On fixed input distributions for noncoherent communication over high SNR Rayleigh fading channels

RongRong Chen, Bruce Hajek, Ralf Koetter and Upamanyu Madhow

August 31, 2004

Abstract

It is well-known that independent and identically distributed Gaussian inputs, scaled appropriately based on the Signal-to-Noise Ratio (SNR), achieve capacity on the Additive White Gaussian Noise (AWGN) channel at all values of SNR. In this correspondence, we consider the question of whether such good input distributions exist for frequency non-selective Rayleigh fading channels, assuming that neither the transmitter nor the receiver has *a priori* knowledge of the fading coefficients. In this *noncoherent* regime, for a Gauss-Markov model of the fading channel, we obtain explicit mutual information bounds for the Gaussian input distribution. The fact that Gaussian input generates bounded mutual information motivates the search for better choices of fixed input distributions for high-rate transmission over rapidly varying channels. Necessary and sufficient conditions are derived for characterizing such distributions for the worst-case scenario of memoryless fading, using the criterion that the mutual information is unbounded as the SNR gets large. Examples of both discrete and continuous distributions that satisfy these conditions are given. A family of fixed input distributions with mutual information growth rate of $O((\log\log \mathrm{SNR})^{1-u}), u > 0$ are constructed. It is also proved that there does not exist a single fixed input distribution that achieves the optimal mutual information growth rate of $\log\log \mathrm{SNR}$.

keywords Channel capacity, fading channels, high signal-to-noise ratio (SNR), noncoherent communication, Rayleigh fading

1 Introduction

In this correspondence, we provide an information-theoretic perspective on the choice of input distributions for high-rate data transmission over rapidly fading, frequency non-selective, wireless channels. The standard practice of tracking the channel based on known pilot symbols is expensive in this setting. Instead, we consider the paradigm of *noncoherent* communication, in

[1]This work was supported by the National Science Foundation under grants NCR96-24008CAR, CCR 99-79381, ITR 00-85929 and by the Army Research Office under grants DAAG55-98-1-0219 and DAAD19-00-1-0567.
R.R. Chen is with the department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112 USA (email: rchen@ece.utah.edu)
B. Hajek and R. Koetter are with the department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801 USA (email: b-hajek@uiuc.edu; koetter@uiuc.edu)
U. Madhow is with the department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 (email: madhow@ece.ucsb.edu)

which the receiver does not have *a priori* knowledge or estimates of the channel. Pilot-aided channel estimation is a suboptimal receiver strategy that falls within this framework.

For the classical AWGN channel, it is well-known that independent and identically distributed (i.i.d.) Gaussian inputs, when appropriately scaled according to SNR, achieve capacity over the entire range of SNR. This motivates us to investigate whether such good fixed input distributions exist for the noncoherent Rayleigh fading channel. Our focus is on the high SNR regime, in order to explore the feasibility of high-rate transmission under rapid channel time variations.

We first study the Gauss-Markov channel and derive the mutual information bounds generated by the Gaussian input. Our results show that the mutual information achieved by the Gaussian input remains bounded at high SNR. This is in contrast to the block fading channel model, studied in detail in [1, 2] for finite SNR: for this model, Gaussian input results in unbounded mutual information in the high SNR limit [3], as long as the channel is constant over a block of at least two symbols. The block fading model, and other blockwise constant channel models, are also attractive from the point of view of design of practical turbo coded modulation schemes using conventional signal constellations, and iterative channel estimation, demodulation, and decoding [4, 2, 5, 6]. However, our results suggest that, even though the block fading channel model is considered to be a good approximation for a continuously varying fading channel with block interleaving (or frequency hopping), the accuracy of the approximation breaks down as the SNR gets large. The boundness of mutual information generated by Gaussian input is also obtained by Lapidoth and Moser [7] for more general fading processes.

The suboptimality of Gaussian input motivates us to consider alternative input distributions. We look for good input distributions that generate unbounded mutual information in the high SNR limit and study the mutual information growth rate as a function of SNR. Necessary and sufficient conditions are derived to characterize such distributions for memoryless fading. Examples of both discrete and continuous distributions that satisfy these conditions are given. In particular, we identify input distributions that are not suitable for high-rate communication due to the boundness of mutual information at high SNR. Such distributions include any continuous distribution that is bounded around 0 (such as the Gaussian distribution) and any constant amplitude distributions (such as the PSK signaling).

For memoryless fading, Taricco and Elia [8] and Lapidoth and Moser [7] proved that the noncoherent channel capacity grows double logarithmically ($\log\log(\mathrm{SNR})$) as a function of SNR. It is shown in [7] that this rate of growth also applies to more general ergodic fading processes. In this correspondence, we propose a special class of discrete input distributions that, when scaled appropriately with SNR, has a mutual information growth rate approaching this optimal capacity growth rate. It is shown by Abou-Faycal *et al.* [9] that the optimal input distribution for memoryless fading is discrete and has finite mass points. However, the exact number and the location of these mass points vary significantly with SNR and have to be computed numerically. In this correspondence, we concentrate on input distributions with closed-form analytical expressions.

This correspondence is organized as follows. Section 2 contains basic definitions. Section 3 contains bounds on the mutual information for i.i.d. Gaussian input on a Gauss-Markov channel. The remainder of the correspondence focuses on characterizing input distributions that generate unbounded mutual information in the high-SNR limit. We restrict attention to memoryless fading for this purpose. Section 4 contains necessary and sufficient conditions for an input distribution to generate unbounded mutual information. Section 5 focuses on discrete distributions, and provides an example of a class of distributions that generates unbounded mutual information, with a rate of growth of SNR approaching the optimal capacity. Concluding remarks are given in Section 6.

2 Entropy, Differential Entropy, and Mutual Information

We first review the basic concepts of entropy, differential entropy, and mutual information. When X is a discrete random variable with a probability distribution $p_i = \mathbf{P}(X = x_i), i = 1, 2, \cdots$, the entropy of X, denoted by $H(X)$, is defined as $H(X) = -\sum_{i=1}^{\infty} p_i \log p_i$. The logarithm is taken to the base e unless otherwise stated. Note that when X is discrete, we have $H(X) \in [0, +\infty]$. When X is a continuous random variable, the differential entropy of X is defined as

$$\begin{aligned} h(X) &= -\int p(x) \log p(x)\, dx \\ &= \left[-\int_{\{x:\, p(x) \geq 1\}} p(x) \log p(x)\, dx \right] + \\ &\quad \left[-\int_{\{x:\, p(x) \leq 1\}} p(x) \log p(x)\, dx \right]. \end{aligned}$$

The last equality is well defined except when the first term is equal to $-\infty$ and the second term is equal to $+\infty$. When X is a continuous random variable, we have $h(X) \in [-\infty, +\infty]$.

For a general measure theoretic definition of the mutual information $I(X; Z)$ between two random variables X and Z, we refer to Pinsker [10]. According to this definition, $I(X; Z)$ take values in $[0, +\infty]$. When X is discrete, we have $I(X; Z) = H(X) - H(X|Z)$, provided that $H(X)$ and $H(X|Z)$ are not both equal to $+\infty$. When X is a continuous random variable, we have $I(X; Z) = h(X) - h(X|Z)$, assuming that the two terms $h(X)$ and $h(X|Z)$ are not both equal to $+\infty$ or both equal to $-\infty$.

3 Mutual Information Bounds for Gaussian Inputs

Consider a Gauss-Markov fading channel model as follows:

$$\begin{aligned} S_{n+1} &= \alpha S_n + \sqrt{1-\alpha^2} U_n \\ Y_n &= S_n X_n + \sigma W_n. \end{aligned} \tag{1}$$

For each time instant n, S_n, X_n, and Y_n represent the fading coefficient, the channel input, and the channel output, respectively. The sequences of random variables $\{U_n\}$ and $\{W_n\}$ are i.i.d. complex Gaussian, $\mathbf{C}N(0, 1)$ distributed. The fading process $\{S_n\}$ is generated by a first-order Markov process with a parameter $0 \leq \alpha < 1$. The constant σ equals to the square root of the noise variance. We assume that noncoherent reception is employed; namely, the receiver has no explicit information about the channel fading coefficients $\{S_n\}$.

Denote the sequence $\{Y_1, \cdots, Y_n\}$ by Y_1^n and the sequence $\{X_1, \cdots, X_n\}$ by X_1^n. Let

$$I_n = \frac{1}{n} I(X_1^n; Y_1^n) = \frac{1}{n}[h(Y_1^n) - h(Y_1^n | X_1^n)]$$

be the average mutual information achieved by the input signals X_1^n through n channel uses. When X_1^n is i.i.d. Gaussian, the following theorem gives upper and lower bounds on I_n.

Theorem 3.1 *Consider the Gauss-Markov channel model defined in Equation* (1)*. Suppose that the channel input X_1^n is i.i.d. complex Gaussian $\mathcal{CN}(0,1)$ distributed. We have the following estimates:*

(1) *Upper bound:*

$$I_n \leq \log\left(1+\sigma^2\right) - \int_0^\infty e^{-x} \log\left[x\left(1-\alpha^2\right)+\sigma^2\right] dx.$$

In particular, we have

$$I_n \leq -\log\left(1-\alpha^2\right) + \gamma, \qquad \text{as } \sigma^2 \to 0, \tag{2}$$

where $\gamma = -\int_0^\infty e^{-x} \log x \, dx = 0.5772\cdots$ is Euler's constant.

(2) *Lower bound:*

$$I_n \geq \int_0^\infty e^{-x} \log\left(x+\sigma^2\right) dx - \frac{1}{n} \log\left|\sigma^2 1_n + A_n\right|, \tag{3}$$

where 1_n is the $n \times n$ identity matrix; A_n is the $n \times n$ covariance matrix of the vector S_1^n. The latter is a Toeplitz matrix with entries $a_{ij} = \alpha^{|i-j|}$. Moreover,

$$\lim_{n\to\infty} I_n \geq \int_0^\infty e^{-x} \log\left(x+\sigma^2\right) dx - f(\alpha), \tag{4}$$

where

$$f(\alpha) = \log\left[\frac{\sqrt{k_1(\alpha)\cdot k_2(\alpha)} + k_3(\alpha)}{2}\right].$$

with $k_1(\alpha) = \sigma^2(1+\alpha)^2+1-\alpha^2$,$k_2(\alpha) = \sigma^2(1-\alpha)^2+1-\alpha^2$ and $k_3(\alpha) = \sigma^2\left(1+\alpha^2\right)+1-\alpha^2$. In particular, we have

$$\lim_{n\to\infty} I_n \geq -\log\left(1-\alpha^2\right) - \gamma, \qquad \text{as } \sigma^2 \to 0. \tag{5}$$

As seen from Equation (2), for any fixed $0 \leq \alpha < 1$, the average mutual information achieved by the Gaussian input is bounded above by a constant at high SNR. The same result was also obtained by Lapidoth and Shamai [11] in the case of $\alpha = 0$. As $\alpha \to 1$, both the upper bound (2) and the lower bound (5) in the high SNR limit approach infinity and differ by a constant 2γ. The lower bound, however, becomes trivial for small values of α.

Proof.

(1) Proof of the upper bound.

By definition,

$$I_n = \frac{1}{n}[h(Y_1^n) - h(Y_1^n | X_1^n)]. \tag{6}$$

For the first term, we have

$$\frac{1}{n}h(Y_1^n) \le \frac{1}{n}\sum_{i=1}^{n} h(Y_i) = h(Y_1) \le \log(\pi e) + \log\left(1+\sigma^2\right). \tag{7}$$

For the second term, we have

$$\begin{aligned}
\frac{1}{n}h(Y_1^n|X_1^n) &= \frac{1}{n}\sum_{i=1}^{n} h(Y_i|Y_1^{i-1}, X_1^n) \\
&\ge \frac{1}{n}\sum_{i=1}^{n} h(Y_i|Y_1^{i-1}, X_1^n, S_{i-1}) \\
&= \frac{1}{n}\sum_{i=1}^{n} h(Y_i|X_i, S_{i-1}) = h(Y_2|X_2, S_1).
\end{aligned}$$

Since $Y_2 = X_2S_2 + \sigma W_2 = X_2\left[\alpha S_1 + \sqrt{1-\alpha^2}U_1\right] + \sigma W_2$, the distribution of Y_2, conditioned on X_2 and S_1, is Gaussian with variance $|X_2|^2\left(1-\alpha^2\right)+\sigma^2$. It follows that

$$\begin{aligned}
\frac{1}{n}h(Y_1^n|X_1^n) &\ge h(Y_2|X_2, S_1) \\
&= \mathbf{E}_{X_2}\log\left[\pi e\left(|X_2|^2\left(1-\alpha^2\right)+\sigma^2\right)\right] \\
&= \log(\pi e) + \int_0^\infty e^{-x}\log\left[x\left(1-\alpha^2\right)+\sigma^2\right]dx.
\end{aligned} \tag{8}$$

Equations (7) and (8) combine to give the upper bound.

(2) Proof of the lower bound.

Again, we start from Equation (6). First, we find a lower bound on $\frac{1}{n}h(Y_1^n)$:

$$\begin{aligned}
\frac{1}{n}h(Y_1^n) &= \frac{1}{n}\sum_{i=1}^{n} h(Y_i|Y_1^{i-1}) \ge \frac{1}{n}\sum_{i=1}^{n} h(Y_i|Y_1^{i-1}, S_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} h(Y_i|S_i) = h(Y_1|S_1) \\
&= \mathbf{E}_{S_1}\log\left[\pi e\left(|S_1|^2+\sigma^2\right)\right] \\
&= \log(\pi e) + \int_0^\infty e^{-x}\log\left(x+\sigma^2\right)dx.
\end{aligned} \tag{9}$$

Next, we derive an upper bound on $\frac{1}{n}h(Y_1^n|X_1^n)$. To simplify notation, denote X_1^n by X. Conditioned on X, Y_1^n is Gaussian with the covariance matrix $Q(X) = \sigma^2 1_n + \Lambda_X A_n \Lambda_X^H$, where Λ_X denotes a diagonal square matrix with diagonal elements $X_1, \cdots, X_n$. We have

$$\begin{aligned}
\frac{1}{n}h(Y_1^n|X_1^n) &= \frac{1}{n}\mathbf{E}_X \log|\pi e Q(X)| \\
&\le \frac{1}{n}\log \mathbf{E}_X|\pi e Q(X)| \quad \text{(Jensen's inequality)} \\
&= \log(\pi e) + \frac{1}{n}\log \mathbf{E}_X\left|\sigma^2 1_n + \Lambda_X A_n \Lambda_X^H\right| \\
&= \log(\pi e) + \frac{1}{n}\log \mathbf{E}_X\left|\sigma^2 1_n + A_n(\Lambda_X^H\Lambda_X)\right| \\
&= \log(\pi e) + \frac{1}{n}\log\left|\sigma^2 1_n + A_n\right|.
\end{aligned} \tag{10}$$

Substituting equations (9) and (10) into Equation (6), we get Equation (3).

Let us define $\Phi_n = \sigma^2 1_n + A_n$. The matrix Φ_n is a Toeplitz matrix. Denote the eigenvalues of Φ_n by $\{\lambda_k^n, k = 1, \cdots, n\}$ and the k-th diagonal element by ϕ_k. Let $\Phi(x) = \sum_{-\infty}^{\infty} \phi_k x^{-j2\pi k}, j = \sqrt{-1}$. We then apply the Toeplitz distribution theorem [12] to obtain

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log \left|\sigma^2 1_n + A_n\right| &= \lim_{n\to\infty} \frac{1}{n} \log \Pi_{k=1}^n \lambda_k^n \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log \lambda_k^n = \int_{-1/2}^{1/2} \log[\Phi(x)]\, dx = f(\alpha).\end{aligned}$$

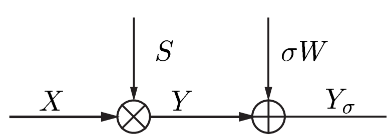
Remark: When the input is i.i.d. and has constant amplitude: $|X_i|^2 = 1$, we have the following upper bound:

$$I_n \leq \log(1+\sigma^2) - \log(1-\alpha^2+\sigma^2). \tag{11}$$

The proof for this upper bound is similar to the proof for the upper bound given in Theorem 3.1 for the Gaussian input. The only difference is that one should replace the last equality in Equation (8) by $\log(\pi e) + \log(1-\alpha^2+\sigma^2)$. By letting $\sigma \to 0$, Equation (11) implies that the mutual information generated by PSK input is also bounded from above by $-\log(1-\alpha^2)$. This suggests that in the high SNR regime, it is power-inefficient to use large PSK constellations.

4 Input Distributions that Generate Unbounded Mutual Information at High SNR

In the remainder of this correspondence, we study the worst-case i.i.d. memoryless fading channel, and characterize input distributions that generate unbounded mutual information in the high-SNR limit.



$$Y_\sigma = SX + \sigma W = Y + \sigma W$$

Figure 1: The i.i.d. fading channel.

4.1 The i.i.d. fading channel model

The i.i.d. fading channel model can be obtained by setting $\alpha = 0$ in the Gauss-Markov channel model defined by Equation (1). To simplify notations, we drop the time index n and model the channel as shown in Figure 1, where Y represents the output signal when the noise vanishes, Y_σ

represents the output signal corresponding to a given noise variance σ^2. Note that the input X satisfies the power constraint $E(|X|^2) = 1$.

A useful property of this channel model can be seen from its channel transition probability:

$$p_{Y_\sigma|X}(y|x) = \frac{1}{\pi\big(|x|^2+\sigma^2\big)} e^{-|y|^2/(|x|^2+\sigma^2)},$$

which depends only on the amplitudes of the input and output signals. This implies that the mutual information between X and Y_σ is the same as the mutual information between $|X|$ and $|Y_\sigma|$:

$$I(X;Y_\sigma) = I(|X|;|Y_\sigma|). \tag{12}$$

Hence, $I(X;Y_\sigma)$ is completely determined by the amplitude distribution of the input signal. This property is used in later sections.

A necessary condition for an input distribution to generate unbounded mutual information at high SNR is given by Lapidoth and Moser (Theorem 4.3, [7]). Using this necessary condition, one can show that any continuous input distribution with a finite density function that is bounded around 0 (including the Gaussian distribution) generates bounded mutual information at high SNR.

Next, we derive necessary and sufficient conditions on the input distribution of X such that

$$\lim_{\sigma\to 0} I(X;Y_\sigma) = +\infty. \tag{13}$$

4.2 Necessary and sufficient conditions

We derive necessary and sufficient conditions to characterize input distributions that satisfy Equation (13). To simplify this problem, we prove that the limit of the mutual information generated by a fixed input distribution, as $\sigma \to 0$, is equal to the mutual information it generates when the noise vanishes ($\sigma = 0$). In other words, the mutual information, as a function of σ, is continuous at $\sigma = 0$.

Theorem 4.1 *For any discrete or continuous distribution of X, we have*

$$\lim_{\sigma\to 0} I(X;Y_\sigma) = I(X;Y).$$

In particular,

$$\lim_{\sigma\to 0} I(X;Y_\sigma) = +\infty \Longleftrightarrow I(X;Y) = +\infty.$$

Theorem 4.1 can be proved by first demonstrating convergence in variation: for any $A \subset \mathbb{C}^2$, we have $\lim_{\sigma\to 0} \mathbf{P}_{X,Y_\sigma}(A) = \mathbf{P}_{X,Y}(A)$, and then apply Pinsker's results [10] (page 13).

Theorem 4.1 states that whether a fixed input distribution leads to unbounded mutual information at high SNR is determined by whether it generates infinite mutual information when the noise vanishes. Hence, it suffices to focus on $I(X;Y)$ with $Y = SX$. As shown in Figure 2, this particular channel model can be reduced to a simple additive noise channel model by transforming the original input and output signals into signals on the logarithm domain. This transformation leads to some useful necessary and sufficient conditions for input distributions that generate unbounded mutual information at high SNR. We summarize these conditions in the following theorem.

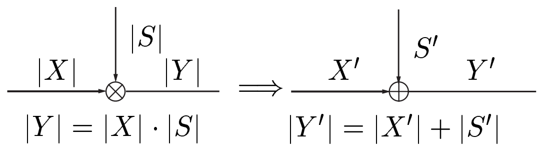


Figure 2: Transformation to the logarithm domain.

Theorem 4.2 *Define* $X' = \log|X|$ *and* $Y' = \log|Y|$. *Let* $\log(0) = -\infty$. *For any discrete or continuous distribution of* X, *the following assertions hold:*

(1) $I(X;Y) = I(X';Y')$. *In particular,* $I(X;Y) = +\infty \Longleftrightarrow I(X';Y') = +\infty$.

(2) If Y' *is a continuous random variable with a finite density function, then*

$$I(X';Y') = +\infty \Longleftrightarrow h(Y') = +\infty.$$

(3) If X *is a continuous random variable with a finite density function, then*

$$h(X') = +\infty \Longrightarrow I(X';Y') = +\infty.$$

Part (2) of this theorem gives a necessary and sufficient condition applicable to any continuous or discrete input distribution. However, it is usually difficult to verify whether or not $h(Y') = +\infty$. Part (3) of this theorem gives a sufficient condition for a continuous distribution to generate unbounded mutual information, which is much easier to use.

Proof. From Equation (12), we know that $I(X;Y) = I(|X|;|Y|)$. Since the mapping between $(|X|,|Y|)$ and (X',Y') is one-to-one, we also have $I(|X|;|Y|) = I(X';Y')$. Part (1) of the theorem follows.

Taking the logarithm of both sides of the Equation $|Y| = |S||X|$, we get

$$\log|Y| = \log|S| + \log|X|. \tag{14}$$

Let $S' = \log|S|$. We can rewrite Equation (14) as

$$Y' = S' + X'.$$

Simple calculations show that $h(S') = h(|S|) - E\log|S|$ is finite, therefore $h(Y'|X') = h(Y' - X'|X') = h(S'|X') = h(S')$ is also finite. Since $I(X';Y') = h(Y') - h(Y'|X')$, we see that $I(X';Y') = +\infty \Longleftrightarrow h(Y') = +\infty$. This proves part (2).

Next, we prove part (3). If X is a continuous random variable, so is X'. Therefore,

$$\begin{aligned} I(X';Y') &= h(X') - h(X'|Y') = h(X') - h(X' - Y'|Y') \\ &= h(X') - h(-S'|Y') \geq h(X') - h(-S'). \end{aligned}$$

Since $h(-S')$ is finite, part (3) of the theorem follows immediately. ■

4.3 An example of a "good" continuous input distribution

We can apply part (3) of Theorem 4.2 to obtain continuous input distributions that generate unbounded mutual information at high SNR. An example of such a distribution is given by

$$p_{|X|}(a) = \begin{cases} \dfrac{c_1}{a\log(1/(c_2 a))[\log\log(1/(c_2 a))]^2} & \text{if } 0 < a < e^{-3}/c_2, \\ 0 & \text{else,} \end{cases}$$

where $c_1 = \log 3$ and $c_2 = 0.0164$ are chosen such that $\int_0^\infty p_{|X|}(a)\,da = 1$ and $\int_0^\infty a^2 p_{|X|}(a)\,da = 1$. A plot of this density function is given in Figure 3. Simple calculations show that $h(X') = h(\log(|X|)) = +\infty$. From part (3) of Theorem 4.2, it follows that $I(X;Y) = I(X';Y') = +\infty$.

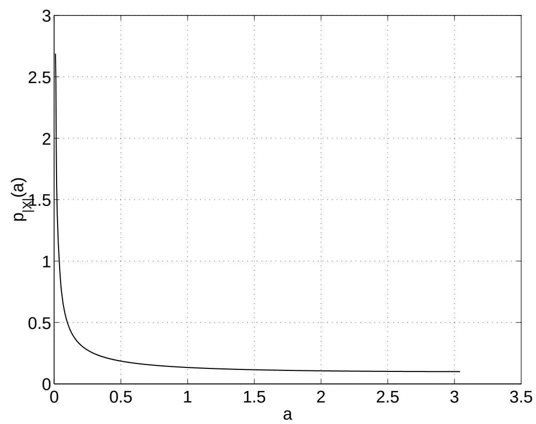


Figure 3: An example of continuous input distributions.

5 Discrete Input Distributions

It is difficult to apply directly the necessary and sufficient conditions for unbounded mutual information in Theorem 4.2 to discrete input distributions, which are of more interest in practice. Instead, we provide in this section an example of a class of discrete distributions that provides unbounded mutual information, and analyze a simple receiver to show that the rate of growth of mutual information with SNR is $\Omega([\log\log \mathrm{SNR}]^u)$, where $0 < u < 1$ ranges over the members of the class. As $u \to 1$, this approaches the $O(\log\log \mathrm{SNR})$ rate of growth of capacity for the memoryless channel.

Since $I(X;Y_\sigma) = I(|X|;|Y_\sigma|)$, in this section, we drop the magnitude notation and denote $|X|$ by X and $|Y_\sigma|$ by Y_σ, respectively.

5.1 Signal constructions

Consider a discrete input distribution of X having mass points at $x_i = \beta L^{-i}$, each with probability $p_i, i = 2, 3, \cdots$. Here $L > 1$ is a fixed constant, β is a scalar constant such that the energy constraint is satisfied: $\sum_{i=2}^{\infty} p_i x_i^2 = 1$. In addition, we assume that

$$H(X) = -\sum_{i=2}^{\infty} p_i \log p_i = +\infty.$$

An example of such a distribution with $L = 2$, $p_i = \frac{t}{i(\log i)^{1.1}}$, where t is a normalization constant, is plotted in Figure 4. Note that if one approximates this discrete distribution by a continuous distribution, the density function of the approximation looks like the one in Figure 3.

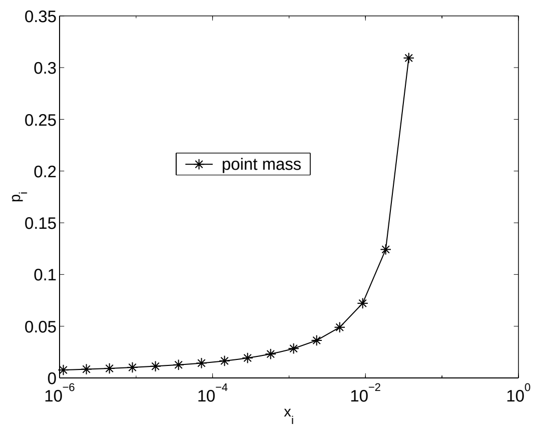


Figure 4: An example of discrete input distributions.

Similar signal constructions were first studied by Taricco and Elia [8], where the number of mass points was finite and only the uniform distribution was used. In [8], the parameter L approaches infinity as the SNR grows, while in this work, we fix L and let the number of mass points equal infinity.

For any $L > 1$, the signals constructed above lead to unbounded mutual information. For brevity, we prove it only for a particular choice of L that satisfies

$$P\big(|S| \notin [L^{-1/3}, L^{1/3}]\big) < \epsilon/2 \tag{15}$$

for a fixed $0 < \epsilon < 1$. Similar proofs also work for other values of L.

5.2 A lower bound for $I(X; Y_\sigma)$ based on a simple receiver

We derive a lower bound for $I(X; Y_\sigma)$ based on a simple receiver structure. Given a fixed noise level σ, the receiver chooses a positive integer N, which can be interpreted as the resolution parameter, and pretends that one of the input signals from the set $\{x_2, \cdots, x_N\}$ was sent.

In other words, define

$$X_N = \begin{cases} X & \text{if } X \in \{x_2, \cdots, x_N\}, \\ 0 & \text{else.} \end{cases}$$

Based on the received signal Y_σ, the receiver tries to estimate of X_N according to some decision rule which we describe later. Since $Y_\sigma \to X \to X_N$ forms a Markov chain, from the data processing inequality [42, p. 32], it follows that $I(X;Y_\sigma) \geq I(X_N;Y_\sigma)$. As the receiver improves its resolution N, $I(X_N;Y_\sigma)$ becomes a tighter lower bound for $I(X;Y_\sigma)$. The main result is summarized in the theorem below.

Theorem 5.1 *For arbitrarily fixed ϵ, L, and σ, let $N_{\max}(\sigma)$ be the largest N such that*

$$P\Big(|\sigma W| > \gamma L^{-N}\Big) < \epsilon/2, \tag{16}$$

where $\gamma = \beta(L^{2/3} - L^{1/3})/2$. Then for any $N \leq N_{\max}(\sigma)$, we have:

$$I(X;Y_\sigma) \geq I(X_N;Y_\sigma) \geq (1-\epsilon)H(X_N) - 1. \tag{17}$$

In particular, for any $\epsilon < 1$,

$$\lim_{\sigma\to 0} I(X;Y_\sigma) = +\infty.$$

Proof. First, we introduce some notation. For each integer $2 \leq i \leq N$, define a decision interval A_i that contains x_i as $A_i = \Big[x_i\, L^{-1/3} - \gamma L^{-N},\ x_i\, L^{1/3} + \gamma L^{-N}\Big]$. An illustration of decision intervals is given in Figure 5.

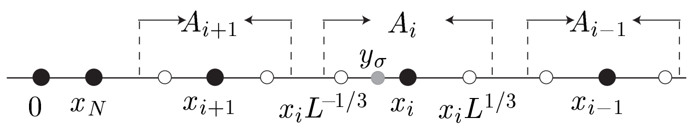


Figure 5: Illustration of decision intervals.

Define a random variable E as

$$E = \begin{cases} 1, & \text{if } |S| \in [L^{-1/3}, L^{1/3}] \text{ and } |\sigma W| \leq \gamma L^{-N}, \\ 0, & \text{else.} \end{cases}$$

Note that for any $N \leq N_{\max}(\sigma)$,

$$\begin{aligned} P(E=0) &\leq P(|S| \notin [L^{-1/3}, L^{1/3}]) + \mathbf{P}(|\sigma W| > \gamma L^{-N}) \\ &\leq \epsilon/2 + \epsilon/2 = \epsilon. \end{aligned}$$

First, we have

$$\begin{aligned} H(X_N \,|\, Y_\sigma) &\leq H(X_N, E \,|\, Y_\sigma) = H(E \,|\, Y_\sigma) + H(X_N \,|\, E, Y_\sigma) \\ &\leq 1 + H(X_N \,|\, E, Y_\sigma) \\ &= 1 + P(E=1)H(X_N \,|\, E=1, Y_\sigma) \\ &\quad\ + P(E=0)H(X_N \,|\, E=0, Y_\sigma) \\ &\leq 1 + H(X_N \,|\, E=1, Y_\sigma) + P(E=0)H(X_N \,|\, E=0, Y_\sigma) \\ &\leq 1 + H(X_N \,|\, E=1, Y_\sigma) + \epsilon\, H(X_N \,|\, E=0) \\ &= 1 + H(X_N \,|\, E=1, Y_\sigma) + \epsilon\, H(X_N). \end{aligned} \tag{18}$$

It is important to realize that under the assumption that $E = 1$, for any possible transmitted signal $X_N = x_i$, $i = 2, \cdots, N$, we must have $Y_\sigma \in A_i$. Since the decision intervals $\{A_i, i = 2, \cdots, N\}$ are nonoverlapping, it follows that when conditioned on Y_σ and $E = 1$, X_N is uniquely determined. Hence, $H(X_N \mid E = 1, Y_\sigma) = 0$. Substituting this into the last equality of Equation (18), we get

$$H(X_N \mid Y_\sigma) \le 1 + \epsilon\, H(X_N). \tag{19}$$

Finally, we obtain

$$\begin{aligned} I(X_N; Y_\sigma) &= H(X_N) - H(X_N \mid Y_\sigma) \\ &\ge H(X_N) - \epsilon\, H(X_N) - 1 \\ &= (1 - \epsilon)\, H(X_N) - 1. \end{aligned}$$

This proves the second inequality in Equation (17). As $\sigma \to 0$, we have $N_{\max}(\sigma) \to \infty$. Therefore, by letting $N \to \infty$, the right side of Equation (17) converges to $(1 - \epsilon)\, H(X) - 1 = +\infty$, assuming that $\epsilon < 1$. ■

5.3 Growth rate of $I(X; Y_\sigma)$ as a function of σ

For some particular choices of the distribution $\{p_i, i = 2, \cdots\}$, we study how fast the mutual information $I(X; Y_\sigma)$ grows as a function of σ. Here we fix ϵ, L, and σ.

Lemma 5.1 *$N_{max}(\sigma)$ grows in the order of $O\big(\log\big(1/\sigma^2\big)\big)$.*

Proof. By solving

$$P\big(|\sigma W| > \gamma L^{-N}\big) = \int_{\gamma L^{-N}/\sigma}^{\infty} 2\, a\, e^{-a^2}\, da = e^{-\gamma^2 L^{-2N}/\sigma^2} < \epsilon/2,$$

we obtain

$$N < \frac{\log(1/\sigma^2)}{2 \log L} - \frac{\log\log(2/\epsilon) - 2\log\gamma}{2\log L} = O(\log(1/\sigma^2)). \quad ■$$

Theorem 5.2 *Let $p_i = \frac{t}{i(\log i)^{1+u}}$, where $0 < u < 1$ and t is a constant such that $\sum_{i=2}^{\infty} p_i = 1$. The mutual information $I(X; Y_\sigma)$ grows in the order of $\Omega\big(\big[\log\log\big(1/\sigma^2\big)\big]^{1-u}\big)$.*

Proof. First, we compute $H(X_N)$:

$$\begin{aligned} H(X_N) \approx -\sum_{i=2}^{N} p_i \log p_i = t \sum_{i=2}^{N} \frac{1}{i(\log i)^u} \\ + t(1+u) \sum_{i=2}^{N} \frac{\log\log i}{i(\log i)^{1+u}} - (t\log t) \sum_{i=2}^{N} \frac{1}{i(\log i)^{1+u}}. \end{aligned}$$

Since the last two terms in the last equality both converge as $N \to \infty$, the growth rate of $H(X_N)$ is determined by the growth rate of the first term.

Because

$$\int_2^N \frac{1}{x(\log x)^u}\,dx = \frac{1}{1-u}(\log N)^{1-u} - \frac{1}{1-u}(\log 2)^{1-u},$$

we have

$$H(X_N) = \frac{t}{1-u}(\log N)^{1-u} + O(1), \qquad \text{as } N \to \infty. \tag{20}$$

From Theorem 5.1 and Lemma 5.1, it follows that

$$\begin{aligned} I(X;Y_\sigma) &\geq I\big(X_{N_{\max}(\sigma)};Y_\sigma\big) \geq (1-\epsilon)H\big(X_{N_{\max}(\sigma)}\big) - 1 \\ &= (1-\epsilon)\left[\frac{t}{1-u}\big(\log N_{\max}(\sigma)\big)^{1-u}\right] + O(1) \\ &= O\big(\big[\log\log\big(1/\sigma^2\big)\big]^{1-u}\big), \qquad \text{as } \sigma \to 0. \quad \blacksquare \end{aligned}$$

From Theorem 5.2 we see that the growth rate of $\log\log(1/\sigma^2)$ (corresponding to $u = 0$) is not achieved because $\{p_i = t/(i \log i)\}$ is not a valid probability distribution. Furthermore, the following theorem (due to A. Lapidoth) shows that no fixed input distribution can achieve the optimal growth rate of $\log\log(1/\sigma^2)$. This theorem is valid for general fading processes.

Theorem 5.3 (Lapidoth) *Let the fading process $\{S_k\}$, be stationary, of finite differential entropy rate; and of unit second moment. Let the input process $\{X_k\}$ be such that the law of X_k does not depend on the time index k and has unit second-moment. Then*

$$\lim_{\sigma^2 \to 0} \frac{\limsup_{n\to\infty} I(X_1,\cdots,X_n;Y_1,Y_2,\cdots,Y_n)/n}{\log\log(1/\sigma^2)} = 0. \tag{21}$$

Proof: By Lemma 4.5 in Lapidoth-Moser [7] it suffices to prove

$$\lim_{\sigma^2\to 0} \frac{I(X_1;Y_1)}{\log\log(1/\sigma^2)} = 0.$$

Consequently, we shall drop indices and prove

$$\lim_{\sigma^2\to 0} \frac{I(X;SX+\sigma W)}{\log\log(1/\sigma^2)} = 0.$$

Define for any $a > 0$ such that $P(|X| > a) > 0$

$$E = \begin{cases} 0 & \text{if } X = 0 \\ 1 & \text{if } 0 < |X| \leq a \\ 2 & \text{if } |X| > a \end{cases} \tag{22}$$

Let $Y_\sigma = SX + \sigma W$. We can now expand $I(X;SX+\sigma W) = I(X;Y_\sigma)$ as

$$\begin{aligned} I(X;Y_\sigma) &= I(X,E;Y_\sigma) \\ &= I(E;Y_\sigma) + I(X;Y_\sigma|E) \\ &\leq \log 3 + I(X;Y_\sigma|E=1)P(E=1) + I(X;Y_\sigma|E=2)P(E=2) \end{aligned} \tag{23}$$

where we have used the fact that $I(X; Y_\sigma | E = 0) = 0$, because conditioned on $E = 0$ the input X is deterministic.

We next study the term $I(X; Y_\sigma | E = 2)$ and show that it is bounded in the noise variance. That is, we shall show that

$$I(X; SX + \sigma W | E = 2) < I(X; SX | E = 2) < \infty. \tag{24}$$

Note that since X is of unit second moment, it follows that, conditioned on $|X| > a$, the second moment of X is at most $1/\Pr(|X| \geq a)$. Consequently we have

$$\begin{aligned} I(X; SX | E = 2) &= h(SX | E = 2) - h(SX | X, E = 2) \\ &\leq \log \left(\pi e \mathbf{E}[|SX|^2 | E = 2] \right) - \mathbf{E}_{X|E=2}[\log |X|^2 + h(S)] \\ &\leq \log \left(\pi e \mathbf{E}[|S|^2] \frac{1}{\Pr(|X| \geq a)} \right) - \log |a|^2 - h(S) < \infty. \end{aligned} \tag{25}$$

Hence, we obtain equation (24).

We next consider the term $I(X; Y_\sigma | E = 1)$. We note that, conditioned on $E = 1$, the input X is upper bounded by a. In particular, the input's conditional second moment is upper bounded by a^2 and hence finite. Consequently (see Lapidoth-Moser [7] Theorem 4.2) it follows that

$$\limsup_{\sigma^2 \to 0} \frac{I(X; Y_\sigma | E = 1)}{\log \log(1/\sigma^2)} \leq 1.$$

We thus conclude that

$$\lim_{\sigma^2 \to 0} \frac{I(X; SX + \sigma W)}{\log \log(1/\sigma^2)} \leq P(E = 1) = P(0 < |X| \leq a).$$

The theorem now follows because a can be chosen as close to zero as we wish and in this way guaranteeing that $P(0 < |X| \leq a)$ is as close to zero as desired.

6 Conclusions

In this correspondence, we derive explicit bounds of mutual information with i.i.d. Gaussian input for Gauss-Markov fading channels. Our results imply that it may not be appropriate to apply standard code designs for the AWGN channel to fading channels. Intuitively, however, we would still expect Gaussian input to work well at moderate SNR and/or slow fading. The regime where this would happen can be roughly characterized based on our bounds, which show that the high-SNR limit of mutual information for Gaussian input is $-\log(1-\alpha^2)$, up to an additive constant independent of α or SNR. This implies that log(SNR) growth in mutual information can occur if $\alpha \to 1$ as SNR gets large, with $\alpha \sim 1 - k/\text{SNR}$ for some $k > 0$. This is also the regime in which the block fading channel model [1], for which conventional signal constellations are known to work well [2, 5, 6], is a good approximation for the continuously varying channel.

We propose a family of fixed input distributions with mutual information growth rate of $O((\log \log \text{SNR})^{1-u}), u > 0$ at high SNR for memoryless fading ($\alpha = 0$). These input distributions have the attractive feature that a fixed constellation, when scaled appropriately, can be employed for a wide range of SNR. Clearly, these constellations also give unbounded mutual information at high SNR for the Gauss-Markov model with memory ($\alpha > 0$) as well, since the use of channel memory at the receiver can only increase the mutual information. We also show that no fixed input distribution can achieve the optimal mutual information growth rate of $\log\log \mathrm{SNR}$. Further investigation is needed on optimizing the choice of constellation for channels with memory (and on determining the rate of growth of capacity for such channels) in the high-SNR limit. In particular, while information can only be conveyed via amplitude for the noncoherent memoryless fading channel, information can be carried by the phase as well when the channel has memory.

Acknowledgments

The authors are greatly indebted to Dr. Amos Lapidoth for generously providing the proof of Theorem 5.3, which led to a significantly improved version of the original manuscript.

References

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

[2] R. R. Chen, R. Koetter, U. Madhow, and D. Agrawal, “Joint noncoherent demodulation and decoding for the block fading channel: a practical framework for approaching shannon capacity,” *IEEE Trans. Commun.*, vol. 51, pp. 1676–1689, Oct. 2003.

[3] Y. Liang and V. V. Veeravalli, “Capacity of noncoherent time-selective block rayleigh flat-fading channels,” Submitted to *IEEE. Trans. on Inform. Theory*, Feb. 2003.

[4] M. Peleg, S. Shamai, and S. Galán, “Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel,” *IEE Proc. Commun*, vol. 147, pp. 87–95, April 2000.

[5] R. R. Chen, R. Koetter, and U. Madhow, “Joint noncoherent demodulation and decoding for fast Rayleigh fading channels,” *Proc. of the 2003 Conf. on Information Sciences and Systems (CISS 2003)*, Mar. 2003.

[6] N. Jacobsen and U. Madhow, “Reduced-complexity noncoherent communication with differential QAM and iterative receiver processing,” *Proc. of the 2003 Conf. on Information Sciences and Systems (CISS 2003)*, Mar. 2003.

[7] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 2426–2467, Oct. 2003.

[8] G. Taricco and M. Elia, “Capacity of fading channels with no side information,” *Electron. Lett.*, vol. 33, pp. 1368–1370, July 1997.

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

[10] M. S. Pinsker, *Information and information stability of random variables and processes.* translation into English by A. Feinstein, Holden-Day, San Francisco, 1964.

[11] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1133, May 2002.

[12] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. 18, pp. 725–730, 1972.