On the Distribution of Mutual Information

J. Nicholas Laneman
Dept. of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
Email: jnl@nd.edu

***Abstract*— In the early years of information theory, mutual information was defined as a random variable, and error probability bounds for communication systems were obtained in terms of its probability distribution. We advocate a return to this perspective for a renewed look at information theory for general channel models and finite coding blocklengths. For capacity-achieving inputs, we characterize the mutual information random variables for several important channel models, including the discrete memoryless binary symmetric channel (BSC), the discrete-time complex additive white Gaussian noise (AWGN) channel, and the discrete-time Rayleigh fading channel with static, flat fading known perfectly to the decoder. Combined with a result known as Feinstein's lemma and its extensions, these characterizations can be employed to obtain bounds on the maximal block error probability of channel codes operating over these channels. Such bounds appear to be particularly useful when operating at transmission rates near the average mutual information.**

I. Introduction

It may be a common view that information theory characterizes fundamental performance limits, *e.g.*, channel capacity, for communication systems only in the regime of infinite coding delay and complexity. Wide classes of channels have been analyzed using approaches based mainly upon probabilistic limit theorems, such as laws of large numbers, ergodic theory, or large deviations [1], [2]. In establishing these results, a limit is often taken as the coding blocklength approaches infinity. This "infinite-blocklength" perspective and its main message, that longer codewords offer better performance, has taken root in the context of many applications and has stimulated development of numerous channel coding schemes of varying performance and complexity. In recent years, the asymptotic performance limits set forth by information theory have essentially been achieved in certain scenarios—with practical complexity no less—through the development of turbo codes and low-density parity-check (LDPC) codes with long blocklengths.

Despite its magnificent success for certain applications, there are applications, particularly in delay-constrained and many network scenarios, in which infinite-blocklength information theory has offered fewer insights, or otherwise the available insights have not been fully integrated into existing systems. This is especially true if delay, fairness, and other quality-of-service goals must be satisfied by the system. We believe that one of the main reasons for this current state of affairs is the lack of a "finite-blocklength" information theory that is general enough to handle a wide class of important channel models but not so complex and subtle so as to be unusable by system designers.

If coding blocklength is restricted so that traditional limit theorems do not apply, then one might infer that the real quantity of interest is not converging to the average mutual information. Both old and new literature suggest that the real quantity of interest is a mutual information random variable, instead of its expectation. Interestingly, this perspective appeared in the early days of information theory [2], [3], [4], [5], [6], [7], [8], receded for some time, and has returned in the context of information spectrum methods [9], [10], [11] and outage probability [12].

Motivated by this perspective, Section II reviews classic terminology and establishes modern notation for mutual information random variables. Section III recalls a generalization of Feinstein's lemma for bounding block error probability in terms of the distribution of mutual information. Section IV characterizes the mutual information for some important channel models, including the discrete memoryless binary symmetric channel (BSC), the discrete-time complex additive white Gaussian noise (AWGN) channel, and the Rayleigh flat-fading channel with static, flat fading known perfectly to the decoder.

II. Mutual Information as a Random Variable

To fix notation for the sequel, consider two random variables[1] x and y on alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, with joint probability law $p_{\mathsf{x},\mathsf{y}}(x,y)$, marginal probability laws $p_{\mathsf{x}}(x)$ and $p_{\mathsf{y}}(y)$, and conditional probability laws $p_{\mathsf{y}|\mathsf{x}}(y|x)$ and $p_{\mathsf{x}|\mathsf{y}}(x|y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. If the alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite or countably infinite, then the probability laws correspond to the appropriate probability mass functions; if $\mathcal{X}$ and $\mathcal{Y}$ are uncountable, the probability laws correspond to the appropriate probability density functions, which we assume exist. However, we note that even more general situations can be considered; see, *e.g.*, [7].

The *mutual information* between x and y is the random variable[2]

$$i(\mathsf{x};\mathsf{y}) := \log \frac{p_{\mathsf{x},\mathsf{y}}(\mathsf{x},\mathsf{y})}{p_{\mathsf{x}}(\mathsf{x})p_{\mathsf{y}}(\mathsf{y})} \ . \tag{1}$$

[1]Random variables are denoted by sans serif font, *e.g.*, x, and sample values are denoted by serif font, *e.g.*, x. Random and sample column vectors of length n are denoted by $\mathsf{x}^n = [\mathsf{x}_1\ \mathsf{x}_2\ \ldots\ \mathsf{x}_n]^\top$ and $x^n = [x_1\ x_2\ \ldots\ x_n]^\top$, respectively.

[2]Unless indicated otherwise, all logarithms in the paper are taken to base e.

This random variable has gone by other names, such as "(mutual) information density" [7], [9], [10], [11]. Because we will focus on the probability density and/or probability distribution of this random variable, our terminology follows that of [2], [4], [6], and refers to the expectation $\mathrm{E}\left[i(x;y)\right]$ as the *average mutual information*. The *distribution of mutual information* $\Pr\left[i(x;y) \leq i\right]$ has also been called the "(mutual) information spectrum" [9], [10], [11].

For three random variables x, y, and z on alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively, with suitable joint, marginal, and conditional probability laws, the *conditional mutual information* between x and y given z is the random variable

$$i(x;y|z) := \log \frac{p_{x,y|z}(x,y|z)}{p_{x|z}(x|z)p_{y|z}(y|z)} \ . \tag{2}$$

Important relations for mutual informations and conditional mutual informations result from basic properties of joint, marginal, and conditional probability laws. As specific examples, with probability one we have [7]

$$\begin{aligned} i(x;y) &= \log \frac{p_{y|x}(y|x)}{p_y(y)} = \log \frac{p_{x|y}(x|y)}{p_x(x)} \ , \\ i(x;y|z) &= \log \frac{p_{y|x,z}(y|x,z)}{p_{y|z}(y|z)} = \log \frac{p_{x|y,z}(x|y,z)}{p_{x|z}(x|z)} \ , \\ i(x,z;y) &= i(z;y) + i(x;y|z) \ . \end{aligned}$$

Also, x and y are independent if and only if $i(x;y) = 0$ with probability one; similarly, x and y are conditionally independent given z if and only if $i(x;y|z) = 0$ with probability one.

Finally, for any mutual information or conditional mutual information, there can be several expectations or conditional expectations of interest, *e.g.*, $\mathrm{E}\left[i(x;y|z)\right]$, $\mathrm{E}\left[\, i(x;y|z)|\, z\right]$, and so forth.

III. Feinstein's Lemma for Communication Channels with Input Constraints

This section recalls a bound for the maximal block error probability for channel codes operating over a communications channel with input constraints. We begin with some standard definitions and then state the result without proof. More details can be found in [8], [13]

A point-to-point *communication system* consists of a channel, inputs to the channel, and outputs from the channel. Let $\mathcal{X}$ and $\mathcal{Y}$ denote the channel input and output alphabets, respectively. For an integer $n > 0$, the channel inputs and outputs are modeled as random vectors x^n and y^n with joint probability law $p_{x^n,y^n}(x^n,y^n)$ and marginal probability laws $p_{x^n}(x^n)$ and $p_{y^n}(y^n)$, respectively, for $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$. The channel is modeled as a conditional probability law $p_{y^n|x^n}(y^n|x^n)$, for $x \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$. We emphasize that these probability laws need not have any structure, such as independent inputs, *i.e.*, $p_{x^n}(x^n) = \prod_{l=1}^n p_{x_l}(x_l)$, or a memoryless channel, *i.e.*, $p_{y^n|x^n}(y^n|x^n) = \prod_{l=1}^n p_{y_l|x_l}(y_l|x_l)$.

Often an application requires that the inputs be constrained. We let $\mathcal{S}_n \subseteq \mathcal{X}^n$ denote the *input constraint set* for such scenarios. As one important example, if the inputs are complex-valued, *i.e.*, $\mathcal{X} = \mathbb{C}$, and constrained to have average power less than P, then

$$\mathcal{S}_n = \left\{ x^n : \sum_{l=1}^n |x_l|^2 \leq nP \right\} \ . \tag{3}$$

To state results precisely, we must define the notion of a *channel code* for the communication system. For any integers $n > 0$ and $m_n > 0$, an $(n, m_n, \mathcal{S}_n, \epsilon_n)$-*code* for the channel $p_{y^n|x^n}(y^n|x^n)$ consists of a message set $\mathcal{M}_n := \{1, 2, \ldots, m_n\}$, a mapping $f_n : \mathcal{M}_n \to \mathcal{X}^n$ called the *encoder*, and a mapping $g_n : \mathcal{Y}^n \to \mathcal{M}_n$ called the *decoder* such that $f_n(l) \in \mathcal{S}_n$ and the conditional probability of error $\Pr[g_n(y^n) \neq l | x^n = f_n(l)] \leq \epsilon_n$ for all $l \in \mathcal{M}_n$.

The following result is a generalization of what is commonly called *Feinstein's Lemma* [3], [5], [8], [10], [11] to incorporate input constraints.

Lemma 1 (Feinstein's Lemma with Input Constraints [13]): Given arbitrary integers $n > 0$ and $m_n > 0$, an input probability law $p_{x^n}(x^n)$, an input constraint set $\mathcal{S}_n$, and a channel probability law $p_{y^n|x^n}(y^n|x^n)$, there exists an $(n, m_n, \mathcal{S}_n, \epsilon_n)$-code with

$$\begin{aligned} \epsilon_n \leq & \Pr\left[\frac{1}{n} i(x^n;y^n) \leq \frac{1}{n}\log m_n + \gamma\right] \\ & + \Pr\left[x^n \notin \mathcal{S}_n\right] + e^{-n\gamma} \ , \end{aligned} \tag{4}$$

where $\gamma > 0$ is an arbitrary constant.

Lemma 1 provides a direct coding theorem for quite general channels, inputs, and constraints, and places the distribution of the mutual information rate $i(x^n;y^n)/n$ in a central role. Originally proven in [3], [5] for discrete channels and inputs, it was extended to continuous channels possessing densities and input constraints in [13], and can be extended to arbitrary measures [14]. Also of note, a result strikingly similar to Lemma 1 without input constraints, *i.e.*, $\mathcal{S}_n = \mathcal{X}^n$, was obtained using random coding arguments in [4] for average error probability of uniform messages over discrete memoryless channels. However, we stress that Lemma 1 bounds the maximum probability of error, not the average probability of error.

Infinite-blocklength information theory relies on the limiting average mutual information rate $\lim_{n\to\infty} \mathrm{E}\left[i(x^n;y^n)\right]/n$ to characterize the capacity of discrete memoryless as well as stationary and ergodic channels [1], [2]. Recent results for much more general channels appear in [10], [11], but the average mutual information is not sufficient for that purpose. Instead, using Feinstein's Lemma as a direct proof, along with a similar lower bound, the quantity $\sup\{t \in \mathbb{R} : \lim_{n\to\infty} \Pr[i(x^n;y^n)/n < t] = 0\}$, called the *liminf in probability* of $i(x^n;y^n)/n$, is needed. Because we will not take the limit as the blocklength goes to infinity, neither of these quantities will be of particular use to us, except as reference points.

To motivate the results to follow, consider the case of memoryless channels with independent and identically distributed (i.i.d.) inputs. In this case, the mutual information rate becomes

$$\frac{1}{n} i(\mathsf{x}^n; \mathsf{y}^n) = \frac{1}{n} \sum_{l=1}^{n} i(\mathsf{x}_l; \mathsf{y}_l) \ . \tag{5}$$

Thus, computing the distribution of the mutual information rate corresponds to finding probabilities for sums of i.i.d. random variables. Beginning with [4], a common tool in classical information theory has been to apply the Chernoff bound, or large deviations arguments, to obtain upper bounds on the distribution of mutual information that decay exponentially in the coding blocklength n. However, as is well known, these bounds are not always sharp for estimates of the distribution near the mean, or when n is limited. As a result, we focus in the sequel on *exactly* characterizing the mutual information rate corresponding to capacity-achieving inputs.

IV. Example Channels

In this section, we determine the density and/or distribution of mutual information for some important special cases, including memoryless binary symmetric channel (BSC), the discrete-time circular complex additive white Gaussian noise (AWGN) channel, and static fading channels with Rayleigh fading and channel state information (CSI) at the decoder. In addition to serving as simulation tools, these results can be used via (4) to obtain bounds on the maximal error probability for finite coding blocklengths. For the BSC in particular, we compare (4) to bounds based upon the random coding error exponent [2].

A. Binary Symmetric Channel

In this section, we consider the discrete memoryless binary symmetric channel (BSC). For this channel, the input and output alphabets are $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and the inputs are unconstrained, *i.e.,* $\mathcal{S}_n = \mathcal{X}^n$. The channel probability law is memoryless with

$$p_{\mathsf{y}_l|\mathsf{x}_l}(y_l|x_l) = \begin{cases} p \ , & y_l \neq x_l \\ (1-p) \ , & y_l = x_l \end{cases} \ , \tag{6}$$

for $l = 1, 2, \ldots, n$, where $0 < p < 1$ is the channel bit-error probability.

For i.i.d. inputs with $p_{\mathsf{x}_l}(x_l) = 1/2$, $x_l \in \mathcal{X}$, we have $p_{\mathsf{y}^n}(y^n) = 1/2^n$, for all $y^n \in \mathcal{Y}^n$. Letting k_n be a binomial random variable corresponding to the number of indices for which $\mathsf{y}_l \neq \mathsf{x}_l$, the mutual information rate can be manipulated into the form

$$\frac{1}{n} i(\mathsf{x}^n; \mathsf{y}^n) = \underbrace{(\log 2 - H(p))}_{C_{\mathrm{BSC}}(p)} + \left(\left(\frac{\mathsf{k}_n}{n} - p \right) \log \frac{p}{(1-p)} \right) \ , \tag{7}$$

where $H(p) := -p \log p - (1-p) \log(1-p)$ is the binary entropy function, and $C_{\mathrm{BSC}}(p)$ is the BSC channel capacity.

It is clear that the second term in (7) has mean zero and is a linear function of the binomial random variable k_n defined above. For $0 \leq p < 1/2$, so that $\log(p/(1-p)) < 0$, the distribution of the mutual information rate satisfies

$$\begin{aligned} &\Pr\left[\frac{1}{n} i(\mathsf{x}^n; \mathsf{y}^n) \leq \frac{1}{n} \log \lceil e^{nR} \rceil + \gamma \right] \\ &= \Pr\left[\mathsf{k}_n \geq \alpha(\gamma) \right] = \Pr\left[\mathsf{k}_n \geq \lceil \alpha(\gamma) \rceil \right] \\ &= \sum_{\lceil \alpha(\gamma) \rceil}^{n} \binom{n}{l} p^l (1-p)^{(n-l)} \\ &= \frac{B_p(\lceil \alpha(\gamma) \rceil, n + 1 - \lceil \alpha(\gamma) \rceil)}{B_1(\lceil \alpha(\gamma) \rceil, n + 1 - \lceil \alpha(\gamma) \rceil)} \end{aligned} \tag{8}$$

where $B_x(a, b) := \int_0^x t^{a-1}(1-t)^{b-1} dt$ is the incomplete Beta function [15], and where

$$\alpha(\gamma) := n \left(\frac{\frac{1}{n} \log \lceil e^{nR} \rceil - \log 2 - \log(1-p) + \gamma}{\log \frac{p}{(1-p)}} \right) \ . \tag{9}$$

Fig. 1 shows the results of (8) in (4) for a memoryless BSC with bit-error probability $p = 10^{-3}$ for various blocklengths n and transmission rates R, so that the code size is $m_n = \lceil e^{nR} \rceil$. The inputs are unconstrained, so that the second term in (4) is zero, and minimization over $\gamma > 0$ is performed numerically.

Fig. 2 shows comparisons of bounds based upon Feinstein's lemma (4) and random coding error exponents [2] for the BSC with i.i.d. uniform inputs. We observe that the bound (4) can be tighter than bounds based upon the random coding error exponent for rates close to capacity and blocklengths up to several hundred bits. For large blocklengths, the bounds based upon error exponents have faster decay with blocklength, as we would expect, and eventually outperform the bound (4). We stress that coding blocklengths on the order of 200 bits are frequently used, *e.g.,* in cellular systems and sensor networks. Fortunately, the closer we operate to the average mutual information, the larger the error probability, and hence computation of (4) is less numerically sensitive than for smaller transmission rates.

B. Gaussian Channels

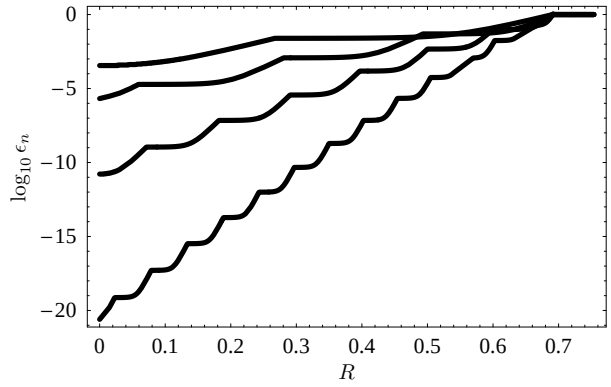
In this section, we consider the discrete-time, scalar complex additive white Gaussian noise (AWGN) channel modeled as

$$\mathsf{y}_l = a\mathsf{x}_l + \mathsf{z}_l \ , \quad l = 1, 2, \ldots, n \ , \tag{10}$$

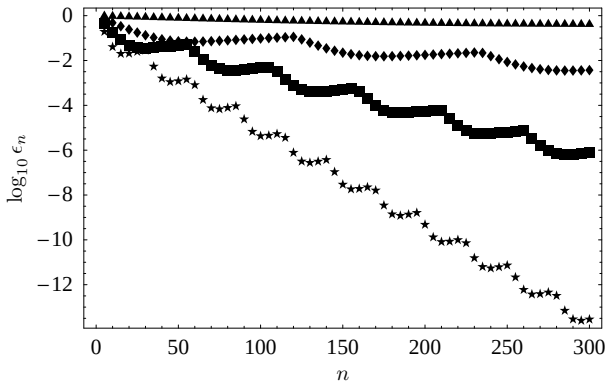
where a is a complex-valued constant known to the transmitter and receiver, and the additive noise z_l is i.i.d. circular complex Gaussian with zero mean and variance N_0. This channel law is memoryless with

$$p_{\mathsf{y}_l|\mathsf{x}_l}(y_l|x_l) = \mathcal{N}(y_l; ax_l, N_0) := \frac{1}{\pi N_0} e^{-|y_l - ax_l|^2 / N_0} \ . \tag{11}$$

For i.i.d. circular complex Gaussian inputs with $p_{\mathsf{x}_l}(x_l) = \mathcal{N}(x_l; 0, \beta P)$, the output is also i.i.d. complex Gaussian with $p_{\mathsf{y}_l}(y_l) = \mathcal{N}(y_l; 0, |a|^2 \beta P + N_0)$. The case of $\beta = 1$ corresponds to the capacity-achieving inputs, but we allow for $\beta < 1$ so that the second term in (4) does not dominate the bound. After some manipulations [7], the mutual information



(a)



(b)

Fig. 1. Plots of the upper bound (4) for the BSC with bit-error probability $p = 10^{-3}$: (a) blocklengths $n = 25, 50, 100, 200$ and rates $0 \leq R \leq 11C_{\mathrm{BSC}}(10^{-3})/10$, (b) blocklengths $n = 10, 20, 40, \ldots, 300$ and rates $R = C_{\mathrm{BSC}}(10^{-3})/2$ (stars), $R = 3C_{\mathrm{BSC}}(10^{-3})/4$ (squares), $R = 9C_{\mathrm{BSC}}(10^{-3})/10$ (diamonds), and $R = 1001C_{\mathrm{BSC}}(10^{-1})/1000$ (triangles).

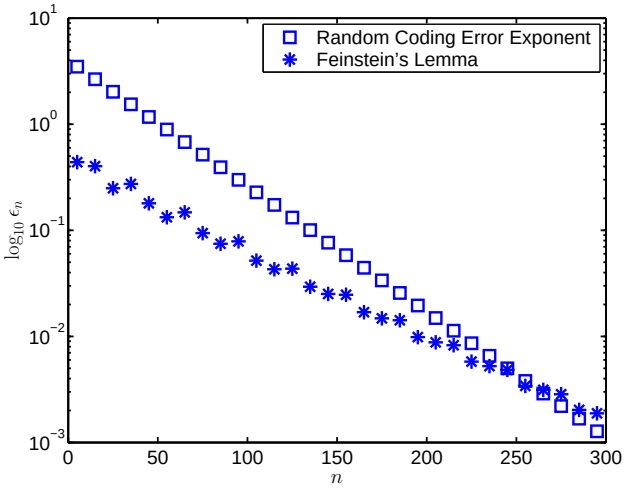


Fig. 2. Comparison of the bound (4) to a bound based upon the random coding error exponent for the BSC with $p = 0.031125$ and $R = 3C_{\mathrm{BSC}}(p)/4$.

rate $i(x^n; y^n)/n$ can be shown to have the same distribution as the random variable

$$\underbrace{\log\left(1+\beta\mathrm{SNR}\right)}_{C_{\mathrm{G}}(\beta\mathrm{SNR})} + \frac{1}{n}\sqrt{\frac{\beta\mathrm{SNR}}{\beta\mathrm{SNR}+1}}\sum_{l=1}^{n} w_l \ , \tag{12}$$

where $\mathrm{SNR} := |a|^2 P/N_0$ is the channel signal-to-noise ratio corresponding to the power constraint P, $C_{\mathrm{G}}(\mathrm{SNR})$ is the AWGN channel capacity in nats per complex channel use at signal-to-noise ratio SNR, and w_l, $l = 1, 2, \ldots, n$, are i.i.d. Laplace random variables with mean zero and variance two, *i.e.*, each is the difference of two independent exponential random variables with mean one.

The sum of n i.i.d. Laplace random variables each with mean zero and variance σ^2 has the Bessel-K distribution, for which the probability density is symmetric about zero and takes the form [16]

$$\mathcal{K}(t;\sigma,n) := \frac{2^{1-n}}{\sqrt{\pi}\Gamma(n)\sigma}\left(\frac{\sqrt{2}|t|}{\sigma}\right)^{n-\frac{1}{2}} K_{n-\frac{1}{2}}\left(\frac{\sqrt{2}|t|}{\sigma}\right) \ , \tag{13}$$

where $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}dt$ is the gamma function and $K_\nu(z)$ is the modified Bessel function of the second kind with index ν [15]. For integer n, $K_{n-1/2}(z)$ can be expressed as

$$K_{n-\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}}e^{-z}\sum_{l=0}^{n-1}\frac{\Gamma(n+l)}{\Gamma(n-l)\Gamma(l+1)}(2z)^{-l} \ . \tag{14}$$

To find the probability distribution corresponding to (13), we utilize (14), integrate each term in the sum, and exploit symmetry to obtain

$$\int_{-\infty}^{t}\mathcal{K}(\tau;\sigma,n)d\tau = \begin{cases} \frac{1}{2} + \sum_{l=0}^{n-1}\frac{\Gamma(n+l)\Gamma(n-l,\sqrt{2}t/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)}2^{-n-l} \ , & t \geq 0 \\ \sum_{l=0}^{n-1}\frac{\Gamma(n+l)\Gamma(n-l,-\sqrt{2}t/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)}2^{-n-l} \ , & t < 0 \end{cases} \ , \tag{15}$$

where $\Gamma(a,z) := \int_z^\infty t^{a-1}e^{-t}dt$ is the incomplete gamma function [15].

Taking into account the non-zero mean and the scale factor in front of the sum in (12), we see that

$$i(x^n;y^n)/n - \log(1+\beta\mathrm{SNR}) \sim \mathcal{K}\left(i;\sqrt{\frac{2}{n^2}\frac{\beta\mathrm{SNR}}{\beta\mathrm{SNR}+1}},n\right) \ . \tag{16}$$

C. Static Fading Channels

In this final section, we briefly consider discrete-time fading channels with flat, static fading know perfectly at the decoder. We focus on Rayleigh fading, but extensions to more general fading distributions can be readily developed. The channel is modeled as

$$y_l = a x_l + z_l \ , \quad l = 1, 2, \ldots, n \ , \tag{17}$$

where a is a $\mathcal{N}(a;0;\sigma_a^2)$ random variable known to the receiver but not the transmitter, and again the additive noise z_l is i.i.d.

circular complex Gaussian noise with mean zero and variance N_0. The fading coefficient a is independent of the additive noise z_l and the input x_l, for $l = 1, 2, \ldots, n$.

For i.i.d. circular complex Gaussian inputs with $p_{\mathsf{x}_l}(x_l) = \mathcal{N}(x_l; 0, \beta P)$, the output is also conditionally i.i.d. complex Gaussian with $p_{\mathsf{y}_l|\mathsf{a}}(y_l|a) = \mathcal{N}(y_l; 0, |a|^2\beta P + N_0)$. Since the fading is known to the decoder, the effective channel output is $(\mathsf{y}^n, \mathsf{a})$. Since a and x^n are independent, the mutual information is, with probability one,

$$\begin{aligned} i(\mathsf{x}^n; \mathsf{y}^n, \mathsf{a}) &= \log \frac{p_{\mathsf{y}^n,\mathsf{a}|\mathsf{x}^n}(\mathsf{y}^n, \mathsf{a}|\mathsf{x}^n)}{p_{\mathsf{y}^n,\mathsf{a}}(\mathsf{y}^n, \mathsf{a})} \\ &= \log \frac{p_{\mathsf{y}^n|\mathsf{x}^n,\mathsf{a}}(\mathsf{y}^n|\mathsf{x}^n, \mathsf{a})}{p_{\mathsf{y}^n|\mathsf{a}}(\mathsf{y}^n|\mathsf{a})} \\ &= i(\mathsf{x}^n; \mathsf{y}^n|\mathsf{a}) \ . \end{aligned} \tag{18}$$

We note that an expression of the form (18) appears in [17]; however, no calculations are provided for any specific fading channel models.

The mutual information rate $i(\mathsf{x}^n; \mathsf{y}^n|\mathsf{a})/n$, conditioned on $\mathsf{a} = a$, can be manipulated into the form (12) for the AWGN channel. Thus, it has the same distribution as the random variable

$$\log(1 + \beta\mathsf{s}) + \frac{1}{n}\sqrt{\frac{\beta\mathsf{s}}{\beta\mathsf{s}+1}}\sum_{l=1}^{n} \mathsf{w}_l \ , \tag{19}$$

where $\mathsf{s} := |\mathsf{a}|^2 P/N_0$ is an exponential random variable corresponding to the realized signal-to-noise ratio, and $\mathrm{SNR} := \sigma_a^2 P/N_0$ is the *average* signal-to-noise ratio. Again, w_l, $l = 1, 2, \ldots, n$ are i.i.d. Laplace random variables with mean zero and variance two. As a result, both the analysis and numerical methods associated with static fading can rely on averaging the results in Section IV-B for the Gaussian case.

From (19), we immediately see that the distribution of mutual information relates to outage probability [12]. Specifically, for $\beta = 1$ and each $\mathsf{s} = s$, the sum in (19) conditionally converges in mean-square to zero, so that the mutual information rate $i(\mathsf{x}^n; \mathsf{y}^n|\mathsf{a})/n$ conditionally converges in mean-square to $\log(1 + s)$. Thus, the distribution of mutual information rate converges to $\Pr[\log(1 + \mathsf{s}) \leq i]$, the well-known outage probability for Gaussian channels with fading. The second term in (19) takes into account the effects of finite blocklength. In fact, (19) suggests that for large average signal-to-noise ratio SNR, the factor $\sqrt{\mathsf{s}/(\mathsf{s}+1)} \approx 1$ with high probability, and the effects of finite blocklength are essentially identical to the AWGN case.

ACKNOWLEDGMENT

This research has been supported in part by NSF Grant Nos. ECS03-29766 and CCF05-15012 and by the University of Notre Dame Faculty Research Program.

REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.
[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.
[3] A. Feinstein, "A New Basic Theorem of Information Theory," *IEEE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.
[4] C. E. Shannon, "Certain Results in Coding Theory for Noisy Channels," *Inform. Contr.*, vol. 1, pp. 6–25, Sept. 1957.
[5] D. Blackwell, L. Breiman, and A. J. Thomasian, "The Capacity of a Class of Channels," *The Annals of Math. Stat.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.
[6] R. M. Fano, *Transmission of Information: A Statistical Theory of Communication*. New York: John Wiley & Sons, Inc., 1961.
[7] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco: Holday-Day, 1964.
[8] R. Ash, *Information Theory*. New York: Interscience Publishers, 1965.
[9] T. S. Han and S. Verdú, "Approximation Theory of Output Statistics," *IEEE Trans. Inform. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
[10] S. Verdú and T. S. Han, "A General Formula for Channel Capacity," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1147–1157, July 1994.
[11] T. S. Han, *Information Spectrum Methods in Information Theory*. Berlin: Springer, 2003.
[12] L. H. Ozarow, S. Shamai (Shitz), and A. D. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 5, pp. 359–378, May 1994.
[13] A. J. Thomasian, "Error Bounds for Continuous Channels," in *Fourth London Symposium on Information Theory*, C. Cherry, Ed. Butterworth, 1961, pp. 46–60.
[14] T. T. Kadota, "Generalization of Feinstein's Fundamental Lemma," *IEEE Trans. Inform. Theory*, vol. 16, no. 6, pp. 791–792, Nov. 1970.
[15] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, June 1974.
[16] S. Kotz, T. Kozubowski, and K. Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Birkhäuser, May 2001.
[17] M. Effros and A. Goldsmith, "Capacity Definitions and Coding Strategies for General Channels with Side Information," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Cambridge, MA, Aug. 1998, p. 39.