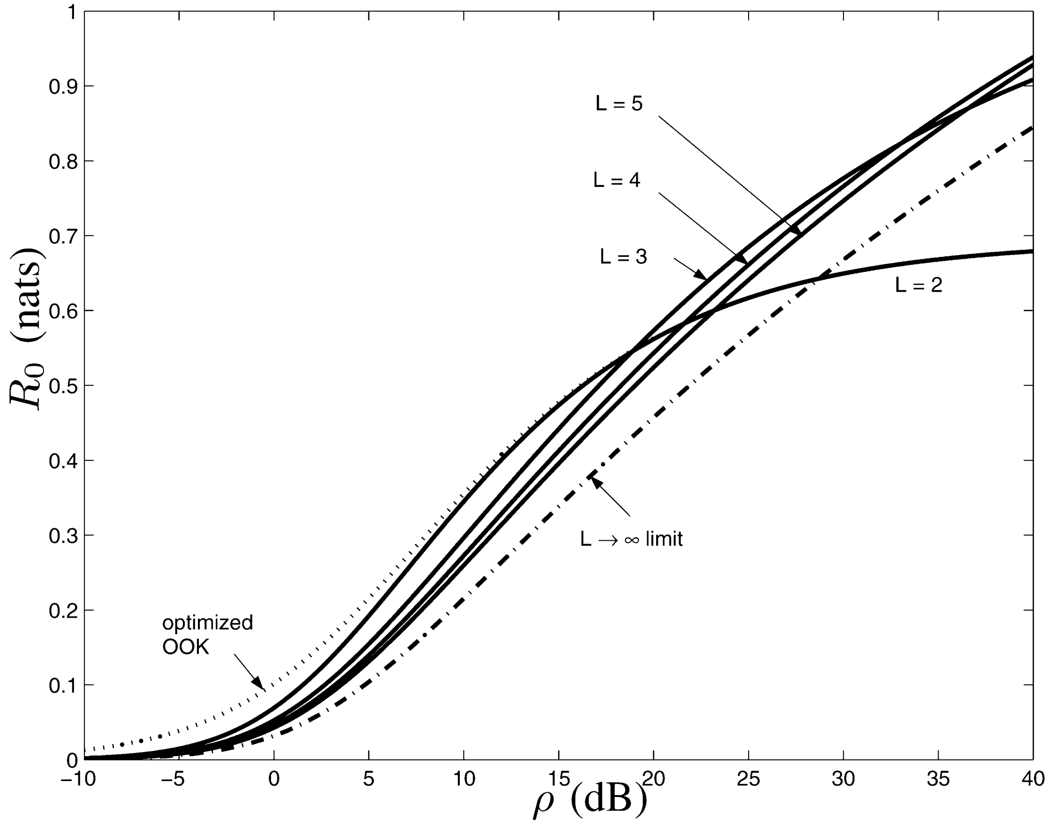

Fig. 4. Cutoff rates R_0 versus SNR for L-ary LDU ($L = 2, 3, 4, 5, \infty$) and optimized OOK inputs.

where

- $\boldsymbol{S} \in \mathbb{C}^{1\times M}$ is the channel input, with an average power constraint
$$\mathcal{E}[\boldsymbol{S}\boldsymbol{S}^\dagger] = P.$$
- $\boldsymbol{X} \in \mathbb{C}^{1\times N}$ is the channel output.
- $\boldsymbol{Z} \in \mathbb{C}^{1\times N}$ is the additive noise vector, which is zero-mean circular complex Gaussian with independent unit-variance elements, i.e., $\boldsymbol{Z} \sim \mathcal{CN}(0, \boldsymbol{I}_{N\times N})$.
- $\boldsymbol{H} \in \mathbb{C}^{M\times N}$ is the fading matrix, which we consider to be of the form
$$\boldsymbol{H} = \boldsymbol{\Phi}_{\mathrm{T}}^{1/2} \cdot \boldsymbol{H}_0$$
where $\boldsymbol{H}_0 \in \mathbb{C}^{M\times N}$ is a random matrix in which the elements are zero-mean unit-variance circular complex Gaussian random variables that are mutually independent. $\boldsymbol{\Phi}_{\mathrm{T}}$ is an $M \times M$ positive semi-definite matrix that characterizes the spatial correlation among the transmit antennas. We normalize the channel model so that the diagonal elements of $\boldsymbol{\Phi}_{\mathrm{T}}$ are all unity.

For a given channel input vector $\boldsymbol{S} = \boldsymbol{s}$, the conditional channel output vector $\boldsymbol{X}$ is circular complex Gaussian

$$\boldsymbol{X} \,|\, \boldsymbol{S} = \boldsymbol{s} \sim \mathcal{CN}\left(0, (1 + \boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger) \cdot \boldsymbol{I}_{N\times N}\right)$$

that is,

$$f_{\boldsymbol{X}|\boldsymbol{S}}(\boldsymbol{x}|\boldsymbol{s}) = \frac{1}{\pi^N \cdot (1 + \boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger)^N} \cdot \exp\left[\frac{\boldsymbol{x}\boldsymbol{x}^\dagger}{(1 + \boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger)}\right].$$

This implies that a scalar sufficient statistic for the channel is $R = \boldsymbol{X}\boldsymbol{X}^\dagger = \sum_{n=1}^N X_n X_n^\dagger$. Conditioned upon a given input $\boldsymbol{S} = \boldsymbol{s}$, R is the sum of N i.i.d. exponentially distributed random variables, i.e., central chi-square distribution with $2N$ degrees of freedom [13]

$$f_{R|\boldsymbol{S}}(r|\boldsymbol{s}) = \begin{cases} \frac{r^{N-1}}{(1+\boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger)^N (N-1)!} \cdot \exp\left[-\frac{r}{1+\boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger}\right], & \text{if } r \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Parallel to the scalar channel case, let us take the logarithm of $\sqrt{R}$. Then we obtain an equivalent sufficient statistic $T = \frac{1}{2}\log R$, for which the conditional pdf is

$$f_{T|\boldsymbol{S}}(t|\boldsymbol{s}) = \frac{2}{(N-1)!} \exp\Big\{2Nt - \log\left(\boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger + 1\right) N - \exp\left[2t - \log\left(\boldsymbol{s}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{s}^\dagger + 1\right)\right]\Big\}.$$

As before, this logarithmic transformation converts the fading channel (30) into an induced additive-noise channel

$$T = U + W \tag{31}$$

where we have the following.

- The channel input is
$$U = \frac{1}{2}\log\left(\boldsymbol{S}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{S}^\dagger + 1\right). \tag{32}$$
From the average power constraint of the fading channel (30), U should satisfy
$$\mathcal{E}[\exp(2U)] = \mathcal{E}[\boldsymbol{S}\boldsymbol{\Phi}_{\mathrm{T}}\boldsymbol{S}^\dagger] + 1. \tag{33}$$
For any given input random variable U, amplifying it by a constant factor greater than one does not decrease the channel mutual information, hence the channel capacity of (31) is achieved when the right-hand side of (33) is maximized. This occurs when $\boldsymbol{S}^\dagger$ is the eigenvector of $\boldsymbol{\Phi}_{\mathrm{T}}$ corresponding to the maximum

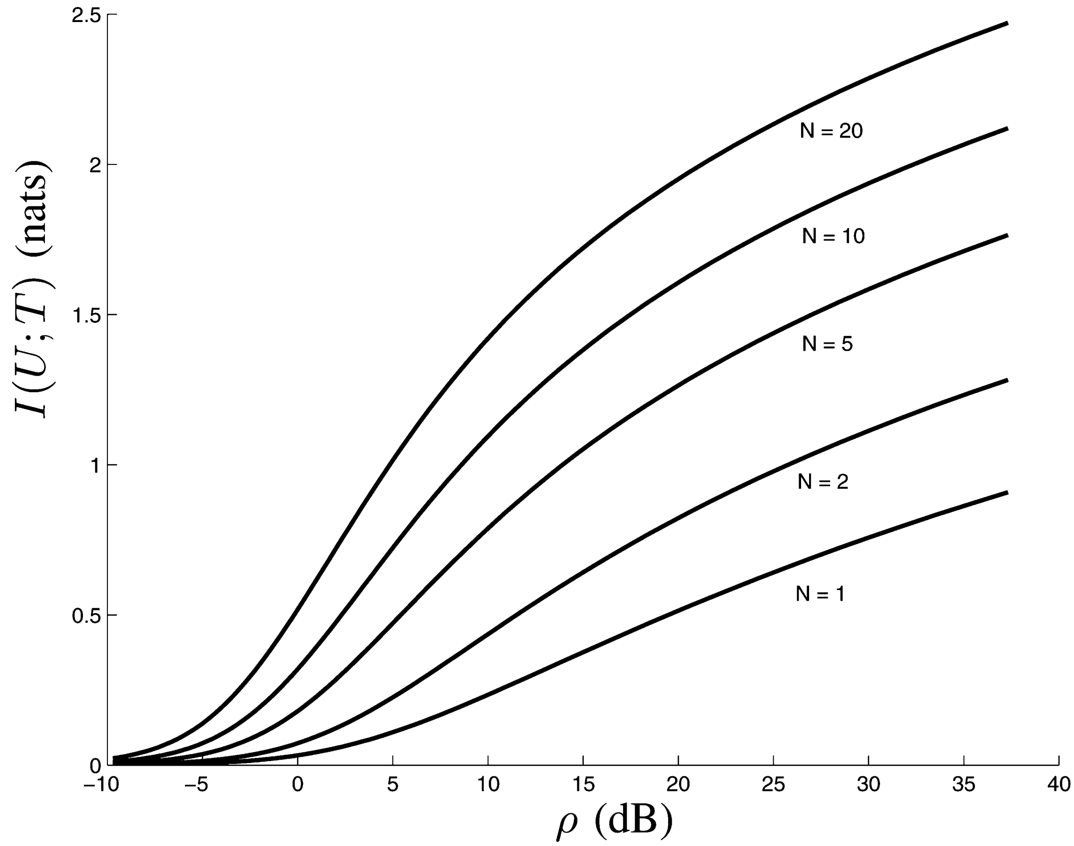

Fig. 5. Lower bounds of $I(U;T)$ (based on the entropy power inequality) under LCU inputs versus SNR, for different N.

eigenvalue $\lambda_{\max}(\mathbf{\Phi}_{\rm T})$, which can be accomplished by beamforming at the transmitter. Thus, in the sequel, we take the input constraint as

$$\mathcal{E}[\exp(2U)] = \lambda_{\max}(\mathbf{\Phi}_{\rm T})P + 1 = \rho + 1 \tag{34}$$

for the support set $U \in [0, \infty)$.

- The channel output is

$$T = \frac{1}{2}\log \boldsymbol{X}\boldsymbol{X}^{\dagger}. \tag{35}$$

- The additive noise W is independent of the channel input U, and has pdf

$$f_W(w) = \frac{2}{(N-1)!}\exp[2Nw - \exp(2w)] \tag{36}$$

for $w \in (-\infty, \infty)$.

Some basic properties of W are summarized as follows.

Lemma 4.1: (Properties of W, the MIMO case) The additive noise W in (36) has the following properties.

1) The cdf of W is

$$F_W(w) = 1 - \exp(-\exp(2w))\sum_{n=0}^{N-1}\frac{\exp(2nw)}{n!}. \tag{37}$$

2) The characteristic function of W is

$$\varphi_W(j\theta) = \frac{1}{(N-1)!}\Gamma(N + j\theta/2). \tag{38}$$

3) The mean of W is

$$\mathcal{E}[W] = \frac{\psi(N)}{2} \tag{39}$$

where $\psi(\cdot)$ is Euler's psi function [9]. Here for integer-valued N

$$\psi(N) = -\gamma + \sum_{n=1}^{N-1}\frac{1}{n} \sim \log N, \qquad \text{as } N \to \infty.$$

On the other hand, the maximum of $f_W(w)$ is attained at $w = \log N/2$.

4) The variance of W is

$$\sigma_W^2 = \frac{\zeta(2,N)}{4} \tag{40}$$

where $\zeta(\cdot,\cdot)$ is Riemann's Zeta function [9]. Here for integer-valued N

$$\zeta(2,N) = \sum_{n=0}^{\infty}\frac{1}{(n+N)^2}.$$

5) The differential entropy of W is

$$h(W) = N\,(1 - \psi(N)) - \log 2 + \sum_{n=1}^{N-1}\log n. \tag{41}$$

Proof: In Appendix C.

For the induced additive-noise channel (31), essentially all of the developments in Section III can be paralleled. The necessary and sufficient conditions for achieving unbounded channel mutual information (Proposition 3.1) and for asymptotically achieving channel capacity (Proposition 3.3) directly apply without change. We can also study performance of LCU and LDU inputs. By noting that the fading number is $\chi = N(\psi(N) - 1) - \sum_{n=1}^{N-1}\log n$ [3], we find that the LCU distribution is again asymptotically capacity-achieving.

Proposition 4.2: For the induced additive-noise channel (31), an LCU input distribution achieves

$$\lim_{\rho\to\infty}(C - I(U;T)) = 0. \tag{42}$$

In Fig. 5 we plot the lower bound of $I(U;T)$ (based on the entropy power inequality) for LCU inputs versus SNR, for different N. Increasing the number of receive antennas leads to quite noticeable rate gain, since the double-logarithmic capacity growth in SNR is rather limited for moderate SNR. By contrast, in coherent channels where the channel capacity grows logarithmically with SNR, the relative rate gain of merely increasing the number of receive antennas is not that significant.

V. Conclusion

In this correspondence we address noncoherent discrete-time memoryless Rayleigh-fading channels. By taking the logarithm of the channel output's magnitude, we transform the fading channel into an equivalent induced channel with additive noise that is independent of the channel input. This additive-noise channel model holds for all SNR. Using this perspective, we revisit several known results and establish several new results. In particular, we examine the moderate and high SNR channel behavior. We obtain simple conditions for channel input distributions to achieve unbounded mutual information as SNR scales, and to asymptotically achieve the channel capacity in the high SNR limit. We show that a simple input distribution called the log-scale (continuous) uniform distribution is asymptotically capacity-achieving. We further reinterpret a class of log-scale discrete uniform input distributions, and demonstrate that for moderate SNR they perform better than the log-scale (continuous) uniform input. Finally, we extend the induced additive-noise channel approach to memoryless multiple-antenna channels possibly with transmit, but without receive, spatial correlation.

A generalization of the log-scale transform approach, if it exists, doubtlessly would enhance our understanding of channel behavior and communication system design for general noncoherent fading channels. However, we note that currently this approach seems to be applicable only to memoryless Rayleigh fading channels; extensions to more general scenarios are not immediate. There are primarily two obstacles. First, for other fading distributions, the "non-Rayleighness" precludes the convenient shift-invariance property of the log-scale transform. For example, in Ricean channels or phase noncoherent Gaussian channels, the induced additive noise after the log-scale transform is input-dependent. More fundamentally, for channels in which the output sufficient statistic is multivariate rather than scalar, we have not yet found an immediate analog of the log-scale transform.

Appendix A
A Capacity Upper Bound Revisited

In this appendix, we obtain a capacity upper bound from [2] using the additive-noise perspective.

In order to apply standard maximum entropy techniques [10], a natural way is to replace the constraint $T = U + W$ by taking moments at both sides. This type of relaxation will yield upper bounds to the channel capacity. In this section let us consider taking their expectations, as

$$\mathcal{E}[T] = \mathcal{E}[U] + \mathcal{E}[W] = \lambda \geq -\frac{\gamma}{2}.$$

Here the last inequality is from $U \in [0, \infty)$. On the other hand, the channel output should have a constrained average power as given by

$$\mathcal{E}[\exp(2T)] = \mathcal{E}[|X|^2] = \rho + 1.$$

Thus following maximum entropy techniques, the resulting entropy-maximizing distribution of T should have pdf

$$f_T^*(t) = 2\left(\frac{\mu}{\rho+1}\right)^{\mu} \frac{1}{\Gamma(\mu)} \exp\left[2\mu t - \frac{\mu}{\rho+1}\exp(2t)\right]$$

where $\mu > 0$ is determined by the equation

$$\psi(\mu) - \log\mu = 2\lambda - \log(\rho+1)$$

and $\psi(\mu) = \frac{d}{d\mu}\log\Gamma(\mu)$ is the psi function [9]. Thus, we can treat $f_T^*(t)$ as the actual output pdf to obtain an upper bound to the channel capacity as

$$\begin{aligned} C &= \sup_{f_T(t)} I(U;T) < h(f_T^*(t)) - h(W) \\ &= \log\Gamma(\mu) - \mu\psi(\mu) + \mu - \gamma - 1 \end{aligned} \tag{43}$$

where

$$\log\mu - \psi(\mu) = \log(\rho+1) - 2\lambda. \tag{44}$$

To tighten the upper bound, we first notice that the right-hand side of (43) is increasing with $\mu > 0$, hence the tightest upper bound is obtained when μ is minimized. We then notice that the left-hand side of (44) is decreasing with $\mu > 0$, so that the minimum allowed μ is attained when $\lambda = -\gamma/2$. This capacity upper bound is exactly the same as that obtained in [2] using variational methods. Here we revisit it from the induced additive-noise channel model. We also note that by introducing more moment constraints, it is possible to obtain tighter capacity upper bounds. However, their derivation and evaluation will become much less tractable.

As a side note, a tighter capacity upper bound is obtained in [3], based upon a dual expression of channel capacity. Such result can also be reproduced by optimizing the output pdf of the induced additive-noise channel, but this approach does not lead to simplification compared to the original one.

Appendix B
Proof of Proposition 3.7

By the definition of cutoff rate (e.g., [13]) we have

$$R_0 = -\log\int_{-\infty}^{\infty}\left[\sum_{l=0}^{L-1}\frac{1}{L}\sqrt{f_W(t-l\Delta)}\right]^2 dt.$$

The inner integral can be evaluated as

$$\begin{aligned} &\int_{-\infty}^{\infty}\left[\sum_{l=0}^{L-1}\frac{1}{L}\sqrt{f_W(t-l\Delta)}\right]^2 dt \\ &= \int_{-\infty}^{\infty}\frac{1}{L^2}\left\{\sum_{l=0}^{L-1} f_W(t-l\Delta)\right. \\ &\quad \left. +2\sum_{l=0}^{L-1}\sum_{m=0}^{l-1}\sqrt{f_W(t-l\Delta)f_W(t-m\Delta)}\right\} dt \\ &= \frac{1}{L^2}\left\{L + 2\sum_{l=0}^{L-1}\sum_{m=0}^{l-1}\int_{-\infty}^{\infty} 2\exp\left[2t-(l+m)\Delta\right.\right. \\ &\quad \left.\left. -\frac{1}{2}\exp(2t)(\exp(-2l\Delta)+\exp(-2m\Delta))\right] dt\right\} \\ &= \frac{1}{L}\left\{1 + \frac{8}{L}\sum_{l=0}^{L-1}\sum_{m=0}^{l-1}\int_0^{\infty} x\exp[-(l+m)\Delta\right. \\ &\quad \left. -x(\exp(-2l\Delta)+\exp(-2m\Delta))]\frac{1}{2x}dx\right\} \\ &= \frac{1}{L}\left\{1 + \frac{4}{L}\sum_{l=0}^{L-1}\sum_{m=0}^{l-1}\frac{\exp(-(l+m)\Delta)}{\exp(-2l\Delta)+\exp(-2m\Delta)}\right\}. \end{aligned}$$

The cutoff rate R_0 then follows immediately. Q.E.D.

APPENDIX C
PROOF OF LEMMA 4.1

1) cdf:

$$F_W(w) = \int_{-\infty}^{w} \frac{2}{(N-1)!} e^{2Nw_1} e^{-e^{2w_1}} dw_1$$
$$= \frac{2}{(N-1)!} \int_0^{e^{2w}} w_2^N e^{-w_2} \frac{1}{2w_2} dw_2$$
$$= \frac{1}{(N-1)!} \left[(N-1)! - e^{-e^{2w}} \sum_{n=0}^{N-1} \frac{(N-1)!}{n!} e^{2nw} \right]$$
$$= 1 - e^{-e^{2w}} \sum_{n=0}^{N-1} \frac{e^{2nw}}{n!}$$

where we use

$$\int_0^u x^n e^{-\mu x} dx = \frac{n!}{\mu^{n+1}} - e^{-u\mu} \sum_{k=0}^{n} \frac{n!}{k!} \frac{u^k}{\mu^{n-k+1}}$$

for $u > 0$, $\Re\mu > 0$ [9, Sec. 3.381].

2) Characteristic function:

$$\varphi_W(j\theta) = \int_{-\infty}^{\infty} \exp(j\theta w) \frac{2}{(N-1)!} e^{2Nw} e^{-e^{2w}} dw$$
$$= \frac{1}{(N-1)!} \int_0^{\infty} w_1^{N-1+j\theta/2} \exp(-w_1) dw_1$$
$$= \frac{1}{(N-1)!} \Gamma(N + j\frac{\theta}{2})$$

where we use $\int_0^{\infty} x^{\nu-1} \exp(-\mu x) dx = \frac{1}{\mu^\nu}\Gamma(\nu)$ for $\Re\mu > 0$, $\Re\nu > 0$ [9, Sec. 3.381].

3) Mean:

$$\mathcal{E}[W] = (-j) \left. \frac{d\varphi_W(j\theta)}{d\theta} \right|_{\theta=0}$$
$$= \frac{1}{2(N-1)!} \int_0^{\infty} w^{N-1} \exp(-w) \log w dw$$
$$= \frac{1}{2(N-1)!} \Gamma(N)\psi(N) = \frac{\psi(N)}{2}$$

where we use

$$\int_0^{\infty} x^{\nu-1} e^{-\mu x} \log x dx = \frac{1}{\mu^\nu} \Gamma(\nu)[\psi(\nu) - \log \mu]$$

for $\Re\mu > 0$, $\Re\nu > 0$ [9, 4.352].

The maximization of $f_W(w)$ at $w = \log N/2$ is easily verified by taking derivative.

4) Variance:

$$\sigma^2{}_W$$
$$= \mathcal{E}[W^2] - \mathcal{E}[W]^2\}$$
$$= - \left. \frac{d^2\varphi_W(j\theta)}{d\theta^2} \right|_{\theta=0} - \frac{\psi(N)^2}{4}$$
$$= \frac{1}{4(N-1)^2} \Gamma(N) \left[\psi(N)^2 + \zeta(2, N) \right] - \frac{\psi(N)^2}{4}$$
$$= \frac{\zeta(2, N)}{4}$$

where we use

$$\int_0^{\infty} x^{\nu-1} e^{-\mu x} (\log x)^2 dx = \frac{\Gamma(\nu)}{\mu^\nu} \{[\psi(\nu) - \log \mu]^2 + \zeta(2, \nu)\}$$

for $\Re\mu > 0$, $\Re\nu > 0$ [9, Sec. 4.358].

5) Differential entropy:

$$h(W) = - \int_{-\infty}^{\infty} \frac{2}{(N-1)!} \exp(2Nw - \exp(2w))$$
$$\left[\log \frac{2}{(N-1)!} + 2Nw - \exp(2w) \right] dw$$
$$= \sum_{n=1}^{N-1} \log n - \log 2 - N \cdot \psi(N)$$
$$+ \frac{1}{(N-1)!} \int_0^{\infty} w^N \exp(-w) dw$$
$$= \sum_{n=1}^{N-1} \log n - \log 2 + N[1 - \psi(N)]. \qquad \text{Q.E.D.}$$

ACKNOWLEDGMENT

The authors wish to thank Amos Lapidoth and the anonymous reviewers for their valuable comments on this work, and Shlomo Shamai for pointing out the additive noise perspective in [1].

REFERENCES

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[2] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, Jul. 1997.

[3] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2567, Oct. 2003.

[4] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[5] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high SNR Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.

[6] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[7] N. T. Gaarder, "Signal design for fast-fading Gaussian channels," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 3, pp. 247–256, May 1971.

[8] J. F. Pieper, J. G. Proakis, R. R. Reed, and J. K. Wolf, "Design of efficient coding and modulation for a Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 457–468, Jul. 1978.

[9] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 5th ed. New York: Academic, 1994.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[11] V. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 809–827, Apr. 2002.

[12] M. J. Borran, A. Sabharwal, and B. Aazhang, "On design criteria and construction of noncoherent space-time constellations," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2332–2351, Oct. 2003.

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