Asymptotic Behavior of Error Exponents in the Wideband Regime [*]

Xinzhou Wu and R. Srikant
Coordinated Science Lab
and
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
xwu@uiuc.edu, rsrikant@uiuc.edu

December 10, 2006

Abstract

In this paper, we complement Verdú's work on spectral efficiency in the wideband regime by investigating the fundamental tradeoff between rate and bandwidth when a constraint is imposed on the error exponent. Specifically, we consider both AWGN and Rayleigh-fading channels where the input symbols are assumed to have a peak constraint. For the AWGN channel model, the optimal values of $R_z(0)$ and $\dot{R}_z(0)$ are calculated, where $R_z(1/B)$ is the maximum rate at which information can be transmitted over a channel with bandwidth B when the error-exponent is constrained to be greater than or equal to z. The computation of $R_z(0)$ follows Gallager's infinite-bandwidth reliability function computation in [5], while the computation of $\dot{R}_z(0)$ is new and parallels Verdu's second-order calculation for channel capacity in [14]. Based on these calculations, we say that a sequence of input distributions is near optimal if both $R_z(0)$ and $\dot{R}_z(0)$ are achieved. We show that QPSK, a widely-used signaling scheme, is near-optimal within a large class of input distributions for the AWGN channel. Similar results are also established for a fading channel where full CSI is available at the receiver.

1 Introduction

Communications in the wideband regime with limited power has attracted much attention recently. An important characteristic of such communication systems is that they operate at relatively low spectral efficiency (bits per

---

[*] Research supported by DARPA grant F30602-00-2-0542, AFOSR URI grant F49620-01-1-0365 and NSF ITR grant 00-85929

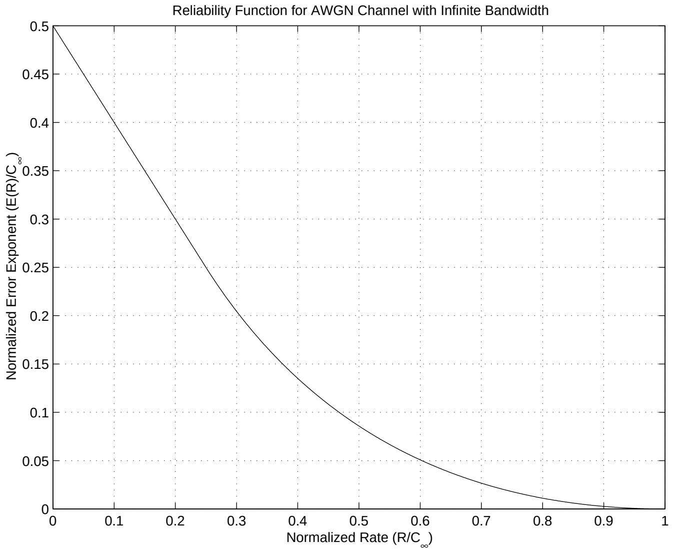

Figure 1: The reliability function for AWGN channel with infinite bandwidth

second per Hz) and energy per bit. The advantages of communication over large bandwidth are many-fold: power savings, higher data rates, more diversity to combat frequency-selective fading, etc. Thus, it is important to understand the ultimate limits of communications in this regime from an information-theoretic point of view, and develop guidelines to design good signaling schemes.

Communications without a bandwidth limit, i.e., the available bandwidth is infinite, is well understood. For the additive white Gaussian noise (AWGN) channel, the capacity, measured in nats per second, converges to a constant P/N_0 when the available bandwidth B goes to infinity. Here P (joule per second) denotes the average power constraint at the input of the channel and $N_0/2$ is the power-spectral density of the Gaussian noise. Furthermore, a Gaussian signaling scheme is not mandatory to achieve this limit. Nearly all signaling schemes are equally good in the sense that the corresponding mutual information converges to the same value in the infinite bandwidth limit. For example, a simple on-off signaling scheme with low duty cycle is capacity-achieving in the infinite bandwidth limit. In [7], Massey showed that all mean zero signaling schemes can achieve this limit.

The *reliability function* $E(R)$, as defined in [4], characterizes the exponential dependence of the probability of decoding error on the codeword length for any coding rate R. Generally, the reliability function of a channel is difficult to compute and is known for all rates only for a few channels. The infinite-bandwidth AWGN channel is one of these channels and its reliability function has the following form[15, 4]

$$E(R) = \begin{cases} \frac{C_\infty}{2} - R & 0 \le R \le \frac{C_\infty}{4}; \\ (\sqrt{C_\infty} - \sqrt{R})^2 & \frac{C_\infty}{4} \le R \le C_\infty, \end{cases} \tag{1}$$

where $C_\infty = P/N_0$ denotes the infinite-bandwidth capacity, as shown in Figure 1. More recently, Gallager [5] investigated the infinite-bandwidth channel reliability function for a broader class of channels. Specifically, it was shown in [5] that a simple expression exists for the infinite-bandwidth reliability function of any discrete-time memoryless channel, if the input is constrained to use a fixed (does not change as the bandwidth increases) set of discrete symbols.

Naturally, the results in the infinite bandwidth regime can be considered as guidelines for designing signaling schemes in the wideband regime as well. However, in the wideband regime (when the available bandwidth is large, but finite), the result based on the infinite bandwidth calculations can be quite misleading. In [14], Verdú pointed out that to understand the performance limit in the wideband regime, two quantities need to be studied: the minimum energy per information bit ($\frac{E_b}{N_0}_{min}$) required to sustain reliable communication, and the slope of spectral efficiency (bits/s/Hz) at the point $\frac{E_b}{N_0}_{min}$. If we treat $C(\cdot)$ as a function of $x = 1/B$, it is easy to see that studying these two quantities is equivalent to studying the optimal values of the following two quantities: infinite-bandwidth capacity $C(0)$ and the first-order derivative of capacity with respect to x, $\dot{C}(0)$. In other words, we need to study both the infinite-bandwidth capacity, and the rate at which this capacity is reached. In [14], it is shown that, while many signaling schemes achieve $C(0)$, only some of these reach the capacity at the fastest possible slope given by $\dot{C}(0)$. We will refer to signaling schemes that achieve both $C(0)$ and $\dot{C}(0)$ as *near-optimal* input distributions in the wideband regime. Further, although $C(0)$ always has the same value for non-fading or fading channels with different CSI, $\dot{C}(0)$ is determined by the CSI and can be very different for different channels.

This paper complements the work of Gallager [5] and Verdú [14] and considers the relationship between probability of decoding error (represented by the reliability function), coding rate, and bandwidth for both AWGN channels and multi-path fading channels. Specifically, we study the maximum rate at which information can be transmitted over a channel, as a function of the available bandwidth, under a certain constraint on the reliability function. For AWGN channels, instead of characterizing the capacity C as a function of $1/B$ as in [14], we are interested in characterizing R_z as a function of $1/B$, where R_z is the maximum rate such that $E(R_z) \ge z$ and $E(R)$ is the reliability function of the channel. In the infinite bandwidth regime, we characterize the optimal rate $R_z(0)$ with respect to a certain error-exponent constraint and study the conditions under which a signaling scheme can achieve this optimal rate. In the wideband regime, both $R_z(0)$ and $\dot{R}_z(0)$ need to be considered. A signaling scheme which can achieve both $R_z(0)$ and $\dot{R}_z(0)$ is said to be *second-order optimal* or *near optimal* with respect to an error-exponent constraint z. In comparison, a signaling scheme which can achieve $R_z(0)$ is said to be *first-order* optimal. Our main contributions in this paper for AWGN channels are two-folds. First, although $R_z(0)$ can be computed from the well-known results [5, 15, 4], we explicitly establish a necessary condition for an input distribution to be first-order optimal and further, we show that a large set of input distributions with mean zero is first-order optimal. This observation is consistent with Massey's proof in [7] from a capacity point of view. Secondly, we provide a closed-form expression for $\dot{R}_z(0)$ for any $z \in (0, \frac{1}{4})$ and identify a widely-used signaling scheme, QPSK, to be second-order optimal.

For fading channels, we use a doubly-block fading model where the available bandwidth spans multiple coherence bandwidth. If we let W_c denote the coherence bandwidth, the total bandwidth of the channel is then assumed to $B = bW_c$ for some $b \geq 1$. Either a large b or a large W_c can lead to a large total bandwidth bW_c. However, these two regimes (the large b regime and the large W_c regime) can have very different channel behavior. Suppose we consider a wireless system with a total bandwidth of 10 MHz and if the delay spread is of the order of 1 $\mu s.$, then W_c would be of the order of 1 MHz and thus, b is of the order of 10. In this paper, we focus on such a system where the coherence bandwidth W_c is large and further, we assume a coherent channel model. By defining R_z to be a function of $1/W_c$, we calculate $R_z(0)$ and $\dot{R}_z(0)$. Further, similar to the AWGN case, for this channel model, we will show that QPSK can achieve both $R_z(0)$ and $\dot{R}_z(0)$ and is thus near-optimal. In the other case where b is large, it may not be appropriate to assume any form of channel side information (CSI) and thus a non-coherent channel model is more suitable. We refer the readers to [16] for first-order asymptotic results for MIMO channels in this regime.

This paper is organized as follows. In section 2, we specify the channel models and formulate the problem that we wish to study. In section 3, we present the main results for both AWGN channels and multipath fading channels. All the proofs will be presented in the Appendix. Section 4 contains concluding remarks and discussions.

2 Channel models and problem formulation

In this section, we will describe the channel models we use to study the behavior of both the AWGN channel and the multipath fading channel in the wideband regime. Further, we will rigorously formulate the problems we consider.

2.1 AWGN channels

We first consider a bandlimited AWGN channel with available bandwidth B :

$$y(t) = x(t) + w(t), \tag{2}$$

where $w(t)$ is a complex symmetric Gaussian random process. We assume that we have an input power constraint P (joules per second) for the channel (2). For notational convenience, we assume the noise power density $N_0/2 = 1/2$. Thus, the average power P also indicates the average SNR of the channel. We pass the channel output through a low-pass filter with bandwidth B and then sample the output of the filter at rate $1/B$. Thus, we can represent the channel as a discrete-time memoryless channel as follows:

$$y = x + w, \tag{3}$$

where w is a complex symmetric Gaussian random variable with variance 1, i.e., $w \in \mathcal{C}N(0,1)$. The power constraint for this discrete-time channel is

$$E\left(|x|^2\right) \leq \frac{P}{B}. \tag{4}$$

We want to study the asymptotic behavior of the communication rate R (nats per second) in terms of the available bandwidth B under this power constraint and an error exponent constraint, which is described below.

Let $P_e(N_t, R, P, B)$ be the minimum probability of decoding error for any block code with codeword length N seconds (or equivalently, $N_t B$ symbols) and coding rate R. The error exponent at communication rate R (also called *reliability function*) of this channel is defined as

$$E(R, P, B) = \lim_{N_t \to \infty} -\frac{\ln P_e(N_t, R, P, B)}{N_t}. \tag{5}$$

We desire a lower bound for $E(R, P, B)$ and denote it by Pz. (Without loss of generality, we scale the desired minimum value for the error exponent by P for mathematical convenience.) Let $R_z(1/B)$ denote the maximum possible rate at which communication is possible given this desired error exponent when the available bandwidth is B. Since $E(P, R, B)$ is a decreasing function of R, $R_z(1/B)$ is the solution to the equation

$$E(P, R, B) = Pz. \tag{6}$$

Our goals for AWGN channels are threefold:

(i) calculate $R_z(0)$ and $\dot{R}_z(0)$;

(ii) characterize the properties of *first-order optimal* signaling schemes, i.e., those that achieve $R_z(0)$; and

(iii) find *near-optimal* or *second-order optimal* signaling schemes in the wideband regime such that both $R_z(0)$ and $\dot{R}_z(0)$ can be achieved.

In the rest of the paper, we drop the subscript and simply refer to R_z as R. From the context, it should be clear that R is a function of z.

2.2 Coherent fading channels

To characterize a multi-path fading channel, we use a *doubly-block* Rayleigh fading model. Specifically, we assume block fading in both the time and frequency domains. Further, we assume that we have a rich-scattering environment such that all the fading gains are Gaussian distributed. We assume that the fading is fixed in each frequency-time block of duration T_c and bandwidth W_c, and independent from one block to another. In each block, we can transmit roughly W_cT_c symbols, from the *dimensionality theorem* [15]. We let $D = W_cT_c$ and refer to D as the *coherence dimension* of the channel.

For this channel model, we can represent the channel by

$$\mathbf{y}_l = H_l\mathbf{x}_l + \mathbf{w}_l, \quad 1 \leq l \leq b, \tag{7}$$

where $\mathbf{x}_l, \mathbf{y}_l, \mathbf{w}_l \in \mathcal{C}^D$. We also assume that all entries of $\mathbf{w}_l$ are i.i.d. complex Gaussian with distribution determined by $\mathcal{C}N(0,1)$. In other words, we have b parallel vector channels each with dimension D. Similar to the AWGN channel, we assume there is power constraint P (joules per second) for the fading channel, i.e., we have the following constraint on the input of the channel (7):

$$\sum_{l=1}^{b} E[\|\mathbf{x}_l\|^2] \leq PT_c. \tag{8}$$

The doubly-block fading model is a simple approximation of the physical multipath fading channel. However, it retains most of the important characteristics of channels in a fading environment. For a derivation of such a model, we refer the interested reader to [12]. This model has been used in [9] to achieve the lower bound for the optimal bandwidth where spreading still increases non-coherent channel capacity. In [6], Hajek and Subramanian use this model to calculate the reliability function and capacity for a non-coherent fading channel with a small peak constraint on the input signals. However, this model is simpler than the model used by Médard and Gallager [8], which allows correlation in both time and frequency blocks, or the model used Telatar and Tse [11], which allows correlation in frequency blocks.

In the wideband regime, we know the available bandwidth $bW_c >> 1$ and the energy available per degree of freedom is small, i.e., $\frac{P}{bW_c} << 1$. Obviously, a large bandwidth can be a result of either a large b or a large W_c.

However, b and W_c have different impacts on the channel performance and the asymptotic results in b and W_c can be very different from each other and can lead to different conclusions. In this paper, we will focus on the case where W_c is large. In this regime, we have large degrees of freedom in each coherence block although the energy per degree of freedom is small. Thus, we might still be able to measure the channel accurately and therefore, we assume a coherent fading channel model in this regime. However, to accurately illustrate the coherence level of this channel model from an error exponent point of view is still a research topic for now. We refer the reader to [18] for a discussion on the relationship between coherence level and coherence length from a capacity point of view.

The ergodic capacity of such channels under full receiver side CSI is well known and is determined by the following expression

$$C = bW_c E_H[\ln(1 + \frac{|H|^2 P}{bW_c})] \quad \textit{nats per second.} \tag{9}$$

The reliability function $E(R, P, W_c)$ of this channel can be defined as below

$$E(R, P, W_c) = \lim_{N \to \infty} -\frac{1}{T_c} \frac{\ln P_e(N, R, P, W_c)}{N}, \tag{10}$$

where $P_e(N, R, P, W_c)$ is the minimum probability of decoding error for all block codes with codeword length NT_c seconds and coding rate R (nats per second).

Let $R_z(1/W_c)$ denote the maximum possible rate at which communication is possible given this desired error exponent $E(R, P, W_c) \geq z$. Our goal in studying this channel model in the wideband regime is still the same as in the AWGN case: calculate both $R_z(0)$ and $\dot{R}_z(0)$ and identify signaling schemes that can achieve $R_z(0)$ and $\dot{R}_z(0)$.

3 Main results

In this section, we will present our main results for AWGN channels and coherent fading channels in two separate sections without proof. Due to the technical nature of the proofs, we will present them in the Appendix.

3.1 AWGN channels

We begin by first carefully describing the set of signaling schemes that we will consider in this paper. Due to the technicality in applying the sphere-packing bound, we only consider input distributions with a finite alphabet. In addition, for a given p, we restrict ourselves to the alphabet with a peak constraint defined below.

Definition 1 *Given positive constants α, K_m, p let $A(p)$ denote the set of finite subsets of complex numbers whose magnitude does not exceed $K_m p^{\alpha}$.*

In other words, we constrain the input such that the largest-magnitude symbol has to decrease as p decreases, although it can decrease at an arbitrarily slow rate. As we will show later, the choice of the parameters K_m and α are not relevant to the result. Thus, K_m can be an arbitrary large number and α can be an arbitrary small positive number. We refer to alphabet satisfying the conditions above to be $\mathcal{A}(p)$.

Definition 2 *Define*

$$\mathcal{D}(p) = \{q : E_q[|x|^2] = p; \text{support of q is an element in } \mathcal{A}(p).\}.$$

A signaling scheme is a sequence of input distributions, parameterized by B. For each B, we can only choose an input distribution from the set $\mathcal{D}(P/B)$.

Definition 3 *We define $\mathcal{F}(P)$ to be the set of signaling schemes, which are parameterized by B and satisfy*

$$\mathcal{F}(P) = \left\{ \{q_B\} : q_B \in \mathcal{D}(P/B) \right\}, \tag{11}$$

where $\mathcal{D}(P/B)$ is defined by Definition 2. ◇

By choosing signaling schemes from $\mathcal{F}(P)$, we are ruling out those *peaky* signaling schemes in which one of the input symbols remains constant or goes to ∞, while the average power per degree of freedom goes to 0.

In the following lemma, we state the well-known sphere packing and random coding bounds [3, 10] on the reliability function $E(R, P, B)$ defined by (5) for AWGN channels whose input alphabet is any fixed discrete set of points in $\mathcal{A}(P/B)$.

Lemma 1 *Consider the discrete-time additive Gaussian channel (3) with bandwidth B and input signaling schemes constrained by $\mathcal{F}(P)$. Then the reliability function for this channel satisfies*

$$E_r(R, P, B) \le E(R, P, B) \le E_{sp}(R, P, B), \tag{12}$$

with

$$\begin{aligned} E_r(R,P,B) &= \sup_{0 \le \rho \le 1} -\rho R + B E_o(P/B, \rho), && (13) \\ E_{sp}(R,P,B) &= \sup_{\rho \ge 0} -\rho R + B E_o(P/B, \rho), \\ E_o(P/B, \rho) &= \sup_{q \in \mathcal{D}(P/B)} \sup_{\beta \ge 0} -\ln \int \left(\int q(x) e^{\beta \left(|x|^2 - P/B \right)} f_w(y-x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy, && (14) \end{aligned}$$

where $f_w(x)$ is the probability density function of a complex Gaussian random variable $\mathcal{C}N(0,1)$. ◇

Remarks: Note there exists a *critical rate* R_{crit}, such that for $R \geq R_{crit}$, the sphere packing bound and the random-coding bound coincide with each other and thus the random-coding exponent (13) with (14) actually is the true reliability function. Thus, if we only focus on this rate region, by characterizing the asymptotic behavior of (13) when B is large, we get the asymptotic behavior of the reliability function. In the following theorem, we obtain closed-form expressions for $R(0)$ and $\dot{R}(0)$.

Theorem 1 *Consider the discrete-time additive Gaussian channel (3) with bandwidth B and input signaling schemes constrained by $\mathcal{F}(P)$. Let $R(1/B)$ be the maximum rate at which information can be transmitted on this channel such that the following error-exponent constraint is satisfied:*

$$E(R, P, B) \geq Pz, \quad 0 < z < \frac{1}{4}. \tag{15}$$

We have

$$R(0) = \lim_{B \to \infty} R(1/B) = P(1 - \sqrt{z})^2, \tag{16}$$

and

$$\dot{R}(0) = -\frac{P^2(1 - \sqrt{z})^3}{2}. \tag{17}$$

◇

Remarks: The constraint on z in (15) arises from the fact that the reliability function is only determined for a certain range of z. Outside this range, the random-coding exponent is not necessarily tight. As we will show later, $z = \frac{1}{4}$ is the error exponent for $R = R_{crit}$ in the infinite bandwidth limit. We now argue that for $0 < z < \frac{1}{4}$, when the bandwidth is sufficiently large, the solution $R(1/B)$ to (15) will exceed $R_{crit}(1/B)$ and thus, the error exponent at $R(1/B)$ is equal to the random-coding exponent. To be precise, we state this argument in the following lemma. It follows from this lemma that we can represent the reliability function by the random-coding exponent if we only consider $z < \frac{1}{4}$.

Lemma 2 *Let $R_r(1/B)$ be the solution to the random-coding exponent constraint $E_r(R, P, B) = Pz$, for a fixed $z \in (0, \frac{1}{4})$. For a fixed $z < \frac{1}{4}$, we must be able to find a $B_z < \infty$, such that for all $B > B_z$, $R(1/B) = R_r(1/B)$.*

Proof: Please refer to [17]. ◇

It should be noted that the constraints on the input signaling are not necessary to obtain the first-order result (16). In other words, introducing *peakiness* or allowing continuous alphabet symbols in the input distributions will not improve the error exponent in the infinite bandwidth limit for the AWGN channel. These constraints only play a role in obtaining the second-order terms in the expansion of $R_z(1/B)$ around $1/B = 0$.

A main goal of our study of the wideband reliability function here is to find good signaling schemes in the sense that they can achieve $R(0)$ and $\dot{R}(0)$. To do that, we first define *first-order optimality* and *near optimality* (or *second-order optimality*) formally of a signaling scheme in the wideband regime, in a similar way as in [14].

Definition 4 *Consider a signaling scheme* $\{q_B(\mathbf{x})\} \in \mathcal{F}(P)$ *parameterized by* B*. Let* $\tilde{R}(1/B)$ *be the solution of*

$$Pz = E(R, q_B, P, B) \tag{18}$$

where $E(R, q_B, P, B)$ *is the reliability function of the channel when the input distribution is fixed to be* q_B*. This signaling scheme is said to be first-order optimal with respect to the normalized error exponent* z*, if*

$$\tilde{R}(0) = R(0).$$

◇

Definition 5 *A signaling scheme* $\{q_B(\mathbf{x})\} \in \mathcal{F}(P)$ *is called second-order optimal or near optimal with respect to the normalized error exponent* z *if*

$$\tilde{R}(0) = R(0); \tag{19}$$

$$\dot{\tilde{R}}(0) = \dot{R}(0), \tag{20}$$

where $\tilde{R}(1/B)$ *is the solution to (18).* ◇

For AWGN channels, we obtain a sufficient condition for a signaling scheme to be first-order optimal. Then, we study the performance of two simple signaling schemes as in [14]: BPSK and QPSK. Specifically, when we say BPSK or QPSK, we mean the following. Let $p = P/B$ be the available power per degree of freedom. For BPSK, we choose the input to be either $\sqrt{p}$ or $-\sqrt{p}$ with equal probability; for QPSK, the input alphabet consists of $\sqrt{\frac{p}{2}}(1+j)$, $\sqrt{\frac{p}{2}}(1-j)$, $\sqrt{\frac{p}{2}}(-1+j)$, and $\sqrt{\frac{p}{2}}(-1-j)$, all chosen with equal probability as well.

Theorem 2 *For AWGN channels, all signaling schemes in* $\mathcal{F}(P)$ *which are symmetric around* 0 *are first-order optimal for any given* $z \in (0, \frac{1}{4})$*. Thus, both BPSK and QPSK are first-order optimal; however, only QPSK is second-order optimal.* ◇

Remarks: From this theorem, we know that it does not take much for a signaling scheme to be first-order optimal. This result is consistent with the capacity result shown by Massey in [7].

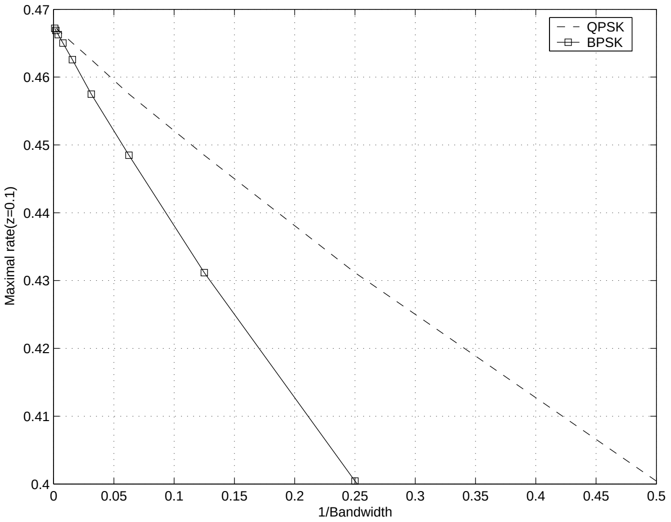

Figure 2: The maximal rate R for BPSK and QPSK for a fixed normalized error exponent $z = 0.1$.

To get a better feel for how differently BPSK and QPSK behave in the wideband regime, we plot R as a function of $1/B$ for both BPSK and QPSK in Figure 2. As shown in Figure 2, as $B \to \infty$, both BPSK and QPSK can achieve the optimal rate $R(0)$. However, only QPSK can achieve $\dot{R}(0)$.

Another way to understand the difference between the performance of BPSK and QPSK is to study the fundamental tradeoff between spectral efficiency and energy per information bit (E_b/N_0), as suggested in [14]. We plot this tradeoff in Figure 3. From this figure, we can see that both BPSK and QPSK can achieve the optimal $\frac{E_b}{N_0}_{min}$, however, only QPSK can achieve the optimal spectral efficiency slope at the point $\frac{E_b}{N_0}_{min}$.

As compared to Figure 2 in [14], the major difference here is that $\frac{E_b}{N_0}_{min}$ in Figure 3 is around $3.3dB$ higher, since we have a more stringent constraint than just reliable communications, as considered in [14]. $\frac{E_b}{N_0}_{min}$ here denotes the minimal energy per information bit such that the probability of error has to decay faster than e^{-Nz} as the codeword length N increases.

3.2 Coherent fading channels

Next, we consider coherent fading channels. As in the case of the AWGN channel, we first describe our assumptions on the input signaling schemes.

Definition 6 *Define $\mathcal{Q}^b_{W_c}(P)$ to be the set of joint input distributions on $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_b)$, where $\{\mathbf{x}_l, \ \ l = 1, 2, \cdots, b\}$ are vectors with dimension $D = W_c T_c$, which satisfy the following*

1. *the average power constraint (8) is satisfied;*

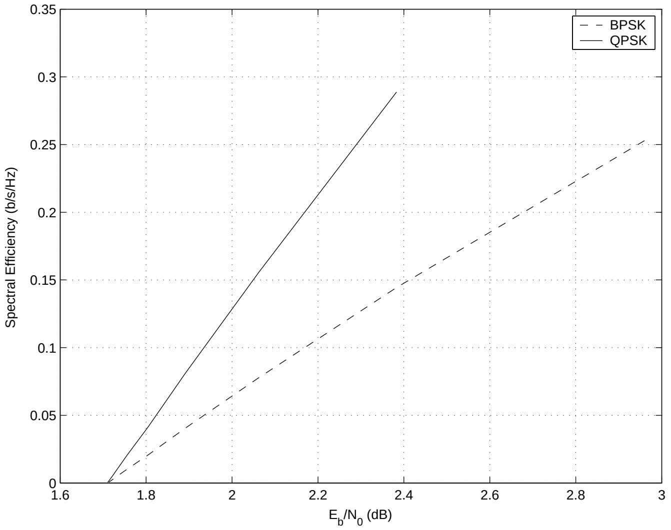

Figure 3: Spectral efficiencies achieved by QPSK and BPSK in the AWGN channel, when the error exponent is constrained by $z = 0.1$.

2. *the distribution has a discrete alphabet, consisting of finite number of symbols;*
3. *each symbol can be chosen from a given set $\mathcal{S}^b_{W_c}$. The set of symbols $\mathcal{S}^b_{W_c}$ is defined as follows:*

$$\mathcal{S}^b_{W_c} = \{\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_b\} : \mathbf{x}_l \in C^D; \max_{d=1,2,\cdots D} |x_{ld}| \le K_m W_c^{-\alpha} \ \forall l = 1, 2, \cdots, b\}, \qquad (21)$$

where K_m and α are allowed to be any positive constants independent of W_c. ◇

The signaling schemes of interest to us are defined as follows.

Definition 7 *We define $\mathcal{F}^b_{W_c}(P)$ to be the set of signaling schemes, which are parameterized by W_c and satisfy*

$$\mathcal{F}^b_{W_c}(P) = \left\{ \{q_{W_c}\} : q_{W_c} \in \mathcal{Q}^b_{W_c}(P) \right\}, \qquad (22)$$

where $\mathcal{Q}^b_{W_c}(P)$ was defined in Definition 6. ◇

The reliability function for our discrete-time channel model (7) is upper and lower bounded as stated in the following lemma.

Lemma 3 *Consider the coherent fading channel model (7) (available bandwidth bW_c) with H known at the receiver. Assume that the input distribution satisfies the average power constraint (4) and the constraint in in $\mathcal{F}^b_{W_c}(P)$. The reliability function $E(R, P, W_c)$ satisfies*

$$E_r(R, P, W_c) \le E(R, P, W_c) \le E_{sp}(R, P, W_c),$$

with

$$\begin{aligned} E_r(R,P,W_c) &= \sup_{0\le\rho\le 1} -\rho R + E_o(P,\rho,W_c), \\ E_{sp}(R,P,W_c) &= \sup_{\rho\ge 0} -\rho R + E_o(P,\rho,D), \\ E_o(P,\rho,W_c) &= \sup_{q\in\mathcal{F}^b_{W_c}(P)} \sup_{\beta\ge 0} -\frac{1}{T_c}\ln E_H \int\left(\int q(\mathbf{X})e^{\beta(\|\mathbf{X}\|^2-PT_c)} f(\mathbf{Y}|\mathbf{X},\mathbf{H})^{\frac{1}{1+\rho}}d\mathbf{X}\right)^{1+\rho} d\mathbf{Y}. \quad (23) \end{aligned}$$

Proof: We can apply the random-coding and sphere-packing exponents [10] to this channel model by viewing the channel as a memoryless channel with output $\hat{\mathbf{Y}} = \{\mathbf{Y},\mathbf{H}\}$. The term $\frac{1}{T_c}$ in (23) is to account for the fact that the rate R here is defined to be nats per second, rather than nats per discrete time slot. ◇

The constraint on the error exponent is

$$E(R,P,W_c) \ge Pz, \tag{24}$$

and we need to solve for $R(0)$ and $\dot{R}(0)$ where R is a function for $\frac{1}{W_c}$ for a fixed b. We have the following theorem.

Theorem 3 *Consider a coherent Rayleigh-fading vector channel (7) with the input signaling constrained by* $\mathcal{F}^b_{W_c}(P)$*. Let* $R(1/W_c)$ *be the maximum rate at which information can be transmitted on this channel such that the following error-exponent constraint is satisfied:*

$$E(R,P,W_c) \ge Pz, \quad 0 < z < z^*, \tag{25}$$

where z^* *is defined as follows*

$$z^* = \frac{b}{PT_c}\ln(1+\frac{PT_c}{2b}) - \frac{1}{4+2PT_c/b}. \tag{26}$$

We have

$$R(0) = \lim_{W_c\to\infty} R_b(1/W_c) = \sup_{0\le\rho\le 1} -\frac{Pz}{\rho} + \frac{1}{T_c}\frac{b\ln\left(1+\frac{\rho PT_c}{b(1+\rho)}\right)}{\rho}, \tag{27}$$

and

$$\dot{R}(0) = -\frac{P^2}{b(1+\rho^*)(1+\rho^*+\frac{\rho^* PT_c}{b})^2}, \tag{28}$$

where ρ^* *is the optimizing* ρ *in (27).* ◇

The constraint on z in (25) again comes from the fact that the reliability function is only known when $R \ge R_{crit}$. Now we show that z^* given by (26) is the corresponding error exponent at R_{crit} when W_c goes to infinity.

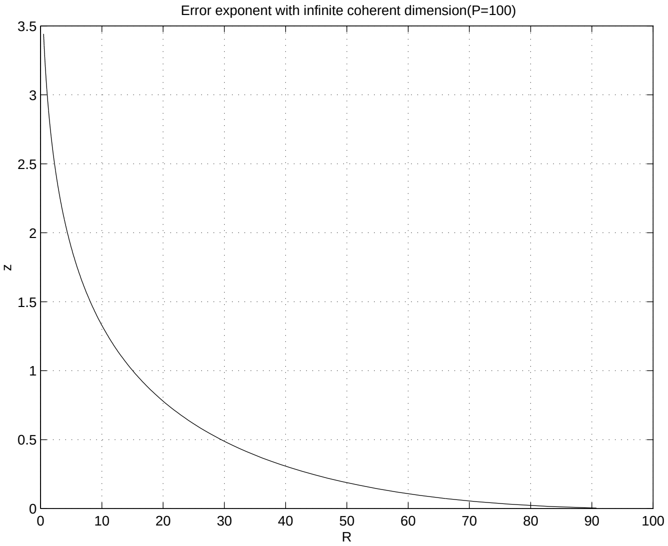

Figure 4: The error exponent curve from R_{crit} to capacity for the channel with infinite coherence dimension. $b = T_c = 1$. $P = 100$.

From the property of the critical rate R_{crit}, we know the optimizing ρ in (27) at the corresponding error exponent z_{crit} is 1. Thus, taking derivative of the right side of (27) with respect to ρ, we must have

$$\frac{z_{crit}}{\rho^2} - \frac{b}{T_c}\frac{\ln(1+\frac{\rho PT_c}{b(1+\rho)})}{\rho^2} + \frac{b}{T_c}\frac{PT_c/b}{\rho(1+\frac{\rho PT_c}{b(1+\rho)})}\frac{1}{(1+\rho)^2}\Bigg|_{\rho=1} = 0.$$

By solving this, it is straightforward to have $z_{crit} = z^*$ with z^* determined by (26). The corresponding rate R_{crit} can be obtained as follows

$$\begin{aligned} R_{crit} &= -z_{crit} + \frac{b}{T_c}\ln(1+\frac{PT_c}{2b}) \\ &= \frac{P}{4+2\frac{PT_c}{b}}. \end{aligned}$$

Using a similar argument as in the AWGN channel case, we can argue that for $z \in (0, z^*)$, the reliability function coincides with the random-coding exponent for sufficiently large W_c . Thus, the calculation of $R(0)$ and $\dot{R}(0)$ can be carried out by using the random-coding exponent.

Another observation here is that the applicable region (in terms of R), where the random-coding exponent coincides with the sphere-packing exponent, actually covers most of the rate region from 0 to capacity, when the available energy per coherence block $\frac{PT_c}{b}$ is fairly large. To see this, we first notice that as W_c goes to infinity, our capacity C_∞ in (9) is P. Thus, the critical rate R_{crit} can be also written as $\frac{1}{4+2\frac{PT_c}{b}}C_\infty$. When $\frac{PT_c}{b}$ is large, we have $R_{crit} << C_\infty$. This observation is also shown in Figure 4. For simplicity, we choose $b = T_c = 1$ in this numerical example and choose $P = 100$.

Next, we need to identify those signaling schemes which can achieve $R(0)$ and $\dot{R}(0)$. Again, we consider BPSK and QPSK signaling. However, for the fading channel (7), these two signaling schemes have slightly different meanings than what we defined in last section for AWGN channels. Specifically, for both BPSK and QPSK, we spread the available power in each coherent block equally among all the time-frequency coherent blocks and make the distributions in each dimension i.i.d. For BPSK, the symbols for each dimension are $\sqrt{P/bW_c}$ and $-\sqrt{P/bW_c}$, with equal probability. For QPSK, the symbols are $\sqrt{\frac{P}{2bW_c}}(1+j)$, $\sqrt{\frac{P}{2bW_c}}(1-j)$, $\sqrt{\frac{P}{2bW_c}}(-1+j)$ and $\sqrt{\frac{P}{2bW_c}}(-1-j)$. Similar to the AWGN case, we have

Theorem 4 *Both BPSK and QPSK are first-order optimal for any given* $z \in (0, z^*)$*; however, only QPSK is second-order optimal.* $\diamond$

3.3 Implications and discussion

The results that we have obtained for both AWGN channels and coherent fading channels are consistent with the results from a capacity point of view in the seminal work [14]. By letting z go to 0, the quantity R_z becomes the capacity of the channel. Thus, it can be easily checked that by taking z to be 0, we can recover the capacity results by using the expressions in Theorem 1 and Theorem 3. However, we also have to point out that in [14], a very general treatment is provided for a much broader class of channel models. In this paper, due to the complexity of the calculation of the reliability function, we only calculated the first and second order rate approximation for two very specific channel models.

Despite the similarity between our results and Verdu's results regarding near-optimal signaling, the fact that QPSK is *still* near-optimal under a certain error exponent constraint is still somewhat surprising because of the following reason. In general, very little is known about the conditions under which an input distribution achieves the optimal error exponent at a given rate, even in the infinite bandwidth limit. It is not necessarily true that capacity-achieving distributions are also optimal from an error-exponent point of view. One example is the infinite-bandwidth non-coherent Rayleigh fading channel, which is studied in [16]. Thus, it is not obvious that actually QPSK can do well in the wideband regime from an error exponent point of view,even though it is wideband optimal from a capacity point of view.

4 Conclusions

In this paper, we have studied the maximum rate at which information transmission is possible in additive Gaussian noise channels and coherent fading channels, for a given error exponent in the wideband regime. Given a desired error exponent, our main contribution is the calculation of the above rate and its derivative in the limit when the available bandwidth goes to ∞. For fading channels, we focus on the case when the coherence bandwidth W_c is large. This also leads to a notion of near-optimality of input distributions, where a sequence of distributions is defined to be near-optimal if it achieves both the rate and its derivative in the infinite bandwidth limit. As in [14], we show that for both AWGN and coherent fading channels, while QPSK is near-optimal, BPSK is not.

This result is surprising to some extent. Generally, it is not well-understood as to what signaling scheme is optimal, i.e., given a coding rate, it is difficult to find the input distribution that gives the smallest probability of decoding error. In this paper, we consider the problem from an alternate point of view, we fix a given error exponent, and consider optimal signaling schemes that gives the largest communication rate. The capacity-achieving schemes, which corresponds to zero error exponent, are not necessarily the best schemes from the error exponent point of view. However, the results in this paper tell us, in the wideband regime, QPSK is near-optimal with respect to a nonzero error exponent just as it is near-optimal for the capacity case for both AWGN and coherent fading channels. Thus, it can not only achieves capacity, but also achieves the the best probability of decoding error, in the wideband regime.

A Proof of Theorem 1 and Theorem 2

Due to the technical nature of the calculations needed in the proofs of our main results, we first summarize the key steps of the proof to help the reader follow the proof of our main results.

The proof of Theorem 1 can be broken down into the following major steps:

1. We first relate the problem of finding $R(0)$ and $\dot{R}(0)$, where R is the communication rate per second as a function of $1/B$, to the problem of finding $\dot{r}(0)$ and $\ddot{r}(0)$, where r is the communication rate per degree of freedom in (3) as a function of p, where p denotes the SNR per degree of freedom.

2. The calculation of $\dot{r}(0)$ can be related to the optimal value for E_o in the infinite bandwidth limit; an upper bound is derived for E_o using a simple inequality; this bound is further shown to be achievable;

3. $\ddot{r}(0)$ can also be related to certain derivatives of E_o; a better upper bound is derived for E_o which yields an upper bound for $\ddot{r}(0)$; this bound is also shown to be achievable.

The next several subsections will prove the main results following these three steps.

A.1 Communication rate and error exponent per degree of freedom

It is shown in [14] that the capacity C in a bandlimited channel with limited available power P, but large available bandwidth b, can be related to the capacity c in a scalar channel with small available power $p = P/B$. Thus, the problem of finding optimal $C(0)$ and $\dot{C}(0)$ can be shown to be equivalent to the problem of finding optimal $\dot{c}(0)$ and $\ddot{c}(0)$. The relationship between $C(0)$ and $\dot{c}(0)$ is also extensively studied in an earlier paper [13], where the notion *capacity per unit cost* was studied. We first show that a similar connection can be made between the error-exponent constrained rates R (nats per second) and r (nats per symbol).

Theorem 5 *Consider a scalar Gaussian channel $y = x + w$ with average power constraint p. Further, the signaling schemes are constrained by $\tilde{\mathcal{F}}(p) = \{\{q_p(x)\} : q_p(x) \in \mathcal{D}(p)\}$. Let r be the maximum rate per symbol at which information can be transmitted through channel (3) such that the error exponent satisfies*

$$\hat{E}(r,p) \geq pz, \quad 0 < z < \frac{1}{4},$$

where $\hat{E}(r,p)$ is the error exponent per symbol of the scalar channel with power constraint p. Consider r as a function of p. Let R (nats per second) be defined as the solution to (15). We have

$$\begin{aligned} R(0) &= P\dot{r}(0); \\ \dot{R}(0) &= \frac{P^2\ddot{r}(0)}{2}. \end{aligned}$$

Proof: Please refer to [17]. ◇

Thus, the original problem of finding $R(0)$ and $\dot{R}(0)$ in the wideband regime is equivalent to finding the optimal values for $\dot{r}(0)$ and $\ddot{r}(0)$, given a constraint on the reliability function $\hat{E}(r,p) \geq pz$. In the rest of this proof, we will deal with this scalar channel problem. For notational convenience, we use $E(r,p)$ to denote the error exponent per symbol of the single channel instead of using $\hat{E}(r,p)$.

A.2 Optimal value of $\dot{r}(0)$ and first-order optimal signaling schemes

We know for the error-exponent constraint in the range of $(0, \frac{1}{4})$ and p sufficiently small, we have

$$E(r,p) = E_r(r,p) = \sup_{0\leq\rho\leq 1} -\rho r + E_o(p,\rho),$$

where

$$E_o(p,\rho) = \sup_{q_p \in \tilde{\mathcal{F}}(p)} \sup_{\beta \geq 0} - \ln \int \left(\int q_p(x) e^{\beta(|x|^2 - p)} f_w(y-x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy. \tag{29}$$

Thus, the constraint on the error exponent can also be written as

$$pz = \sup_{0 \leq \rho \leq 1} -\rho r + E_o(p,\rho). \tag{30}$$

With some effort, we can further show that this constraint is equivalent to the following expression

$$\frac{r}{p} = \sup_{0 \leq \rho \leq 1} -\frac{z}{\rho} + \frac{E_o(p,\rho)}{p\rho}. \tag{31}$$

Now we prove the following lemma:

Lemma 4 *For* $0 < z < \frac{1}{4}$,

$$\dot{r}(0) = (1 - \sqrt{z})^2. \tag{32}$$

Proof: This result can be calculated by using Gallager's calculation in [5] and some additional careful manipulations. Please refer to [17] for a complete proof. ◇

Next we study conditions for a sequence of input distributions to be first-order optimal. It is straightforward to see that

Lemma 5 *Assuming* $0 < z < \frac{1}{4}$*, a sufficient condition for* $\{q_p\} \in \tilde{\mathcal{F}}(p)$ *to be first-order optimal is that*

$$\lim_{p \to 0} \frac{\tilde{E}_o(p, q_p, \rho^*)}{p} = \frac{\rho^*}{1+\rho^*}, \tag{33}$$

where

$$\tilde{E}_o(p, q_p, \rho) = \sup_{\beta \geq 0} - \ln \int \left(\int q_p(x) e^{\beta(|x|^2 - p)} f_w(y-x)^{\frac{1}{1+\rho}} dx \right)^{1+\rho} dy$$

and $\rho^* = \frac{\sqrt{z}}{1-\sqrt{z}}$.

Proof: To show that (33) is sufficient, it suffices to show that given a signaling scheme q_p, the corresponding rate $\tilde{r(p)}$ given an error exponent constraint $\tilde{E}_o(p, q_p, \rho) \geq pz$ satisfies

$$\lim_{p \to 0} \frac{\tilde{r}(p)}{p} = (1 - \sqrt{z})^2.$$

If $\lim_{p \to 0} \frac{\tilde{E}_o(p, q_p, \rho^*)}{p} = \frac{\rho^*}{1+\rho^*}$, we have

$$\begin{aligned}
\liminf_{p \to 0} \frac{\tilde{r}}{p} &\geq \liminf_{p \to 0} -\frac{z}{\rho^*} + \frac{\tilde{E}_o(p, q_p, \rho^*)}{p\rho^*} \\
&= -\frac{z}{\rho^*} + \lim_{p \to 0} \frac{\tilde{E}_o(p, q_p, \rho^*)}{p\rho^*} \\
&= -\frac{z}{\rho^*} + \frac{1}{1+\rho^*} \\
&= (1 - \sqrt{z})^2.
\end{aligned}$$

On the other hand, we know that

$$\begin{aligned}
\limsup_{p\to 0}\frac{\tilde{r}}{p} &= \limsup_{p\to 0}\sup_{0\le\rho\le 1} -\frac{z}{\rho}+\frac{\tilde{E}_o(p,q_p,\rho)}{p\rho}\\
&\le \limsup_{p\to 0}\sup_{0\le\rho\le 1} -\frac{z}{\rho}+\frac{E_o(p,\rho)}{p\rho}\\
&\le \limsup_{p\to 0}\sup_{0\le\rho\le 1} -\frac{z}{\rho}+\frac{1}{1+\rho}\\
&= (1-\sqrt{z})^2.
\end{aligned}$$

Thus, the limit of $\frac{\tilde{r}}{p}$ exists and we have

$$\dot{\tilde{r}}(0)=\lim_{p\to 0}\frac{\tilde{r}}{p}=(1-\sqrt{z})^2.$$

◇

Actually, it does not take much to be first-order optimal.

Lemma 6 *For a fixed* $0 < z < \frac{1}{4}$, *a sequence of input distribution* $q_p \in \tilde{\mathcal{F}}(p)$ *is first-order optimal if it is symmetric around* 0.

Proof: To show this, we need to check (33) for a sequence of mean-zero input distribution $q_p \in \tilde{\mathcal{F}}(p)$. Since it is always true that

$$\limsup_{p\to 0}\frac{\tilde{E}_o(p,q_p,\rho^*)}{p}\le\frac{\rho^*}{1+\rho^*},$$

it suffices to show that

$$\liminf_{p\to 0}\frac{\tilde{E}_o(p,q_p,\rho^*)}{p}\ge\frac{\rho^*}{1+\rho^*}.$$

Note

$$\begin{aligned}
\tilde{E}_o(p,q_p,\rho^*) &= \sup_{\beta\ge 0} -\ln\int\alpha(y)^{1+\rho^*}dy\\
&= \sup_{\beta\ge 0} -\ln\int f_w(y)(1+T(y))^{1+\rho^*}dy,
\end{aligned}$$

where $\alpha(y)$ and $T(y)$ are defined below:

$$\alpha(y)=\int q_p(x)e^{\beta(|x|^2-p)}f_w(y-x)^{\frac{1}{1+\rho}}dx; \tag{34}$$

$$T(y)=\int q_p(x)e^{\beta(|x|^2-p)}\left[\frac{f_w(y-x)}{f(y|0)}\right]^{\frac{1}{1+\rho}}dx-1. \tag{35}$$

To achieve a lower bound, we choose $\beta = \theta = \frac{\rho^*}{1+\rho^*}$. Further, we use the following inequality

$$(1+t)^{1+\rho^*} \leq 1 + (1+\rho^*)t + \frac{\rho^*(1+\rho^*)}{2}t^2.$$

This leads to

$$\tilde{E}_o(p, q_p, \rho^*) \geq -\ln \int f_w(y)(1 + (1+\rho^*)T(y) + \frac{\rho^*(1+\rho^*)}{2}T^2(y))dy.$$

When $\beta = \theta$, it can be shown that

$$\int f_w(y)(1 + (1+\rho^*)T(y))dy = -\rho^* + (1+\rho^*)e^{-\theta p},$$

and

$$\int f_w(y)T^2(y)dy = 1 - 2e^{-\theta p} + E\left[e^{\frac{2Re(x_1 x_2^*)}{(1+\rho^*)^2}}\right]e^{-2\theta p},$$

where x_1 and x_2 are i.i.d. random variables distributed according to $q_p(x)$.

Next we claim

$$\lim_{p\to 0} \frac{\int f_w(y)T^2(y)dy}{p} = 0.$$

Since $\lim_{p\to 0} \frac{(1-e^{-\theta p})^2}{p} = 0$, it suffices to show

$$\lim_{p\to 0} \frac{E\left[e^{\frac{2Re(x_1 x_2^*)}{(1+\rho^*)^2}}\right] - 1}{p} = 0.$$

This can be easily shown (refer to [17]) by using the assumption that $q_p(x)$ is symmetric around 0 and

$$|x|_{max} < K_m p^{\alpha}.$$

Thus, we have

$$\begin{aligned}
\liminf_{p\to 0} \frac{\tilde{E}_o(p, q_p, \rho^*)}{p} &\geq \liminf_{p\to 0} \frac{-\ln(-\rho^* + (1+\rho^*)e^{-\theta p} + o(p))}{p} \\
&= \liminf_{p\to 0} \frac{-\ln(1 - \frac{\rho^* p}{1+\rho^*} + o(p))}{p} \\
&= \frac{\rho^*}{1+\rho^*}.
\end{aligned}$$

◇

A.3 The optimal value of $\ddot{r}(0)$

In this section, we compute the value of $\ddot{r}(0)$. To do this, we first connect $\ddot{r}(0)$ to the second partial derivative of $E_o(p,\rho)$ with respect to p.

Theorem 6 *Assume the second partial derivative of $E_o(p,\rho)$ with respect to p at $p=0$ (denoted as $\ddot{E}_o(0,\rho)$) exists for any $\rho \in [0,1]$. Further, assume that*

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} \to \frac{\ddot{E}_o(0,\rho)}{2\rho} \quad \textit{uniformly for } \rho \in [0,1],$$

and $\frac{\ddot{E}_o(0,\rho)}{\rho}$ is a continuous and bounded function of ρ for $\rho \in [0,1]$. Then $\ddot{r}(0)$ can be determined by

$$\ddot{r}(0) = \frac{\ddot{E}_o(0,\rho^*)}{\rho^*}, \tag{36}$$

where $\rho^* = \frac{\sqrt{z}}{1-\sqrt{z}}$.

Proof: First we show that

$$\ddot{\bar{r}}(0) = \limsup_{p\to 0} \frac{r(p) - p\dot{r}(0)}{p^2/2} \leq \frac{\ddot{E}_o(0,\rho^*)}{\rho^*}.$$

The uniform convergence gives us: for any $\epsilon > 0$, we can find $\eta(\epsilon)$ such that for all $p < \eta(\epsilon)$,

$$\left| \frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} - \frac{\ddot{E}_o(0,\rho)}{2\rho} \right| < \epsilon \quad \textit{for all } \rho \in [0,1].$$

In other words, for $p < \eta(\epsilon)$, we can write

$$E_o(p,\rho) \leq \dot{E}_o(0,\rho)p + \ddot{E}_o(0,\rho)p^2/2 + \rho\epsilon p^2.$$

It is easy to see that

$$r(p) \leq \sup_{0\leq\rho\leq 1} -\frac{pz}{\rho} + \frac{\dot{E}_o(0,\rho)p + \ddot{E}_o(0,\rho)p^2/2}{\rho} + \epsilon p^2. \tag{37}$$

Assume $\rho(p)$ is the optimizing ρ for (37). From the first-order calculation, we already know that

$$\dot{E}_o(0,\rho) = \frac{\rho}{1+\rho}.$$

Since the optimization in (37) is performed over a compact set $[0,1]$ and by assumption $\ddot{E}_o(0,\rho)$ is continuous in ρ, the optimizing ρ must exist.

We must have

$$r(p) \quad \leq \quad \left\{ \sup_{0\leq\rho\leq 1} -\frac{pz}{\rho} + \frac{p\dot{E}_o(0,\rho)}{\rho} \right\} + \frac{\ddot{E}_o(0,\rho(p))\frac{p^2}{2}}{\rho(p)} + \epsilon p^2.$$

We know that

$$\dot{r}(0)p = \sup_{0\leq\rho\leq 1} -\frac{pz}{\rho} + \frac{p\dot{E}_o(0,\rho)}{\rho}.$$

This gives us

$$\frac{r(p)-p\dot{r}(0)}{p^2/2} \leq \frac{\ddot{E}_o(0,\rho(p))}{\rho(p)} + 2\epsilon.$$

Letting ϵ go to 0, we have

$$\begin{aligned} \overline{\ddot{r}}(0) &= \limsup_{p\to 0} \frac{r(p)-p\dot{r}(0)}{p^2/2} \\ &\leq \limsup_{p\to 0} \frac{\ddot{E}_o(0,\rho(p))}{\rho(p)} \\ &= \frac{\ddot{E}_o(0,\rho^*)}{\rho^*}, \end{aligned} \tag{38}$$

where ρ^* is the optimizing ρ of (37) as p goes to zero, and can be shown to be equal to $\frac{\sqrt{z}}{1-\sqrt{z}}$. The last equation (38) can be easily verified given that $\frac{\ddot{E}_o(0,\rho)}{\rho}$ is a continuous function of ρ, if we have $\lim_{p\to 0}\rho(p)=\rho^*$. We refer the readers to [17] for proof of this fact.

To complete the proof of the theorem, it suffices to show

$$\underline{\ddot{r}}(0) = \liminf_{p\to 0} \frac{r(p)-p\dot{r}(0)}{p^2/2} \geq \frac{\ddot{E}_o(0,\rho^*)}{\rho^*}.$$

To see this, we choose $\rho=\rho^*$ in (37) and we have

$$r(p) \geq -\frac{pz}{\rho^*} + \frac{p\dot{E}_o(0,\rho^*)}{\rho^*} + \frac{\ddot{E}_o(0,\rho^*)\frac{p^2}{2}}{\rho^*} - \epsilon p^2.$$

It is easy to see that

$$\dot{r}(0) = -\frac{z}{\rho^*} + \frac{\dot{E}_o(0,\rho^*)}{\rho^*},$$

and thus, we have

$$\frac{r(p)-p\dot{r}(0)}{p^2/2} \geq \frac{\ddot{E}_o(0,\rho^*)}{\rho^*} - 2\frac{\epsilon p^2}{p^2}.$$

Letting $p\to 0$, we will have

$$\underline{\ddot{r}}(0) \geq \frac{\ddot{E}_o(0,\rho^*)}{\rho^*}.$$

◇

Thus, to obtain the optimal value for $\ddot{r}(0)$, we need to verify the uniform convergence assumption in Theorem 6 and calculate $\frac{\ddot{E}_o(0,\rho^*)}{\rho^*}$, which is established in the next theorem.

Theorem 7

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} \to -\frac{1}{2(1+\rho)^3} \quad \textit{uniformly for } \rho \in [0,1], \tag{39}$$

as p goes to 0.

Proof: To show uniform convergence, we both upper and lower bound

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p}$$

by a function of ρ plus a small term $\delta(1)$, which converges to 0 uniformly for $\rho \in [0,1]$, as p goes to 0. Specifically, we want to show that when p is small, we have

$$\frac{\ddot{E}_o(0,\rho)}{2\rho} + \delta_1(1) \leq \frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} \leq \frac{\ddot{E}_o(0,\rho)}{2\rho} + \delta_2(1),$$

where both $\delta_1(1)$ and $\delta_2(1)$ converge to 0 uniformly as p goes to 0. The uniform convergence of

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p}$$

follows easily from here. We will first show an upper bound, then we will obtain a lower bound by using QPSK signaling at the input. Throughout this proof, we will use the notation $\delta(p^m)$ to denote a term satisfying that as p goes to 0, $\frac{\delta(p^m)}{p^m} \to 0$ uniformly for $\rho \in [0,1]$.

We know that

$$E_o(p,\rho) = \sup_{\{q_p\} \in \tilde{\mathcal{F}}(p)} \tilde{E}_o(p, q_p, \rho).$$

However, it is easy to see that we will not lose any optimality if we constraint ourselves to those input distributions which perform at least as good as QPSK. In other words, we have

$$E_o(p,\rho) = \sup_{\{q_p\} \in \tilde{\mathcal{G}}(p)} \tilde{E}_o(p, q_p, \rho), \tag{40}$$

where $\tilde{\mathcal{G}}(p)$ is defined as

$$\tilde{\mathcal{G}}(p) = \Big\{ \{q_p\} \in \tilde{\mathcal{F}}(p) : \tilde{E}_o(p, q_p, \rho) \geq \tilde{E}_o(p, QPSK, \rho), \forall p > 0 \Big\} \tag{41}$$

First we claim that: for any sequence of input distributions $\{q_p(x)\} \in \tilde{\mathcal{G}}(p)$,

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} \leq \frac{-\inf_{\{q_p\} \in \tilde{\mathcal{G}}(p)} \inf_{\beta \geq 0} \int \alpha(y)^{1+\rho} dy + e^{-\frac{\rho p}{1+\rho}}}{\rho p^2}. \tag{42}$$

This claim can be proved by the following series of inequalities:

$$\begin{aligned}
&\frac{\frac{E_o(p,\rho)}{p\rho}-\frac{\dot{E}_o(0,\rho)}{\rho}}{p}\\
=\;&\frac{-\ln\left(\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy\right)-\frac{\rho p}{1+\rho}}{\rho p^2}\\
=\;&\frac{-\ln\left(e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy\right)}{\rho p^2}\\
=\;&\frac{\ln\frac{1}{e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy}}{\rho p^2}\\
\leq\;&\frac{-1+\frac{1}{e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy}}{\rho p^2}\\
=\;&\frac{-e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy+1}{\rho p^2}\frac{1}{e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy}\\
\leq\;&\frac{-e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy+1}{\rho p^2 e^{\frac{\rho p}{1+\rho}}}\qquad(43)\\
=\;&\frac{-\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy+e^{-\frac{\rho p}{1+\rho}}}{\rho p^2}.
\end{aligned}$$

The inequality (43) is true because that the first order results imply

$$\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy=e^{-E_o(p,\rho)}\geq e^{-\frac{\rho p}{1+\rho}},$$

which leads to

$$-e^{\frac{\rho p}{1+\rho}}\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy+1\leq 0.$$

On the other hand,

$$\begin{aligned}
\inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\inf_{\beta\geq 0}\int\alpha(y)^{1+\rho}dy &\leq \inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\int\alpha(y)^{1+\rho}dy|_{\beta=0}\\
&= \inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\int\left(\sum_k q_k f(y|x_k)^{\frac{1}{1+\rho}}\right)^{1+\rho}dy\\
&\leq \inf_{\{q_p\}\in\tilde{\mathcal{G}}(p)}\int\left(\sum_k q_k f(y|x_k)\right)dy\\
&= 1.
\end{aligned}$$

These two bounds together give us (43).

Next, we further bound $\int \alpha(y)^{1+\rho} dy$ for any sequence of input distributions $\{q_p\} \in \tilde{\mathcal{G}}(p)$. We claim the following: for all $q_p(x)$ and all β, we have

$$\begin{aligned} \int \alpha(y)^{1+\rho} dy &= \int f_w(y)(1+T(y))^{1+\rho} dy \\ &\geq 1 + (1+\rho)\int f_w(y)T(y)dy + \frac{\rho(1+\rho)}{2}\int f_w(y)T^2(y)dy + \frac{\rho(1+\rho)(\rho-1)}{6}\int f_w(y)T^3(y)dy. \end{aligned} \tag{44}$$

This claim is quite obvious by noticing that the following inequality is true for all $t \geq -1$ and all $\rho \in [0, 1]$:

$$(1+t)^{1+\rho} \geq 1 + (1+\rho)t + \frac{\rho(1+\rho)}{2}t^2 + \frac{\rho(1+\rho)(\rho-1)}{6}t^3.$$

We will now treat the three terms in (44), which involves $T(y)$, separately and find a bound for each of them. For the first term, we have

$$\begin{aligned} \int f_w(y)T(y)dy &= \int f_w(y)T(y)dy && (45) \\ &= E[e^{\beta(|x|^2-p)}e^{-\theta|x|^2}] - 1 && (46) \\ &\geq e^{-\theta p} - 1, && (47) \end{aligned}$$

where Eq. (47) is obtained by applying Jensen's inequality.

For the second term, we claim that: for any input distribution $\{q_p(x)\} \in \tilde{\mathcal{G}}(p)$, let β^* be the optimizing β, which maximizes

$$\sup_{\beta \geq 0} - \ln \int \alpha(y)^{1+\rho} dy. \tag{48}$$

We have

$$\int f_w(y)T^2(y)dy\bigg|_{\beta=\beta^*} \geq \theta^2 p^2 + \frac{p^2}{(1+\rho)^4} + \delta(p^2).$$

We also claim that, for those input distributions in $\tilde{\mathcal{G}}(p)$, the term with integral over $T^3(y)$ actually does not contribute anything to the second-order calculation, i.e.,

$$\int f_w(y)T^3(y)dy\bigg|_{\beta=\beta^*} = \delta(p^2).$$

for any $\{q_p(x)\} \in \tilde{\mathcal{G}}(p)$. We do not include the calculation regarding the two claims in this paper. For interested users, please refer to [17] for a complete proof.

From here, we have

$$\begin{aligned} \int \alpha(y)^{1+\rho} dy &\geq 1 - \frac{\rho}{1+\rho}p + (1+\rho)\theta^2 p^2/2 + \frac{\rho(1+\rho)}{2}\left(\theta^2 p^2 + \frac{p^2}{(1+\rho)^4}\right) + \rho\delta(p^2) \\ &= 1 - \frac{\rho}{1+\rho}p + \frac{\rho^2 p^2}{2(1+\rho)^2} + \frac{\rho p^2}{2(1+\rho)^3} + \rho\delta(p^2). \end{aligned}$$

and thus

$$\frac{\frac{E_o(p,\rho)}{p\rho} - \frac{\dot{E}_o(0,\rho)}{\rho}}{p} \leq -\frac{1}{2(1+\rho)^3} + \delta(p^2).$$

Later, we will show that by choosing the input distribution to be QPSK, we can establish a lower bound which has the same expression as the upper bound. Thus, we know (39) is true. ◇

Since we know $\rho^* = \frac{\sqrt{z}}{1-\sqrt{z}}$, we now have that for $0 < z < \frac{1}{4}$,

$$\ddot{r}(0) = -(1-\sqrt{z})^3. \tag{49}$$

A.4 BPSK and QPSK

Combining the results regarding $\dot{r}(0)$ and $\ddot{r}(0)$ in the previous subsections and Theorem 5, we have proved Theorem 1. Regarding Theorem 2, the first part of the theorem is a direct consequence of Lemma 6, which has already been proved. For the second part of the theorem regarding BPSK and QPSK signaling, we can again do the calculations in a scalar channel with small power as we have proceeded with the proof of Theorem 1.

Since for both BPSK and QPSK, we have $|x|^2 = p$ with probability 1, the power constraint parameter β does not play a role here and $\tilde{E}_o(p, q_p, \rho)$ can be simplified to

$$\tilde{E}_o(p, q_p, \rho) = -\ln \int \alpha(y)^{1+\rho} dy,$$

with

$$\alpha(y) = \int q_p(x) f_w(y|x)^{\frac{1}{1+\rho}} dx.$$

Again, we use the two inequalities which have been very helpful to us in the general first and second order calculations:

$$\begin{aligned} (1+t)^{1+\rho} &\leq 1 + (1+\rho)t + \frac{\rho(1+\rho)}{2}t^2; & (50) \\ (1+t)^{1+\rho} &\geq 1 + (1+\rho)t + \frac{\rho(1+\rho)}{2}t^2 - \frac{\rho(1+\rho)(1-\rho)}{6}t^3. & (51) \end{aligned}$$

We write $\int \alpha(y) dy$ as follows

$$\int \alpha(y) dy = \int f_w(y)(1+T(y))^{1+\rho} dy, \tag{52}$$

where $T(y)$ denotes

$$T(y) = \sum_k q_k \left(\frac{f_w(y-x)}{f_w(y)} \right)^{\frac{1}{1+\rho}} - 1.$$

From here, we can apply the inequalities (50) and (51) and follow a similar procedure as in the proof of Theorem 1 to tightly bound $\dot{r}(0)$ and $\ddot{r}(0)$ when the input is constrained to be BPSK signaling. We can obtain that:

$$\begin{aligned}\dot{\tilde{r}}(0) &= \sup_{0\le\rho\le 1} -\frac{z}{\rho}+\frac{1}{1+\rho} = (1-\sqrt{z})^2;\\ \ddot{\tilde{r}}(0) &= \frac{\ddot{\tilde{E}}_o(0,BPSK,\rho^*)}{\rho^*} = -2(1-\sqrt{z})^3,\end{aligned}$$

Therefore, BPSK is first-order optimal but not second-order optimal.

The QPSK calculations are very similar to the BPSK calculations and we can show that for QPSK

$$\begin{aligned}\dot{\tilde{r}}(0) &= \sup_{0\le\rho\le 1} -\frac{z}{\rho}+\frac{1}{1+\rho} = (1-\sqrt{z})^2;\\ \ddot{\tilde{r}}(0) &= \frac{\ddot{\tilde{E}}_o(0,QPSK,\rho^*)}{\rho^*} = -(1-\sqrt{z})^3,\end{aligned}$$

which implies that QPSK is near-optimal.

B Proof of Theorem 3 and Theorem 4

In this section, we will prove Theorem 3 and Theorem 4. For simplicity, we only prove the case for $B=1$, i.e., we focus on one of the b parallel channels in the channel model (7). The extension to the general case with b parallel channels is quite straightforward. Since $b=1$, we drop the subscript of l in (7) and we have

$$\mathbf{y} = H\mathbf{x} + \mathbf{w}. \tag{53}$$

We assume the average power available in each block is PT_c, i.e.,

$$E[\|\mathbf{x}\|^2] = PT_c. \tag{54}$$

Thus, the energy per degree of freedom is $\frac{P}{W_c}$, which is small when W_c is large.

In this proof, we will use the results for AWGN channels extensively. To avoid confusion in the notation, we will use a superscript "NF" (Non-Fading) to denote any quantity that was computed for the AWGN channel.

B.1 $R(0)$ and first-order optimal condition

In the near capacity region ($R > R_{crit}$), where the random-coding exponent and sphere-packing exponent are tight, the reliability function constraint can be written as

$$\sup_{0\le\rho\le 1} -\rho R + E_o(P,\rho,W_c) = z,$$

and

$$E_o(P,\rho,W_c) = \frac{1}{T_c}\sup_{q\in\mathcal{F}_{W_c}(P)}\sup_{\beta\geq 0} -\ln E_H\left[\int(\int q(\mathbf{x})e^{\beta(\|\mathbf{x}\|^2-PT_c)}f(\mathbf{y}|\mathbf{x},H)^{\frac{1}{1+\rho}}d\mathbf{x})^{1+\rho}d\mathbf{y}\right]. \tag{55}$$

In this section, we first compute $R(0)$ in the case of $b=1$, i.e., Eq. (27) in Theorem 3 is true. Similar to the AWGN case, we can rewrite the error-exponent constraint as

$$R(1/W_c) = \sup_{0\leq\rho\leq 1} -\frac{z}{\rho} + \frac{E_o(P,\rho,W_c)}{\rho}, \tag{56}$$

given that $E_o(P,\rho,W_c)$ is a bounded quantity. However, to provides bound for $E_o(P,\rho,W_c)$ is a bit more involved as compared to the AWGN case, since Gallager's first-order calculations can not be applied here directly due to the fading coherence between the degrees of freedom. For clarity, we establish the bounds in the following lemma:

Lemma 7 *For any* $\rho\in[0,1]$,

$$0\leq E_o(P,\rho,W_c)\leq\frac{1}{T_c}\ln(1+\frac{\rho PT_c}{1+\rho}). \tag{57}$$

Proof: The lower bound is obvious due to the positive property of E_o [4]. The upper bound can be obtained by using the fact that

$$E_o^{NF}(p,\rho)\leq\frac{p\rho}{1+\rho}.$$

Refer to [17] for a complete proof. ◇

Similar to the AWGN case, the key part in the first-order calculation is to establish the *uniform convergence* of $E_o(P,\rho,W_c)$ to $E_o(P,\rho,\infty)$, which further validates the exchange of $limit$ and $supremum$ in the calculation of $R(0)$. The uniform convergence is stated in the lemma below.

Lemma 8 *When* W_c *goes to infinity,*

$$E_o(P,\rho,W_c)\rightarrow\frac{1}{T_c}\ln(1+\frac{\rho PT_c}{1+\rho}) \text{ uniformly for } \rho\in[0,1]. \tag{58}$$

Proof: Because of (57), it suffices to show that for any $\epsilon>0$, we can find $W_c^{(\epsilon)}$, such that

$$E_o(P,\rho,W_c)\geq\frac{1}{T_c}\ln(1+\frac{\rho PT_c}{1+\rho})-\epsilon,$$

for any $W_c\geq W_c^{(\epsilon)}$ and for all $\rho\in[0,1]$.

From the definition of $E_o(P,\rho,W_c)$, we know for any specific choice of $\{q^*\}\in\mathcal{F}_{W_c}(P)$, we have

$$E_o(P,\rho,W_c)\geq\tilde{E}_o(P,q^*,\rho,W_c),$$

where $\tilde{E}_o(P, q^*, \rho, W_c)$ is defined as follows

$$\tilde{E}_o(P, q^*, \rho, W_c) = \frac{1}{T_c} \sup_{\beta \geq 0} - \ln E_H \left[\int (\int q^*(\mathbf{x}) e^{\beta(\|\mathbf{x}\|^2 - PT_c)} f(\mathbf{y}|\mathbf{x}, H)^{\frac{1}{1+\rho}} d\mathbf{x})^{1+\rho} d\mathbf{y} \right]. \tag{59}$$

Now we choose q^* to be QPSK. Since now $\|\mathbf{x}\|^2 = PT_c$ with probability 1, the power-constraint parameter β does not affect $\tilde{E}_o(P, QPSK, \rho, W_c)$ and we have

$$\tilde{E}_o(P, QPSK, \rho, W_c) = -\frac{1}{T_c} \ln E_H \left[\exp\{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)\} \right], \tag{60}$$

where $\tilde{E}_o^{NF}(p, QPSK, \rho)$ is

$$\tilde{E}_o^{NF}(p, QPSK, \rho) = -\ln \int E_x[f_w(y-x)^{\frac{1}{1+\rho}}]^{1+\rho} dy.$$

Next we show that for any $\epsilon > 0$, we can find $W_c^{(\epsilon)}$, such that

$$\tilde{E}_o(P, QPSK, \rho, W_c) \geq \frac{1}{T_c} \ln(1 + \frac{\rho PT_c}{1+\rho}) - \epsilon.$$

From (60), it suffices to show that

$$(1 + \frac{\rho P}{1+\rho}) E_H \left[\exp\{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)\} \right] < e^{\epsilon T_c}. \tag{61}$$

In last section, we have already shown that as $p \to 0$, $\frac{\tilde{E}_o^{NF}(p, QPSK, \rho)}{p\rho} \to \frac{1}{1+\rho}$ uniformly. In other words, for any $\epsilon' > 0$, we can find $\xi > 0$, such that for all $p \leq \xi$,

$$\frac{\tilde{E}_o^{NF}(p, QPSK, \rho)}{p\rho} > \frac{1}{1+\rho} - \epsilon', \quad \text{for all } \rho \in [0, 1],$$

or equivalently,

$$\tilde{E}_o^{NF}(p, QPSK, \rho) > \frac{p\rho}{1+\rho} - \epsilon' p\rho, \quad \text{for all } \rho \in [0, 1], \tag{62}$$

Note that

$$\begin{aligned}
& E_H \left[\exp\{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)\} \right] \\
= \; & E_H \left[e^{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)} \middle| |H|^2 \leq \frac{\xi W_c}{P} \right] + E_H \left[e^{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)} \middle| |H|^2 > \frac{\xi W_c}{P} \right] \\
\leq \; & E_H \left[e^{-D(\frac{\rho}{1+\rho} - \epsilon'\rho)\frac{P|H|^2}{W_c}} \middle| |H|^2 \leq \frac{\xi W_c}{P} \right] + Pr(|H|^2 > \frac{\xi W_c}{P}) \qquad (63) \\
\leq \; & E_H \left[e^{-(\frac{\rho}{1+\rho} - \epsilon'\rho)PT_c|H|^2} \right] + Pr(|H|^2 > \frac{\xi W_c}{P}). \qquad (64)
\end{aligned}$$

The inequality in (63) comes from (62) and the fact that $E_o(p, QPSK, \rho) \geq 0$. For Rayleigh fading, we can compute (64) and we have

$$E_H\left[\exp\{-D\tilde{E}_o^{NF}(\frac{P|H|^2}{W_c}, QPSK, \rho)\}\right] \leq \frac{1}{1+(\frac{\rho}{1+\rho}-\epsilon'\rho)PT_c} + e^{-\frac{\xi W_c}{P}}.$$

We choose ϵ' such that $\epsilon' = \frac{\epsilon}{2P}$. We can then find the corresponding ξ with respect to this choice of ϵ'. We then choose $W_c^{(\epsilon)}$ such that

$$e^{-\frac{W_c^{(\epsilon)}\xi}{P}} < \frac{\epsilon}{2(1+P)}.$$

It is straightforward to check that for all $W_c \geq W_c^{(\epsilon)}$, (61) will be held and thus complete the proof of this Lemma. ◇

From here, it is straightforward to obtain

$$\begin{aligned} R(0) &= \lim_{W_c\to\infty} R(1/W_c) && (65)\\ &= \sup_{0\leq\rho\leq 1} -\frac{z}{\rho} + \frac{1}{T_c}\frac{\ln\left(1+\frac{\rho PT_c}{1+\rho}\right)}{\rho}. && (66)\end{aligned}$$

Next we present a sufficient condition for a sequence of input distributions $q_{W_c}(\mathbf{x})$ to be first order optimal. Due to the similarity to the AWGN case, we present the following lemmas on first-order optimal signaling schemes without proof.

Lemma 9 *Assuming $0 < z < z^*$, where z^* is defined by (26), a sufficient condition for $\{q_{W_c}\}$ to be first-order optimal is that*

$$\lim_{W_c\to\infty} \tilde{E}_o(P, q_{W_c}, \rho^*, W_c) = \frac{1}{T_c}\ln(1+\frac{\rho^* PT_c}{1+\rho^*}), \qquad (67)$$

where ρ^* is the optimizing ρ for (66).

Similar to the AWGN channel, in the fading channel with large coherence bandwidth W_c, it does not take much to be first-order optimal. We restrict ourselves to those vector input distributions which are i.i.d. in each dimension.

Lemma 10 *For i.i.d. input distributions, such that $q_{W_c}(\mathbf{x}) = \Pi_{d=1}^{D} q(x_d)$, a sufficient condition for $\{q_{W_c}(\mathbf{x})\} \in \mathcal{F}_{W_c}(P)$ to be first-order optimal is that $q(x)$ is symmetric around zero, i.e.*

$$q(x) = q(-x).$$

B.2 $\dot{R}(0)$ and second-order optimal condition

To compute $\dot{R}(0)$, we first establish a relationship between $\dot{R}(0)$ and the derivative of $E_o(P,\rho,W_c)$ with respect to $1/W_c$. We omit the proof of the following results due to the similarity to the AWGN case.

Theorem 8 *If as W_c goes to infinity, for each $\rho \in [0,1]$, the limit of of $W_c\left[E_o(P,\rho,W_c) - E_o(P,\rho,\infty)\right]$ exists, which we denote as $\dot{E}_o(P,\rho,\infty)$ and is a continuous function in ρ, and further,*

$$W_c\left[E_o(P,\rho,W_c) - E_o(P,\rho,\infty)\right] \to \dot{E}_o(P,\rho,\infty) \quad \textit{uniformly for all } \rho \in [0,1], \tag{68}$$

$\dot{R}(0)$ can be determined as

$$\dot{R}(0) = \frac{\dot{E}_o(P,\rho^*,\infty)}{\rho^*}, \tag{69}$$

where ρ^* is the optimizing ρ in (66).

◇

Next we verify the uniform convergence assumption needed in Theorem 8.

Lemma 11 *As W_c goes to infinity, we have*

$$W_c\left[E_o(P,\rho,W_c) - E_o(P,\rho,\infty)\right] \to -\frac{\rho P^2}{(1+\rho)(1+\rho+\rho P T_c)^2} \quad \textit{uniformly for } \rho \in [0,1]. \tag{70}$$

Proof: Similar to the proof of Theorem 7. Refer to [17] for a complete proof. ◇

B.3 BPSK and QPSK

Now we prove the statement in Theorem 4 regarding BPSK and QPSK. The first-order optimality of BPSK and QPSK can be easily seen from Lemma 10. In the proof of Lemma 11, we essentially showed that by choosing the input distribution of QPSK,

$$\tilde{E}_o(P,QPSK,\rho,W_c) \geq \frac{1}{T_c}\ln(1+\frac{\rho P T_c}{1+\rho}) - \frac{\rho P^2}{(1+\rho)(1+\rho+\rho P T_c)^2 W_c} + \delta(\frac{1}{W_c}).$$

On the other hand, it was also shown in the proof of Lemma 11 that

$$\tilde{E}_o(P,QPSK,\rho,W_c) \leq E_o(P,\rho,W_c) \leq \frac{1}{T_c}\ln(1+\frac{\rho P T_c}{1+\rho}) - \frac{\rho P^2}{(1+\rho)(1+\rho+\rho P T_c)^2 W_c} + \delta(\frac{1}{W_c}).$$

Thus, we must have

$$W_c[\tilde{E}_o(P,QPSK,\rho,W_c) - \frac{1}{T_c}\ln(1+\frac{\rho P T_c}{1+\rho})] \to -\frac{\rho P^2}{(1+\rho)(1+\rho+\rho P T_c)^2} \quad \textit{uniformly for } \rho \in [0,1].$$

Following a similar argument as in Theorem 8, we can easily obtain

$$\dot{\tilde{R}}(0) = \dot{R}(0) = -\frac{P^2}{(1+\rho^*)(1+\rho^*+\rho^* PT_c)^2}.$$

For BPSK, using the result in last section regarding BPSK, we can obtain that

$$W_c[\tilde{E}_o(P, QPSK, \rho, W_c) - \ln(1+\frac{\rho P}{1+\rho})] \to -\frac{2\rho P^2}{(1+\rho)(1+\rho+\rho PT_c)^2} \quad \textit{uniformly for } \rho \in [0,1].$$

Thus,

$$\dot{\tilde{R}}(0) = -\frac{2P^2}{(1+\rho^*)(1+\rho^*+\rho^* PT_c)^2} < \dot{R}(0).$$

Therefore, QPSK is near optimal while BPSK is not. ◇

References

[1] R. E. Blahut. *Principles and Practise of Information Theory*. Addison-Wesley, Owego, NY, 1987.

[2] R. M. Fano. *Transmission of Information*. M.I.T. Press, and Wiley, New York, NY, 1961.

[3] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11(3):3–18, January 1965.

[4] R. G. Gallager. *Information Theory and Reliable Communication*. John Wiley and Sons, New York, NY, 1968.

[5] R. G. Gallager. Energy limited channels: Coding, multiaccess, and spread spectrum, November 1987. Tech. Report LIDS-P-1714, LIDS, MIT, Cambridge, Mass.

[6] B. Hajek and V. G. Subramanian. Capacity and reliability function for small peak signal constraints. *IEEE Transactions on Information Theory*, 48(4):828–839, April 2002.

[7] J. L. Massey. All signal sets centered about the origin are optimal at low energy-to-noise ratios on the AWGN channel. 1976. in Proc. 1976 IEEE Int. Symp. Information Theory, June 21-24.

[8] M. Médard and R. G. Gallager. Bandwidth scaling for fading multipath channels. *IEEE Transactions on Information Theory*, 48(4):840–853, April 2002.

[9] M. Médard and D. N. Tse. Spreading in block-fading channels. volume 2, 2000. Conference Record of the 34th *Asilomar Conference on Signal, Systems and Computers*.

[10] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. I. *Informaion and Control*, 10:65–103, 1967.

[11] E. Telatar and D. N. Tse. Capacity and mutual information of wideband multipath fading channels. *IEEE Transactions on Information Theory*, 46(7):1384–1400, July 2000.

[12] D. Tse and P. Viswanath. Fundamentals of Wireless Communications. *Preprint*, 2004.

[13] S. Verdú. On channel capacity per unit cost. *IEEE Transactions on Information Theory*, 36(9):1019–1030, September 1990.

[14] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Transactions on Information Theory*, 48(6):1319–1343, June 2002.

[15] J. M. Wozencraft and I. M. Jacobs. *Principles of Communications Engineering*. John Wiley and Sons Inc., 1965.

[16] X. Wu and R. Srikant. MIMO channels in the low-SNR regime: communication rate, error exponent and signal peakiness, 2004. A shorter version of this paper appeared in the *Proceedings of the IEEE Information Theory Workshop*, San Antonio, TX, Oct. 2004. Preprint available at http://www.uiuc.edu/˜xwu.

[17] X. Wu and R. Srikant Near-optimal Input Distributions in the Wideband Regime. Technical report. Available at http://www.comm.csl.uiuc.edu/˜xwu/ and http://arXiv.org.

[18] L. Zheng, M. Mèdard, D.N.C. Tse, and C. Luo. On channel coherence in the low SNR regime. In *Proc. of the 41st Allerton Annual Conference on Communication, Control and Computing*, October 2003.