The Information-theoretic Optimality of QPSK

Sergio Verdú
Dept. Electrical Engineering
Princeton University
Princeton, New Jersey 08544
Verdu@Princeton.edu

Abstract— **Quadrature Phase Shift Keying (QPSK) finds ubiquitous use in coherent communication systems. In this paper we show that in the low-power wideband regime, QPSK is a capacity-maximizing strategy for additive-noise Gaussian complex-valued channels subject to fading where the receiver has full information about the channel realization. The traditional asymptotic optimality criterion used since Shannon (1948) is that the derivative at zero SNR be optimal. However, this optimality criterion is not strong enough to withstand the test of spectrally efficient finite-bandwidth communication. In contrast, in this paper we propose a new asymptotic optimality criterion where both the first and second derivatives at zero SNR are required to be optimal. The setting of this paper is general enough to encompass multiantenna channels, frequency selectivity, and multiaccess channels. We also illustrate the robustness of QPSK in the absence of full receiver knowledge of the channel.**

I. Optimality criterion

As long as the receiver has full knowledge of the channel, it is well-known that even in the presence of fading, the input signaling that maximizes the input-output mutual information of additive Gaussian noise channels subject to a power constraint is Gaussian. The second-order statistics of the optimum Gaussian distribution depend on the knowledge of the channel available at the transmitter. Furthermore, at any fixed signal-to-noise ratio any non-Gaussian distribution is strictly suboptimal.

In practice, however, Gaussian-like inputs have the shortcoming of dictating large transmitted peak-to-average ratios which are prohibitive in battery-operated systems with current amplifier technology. Starting with Shannon [1] in 1948, and continuing, most notably, with Smith [2] in 1971 and Shamai and Bar-David [3] in 1995 (see also [4], [5], [6]), information theorists have studied the capacity of amplitude (or, so-called, peak-power) constrained systems. The ensuing optimization problem is considerably more involved than in the power-constrained case, and, in particular, the optimizing distribution is discrete with a finite number of masses that grows with the signal-to-noise ratio. Closed-form solutions for the location and weight of each mass are not available and numerical optimization is necessary.

Shannon [1] was the first to point out that as the signal-to-noise ratio goes to zero, binary antipodal inputs incur no asymptotic loss in capacity in the sense that the ratio of binary-input mutual information to Gaussian-input mutual information goes to one, i.e.

$$\lim_{\mathsf{SNR}\to 0} \frac{I(X_{\mathsf{SNR}}; Y_{\mathsf{SNR}})}{C(\mathsf{SNR})} = 1 \tag{1}$$

where $C(\mathsf{SNR})$ stands for the capacity at signal-to-noise ratio SNR. This result on the asymptotic optimality of binary phase-shift keying (BPSK), obtained in the context of the real-valued channel, would seemingly lead one to conclude that in the complex-valued channel, and therefore in carrier-modulated communication, QPSK is optimal.

The reason why this is not the end of the story is that the asymptotic optimality criterion used by Shannon and subsequent works (namely, ratio of mutual information to capacity approaching one) is, surprisingly, too weak to be of practical relevance. We can get an inkling of the weakness of this "traditional" optimality criterion from the fact that according to it, BPSK is asymptotically optimum not only in the real-valued channel but in the complex-valued channel. Yet, BPSK requires twice the bandwidth of QPSK to send the same data rate at the same power. Thus, unless bandwidth is free, BPSK is a very poor signaling choice.

The ratio of mutual information (achieved with a particular input signaling format) to capacity approaches one with vanishing signal-to-noise ratio if and only if the first derivative with respect to signal-to-noise ratio (at zero signal-to-noise ratio) of the mutual information coincides with that of the capacity. The derivative at zero signal-to-noise ratio of the mutual information (in bits) is equal to the reciprocal of $\frac{E_b}{N_0}_{\min}$, the energy per bit normalized to the white noise spectral level required for reliable communication using that particular signal strategy. In particular, for the capacity-achieving input we have

$$\frac{E_b}{N_0}_{\min} = \lim_{\mathsf{SNR}\to 0} \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \tag{2}$$

$$= \frac{\log_e 2}{\dot{C}(0)} \tag{3}$$

where

$$\dot{C}(0) = \text{derivative at 0 of } C(\mathsf{SNR})$$

computed in nats/dimension.

Thus, the traditional low-power asymptotic optimality criterion is equivalent to requiring that the signaling strategy achieve the minimum energy per bit achievable with

This work was supported by the National Science Foundation

the best signaling strategy. Several previous contributions have noticed that it takes very little for the input signaling to satisfy this optimality criterion. Golay [7] noticed that the infinite-bandwidth limit of Shannon's AWGN capacity formula can be approached by on-off keying (pulse position modulation) with very low duty cycle. For the AWGN channel, Massey [8] and Lapidoth and Shamai [9] showed that any signal constellation with zero mean achieves the optimum derivative of the cutoff rate and capacity with respect to signal-to-noise ratio. The necessary and sufficient condition for the input signal to achieve minimum energy per bit was recently been found in [10]: When the receiver has perfect knowledge of the channel any input that wastes negligible power in the mean achieves minimum energy per bit. The basic problem with the traditional optimality criterion of matching the slope at zero signal-to-noise ratio of the capacity, or equivalently, achieving minimum energy per bit, is that it is only relevant in the zero spectral efficiency infinite-bandwidth limit, and thus leads to the practically useless conclusion that BPSK is asymptotically optimal.

In this paper (which presents a subset of the results reported in the upcoming journal publication [10])) we show that the sensible optimality criterion in the wideband regime is that *both the first and second derivatives of the capacity* $C(\mathsf{SNR})$ *at* $\mathsf{SNR}=0$ *are achieved by the input signaling.*

Suppose that the transmit power is fixed to P and the information data rate is fixed to R the required bandwidth is given by

$$B = \frac{R}{\mathsf{C}\left(\frac{1}{R}\frac{P}{N_0}\right)}. \tag{4}$$

where we have denoted

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) = \text{spectral efficiency (b/s/Hz)}. \tag{5}$$

We see from (4) that it makes sense to compare signaling strategies by looking at the ratio of their respective spectral efficiencies, since this determines their respective bandwidth requirements for given R and P. As the power vanishes so does the spectral efficiency, leading to the *wideband* regime. So, our focus will be the *wideband slope* of the spectral efficiency function. Since each second × hertz requires one complex dimension, the spectral efficiency is equal to the conventional channel capacity measured in bits per channel use:

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) = C(\mathsf{SNR}) \tag{6}$$

where SNR is the solution to

$$\frac{E_b}{N_0}C(\mathsf{SNR}) = \mathsf{SNR}. \tag{7}$$

Our approach is to analyze the first-order behavior of the spectral efficiency vs $\frac{E_b}{N_0}$ function in the wideband limit. The first-order Taylor series expansion of the inverse of (5) is

$$\begin{aligned} 10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) &= 10\log_{10}\frac{E_b}{N_0}_{\min} \\ &+ \frac{\mathsf{C}}{\mathcal{S}_0}10\log_{10}2 \\ &+ o(\mathsf{C}), \quad \mathsf{C}\to 0 \end{aligned} \tag{8}$$

where $\frac{E_b}{N_0}_{\min}$ denotes the minimum $\frac{E_b}{N_0}$ required for reliable communication, and $\mathcal{S}_0$ denotes the slope of spectral efficiency in b/s/Hz/(3 dB) at the point $\frac{E_b}{N_0}_{\min}$:

$$\mathcal{S}_0 \stackrel{\text{def}}{=} \lim_{\frac{E_b}{N_0}\downarrow\frac{E_b}{N_0}_{\min}} \frac{\mathsf{C}(\frac{E_b}{N_0})}{10\log_{10}\frac{E_b}{N_0} - 10\log_{10}\frac{E_b}{N_0}_{\min}} 10\log_{10}2 \tag{9}$$

A general result from [10] gives the *wideband slope* in terms of both the first and second derivatives of the capacity-SNR function:

Theorem 1: At $\frac{E_b}{N_0}_{\min}$, the slope of the spectral efficiency vs. $\frac{E_b}{N_0}$ in b/s/Hz/(3 dB) is given by

$$\mathcal{S}_0 = \frac{2\left[\dot{C}(0)\right]^2}{-\ddot{C}(0)} \tag{10}$$

with $\dot{C}$ and $\ddot{C}$, the first and second derivative, respectively, of the function $C(\mathsf{SNR})$ computed in nats.

Therefore, it makes sense to define an input distribution to be first-order optimal if it achieves $\dot{C}(0)$, and second-order optimal if it achieves both $\dot{C}(0)$ and $\ddot{C}(0)$, and consequently, both $\frac{E_b}{N_0}_{\min}$ and the maximum wideband slope.

II. Additive white Gaussian Noise channel

Although we consider here the simplest channel, our result on the optimality of QPSK is new even in this setting. Consider the complex-valued memoryless channel with additive white Gaussian noise

$$\mathsf{y} = \mathsf{x} + \mathsf{n} \tag{11}$$

whose capacity is

$$C(\mathsf{SNR}) = \log_2(1+\mathsf{SNR}) \tag{12}$$

where

$$\mathsf{SNR} = E[|\mathsf{x}|^2]/E[|\mathsf{n}|^2].$$

From (12), the transmitted $\frac{E_b}{N_0}$ required to achieve a given spectral efficiency is equal to:

$$\frac{E_b}{N_0}(\mathsf{C}) = \frac{2^{\mathsf{C}}-1}{\mathsf{C}} \tag{13}$$

Figure 1 plots (13) and its affine approximation. Direct computation shows that for the additive white Gaussian noise channel, the wideband slope is

$$\mathcal{S}_0 = 2 \text{ b/s/Hz/(3 dB)}, \tag{14}$$

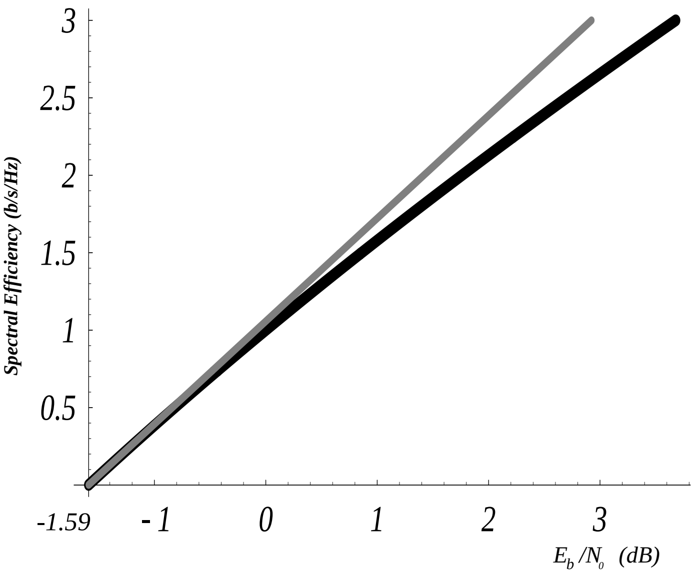

Fig. 1. Spectral Efficiency of the Additive White Gaussian Noise Channel and its wideband approximation.

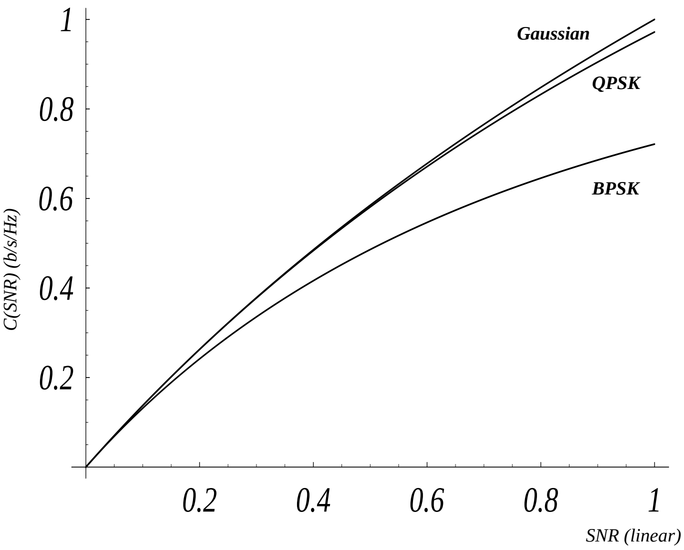

Fig. 2. Capacity achieved by complex Gaussian inputs, QPSK and BPSK in the additive white Gaussian noise channel.

which, in this case, is the highest slope achieved for any $\frac{E_b}{N_0}$. Note that although convex for the AWGN channel, in general, the function $10 \log_{10} \frac{E_b}{N_0}(\mathsf{C})$ need not be convex.

Let us denote the mutual informations achieved by BPSK and QPSK as a function of SNR by $C_{BPSK}(\mathsf{SNR})$ and $C_{QPSK}(\mathsf{SNR})$, respectively. With QPSK at $A(\pm 1 \pm j)$ the mutual information is equal to that achieved by two independent channels with BPSK inputs $\pm A$. Since the SNR of the latter channels is half of that of the original channel we have the relationship:

$$C_{QPSK}(\mathsf{SNR}) = 2\, C_{BPSK}(\mathsf{SNR}/2) \tag{15}$$

Furthermore, the evaluation of the binary-input mutual information results in

$$\begin{aligned} & C_{BPSK}(\mathsf{SNR}) = 2\,\mathsf{SNR}\, \log_2 e \\ - & \frac{e^{-\mathsf{SNR}}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2} \cosh(2\sqrt{\mathsf{SNR}} z) \log_2(\cosh(2\sqrt{\mathsf{SNR}} z)) dz \\ = & \ \mathsf{SNR}\, \log_2 e - \mathsf{SNR}^2 \log_2 e + o(\mathsf{SNR}^2) \end{aligned} \tag{16}$$

Figure 2 plots $C_{BPSK}(\mathsf{SNR})$, $C_{QPSK}(\mathsf{SNR})$ and $C(\mathsf{SNR})$ with SNR in linear scale. Note that the derivative at the origin is identical for all three curves. Thus, BPSK and QPSK are first-order optimal.

From (16) we get the following derivatives

$$\dot{C}_{BPSK}(0) = 1 \tag{17}$$

and

$$\ddot{C}_{BPSK}(0) = -2 \tag{18}$$

Thus, according to (10) , the wideband slope of BPSK is

$$\mathcal{S}_{BPSK} = 1 \text{ b/s/Hz/(3 dB)} \tag{19}$$

Moreover, it follows from (15) and (7) that the spectral efficiencies are related by:

$$\mathsf{C}_{QPSK}\left(\frac{E_b}{N_0}\right) = 2\, \mathsf{C}_{BPSK}\left(\frac{E_b}{N_0}\right), \tag{20}$$

and consequently the wideband slope of QPSK is twice that of BPSK. Thus, QPSK is wideband optimal. Furthermore, note that any signaling distribution that can be written as a mixture of (rotated and scaled) QPSK distributions is also wideband optimal because mutual information is concave in the input distribution.

Figure 3 plots the spectral efficiencies as a function of $\frac{E_b}{N_0}$ achieved by optimum (complex Gaussian) signaling, QPSK and BPSK. We can see that the wideband affine approximation is accurate for a fairly large range of $\frac{E_b}{N_0}$.

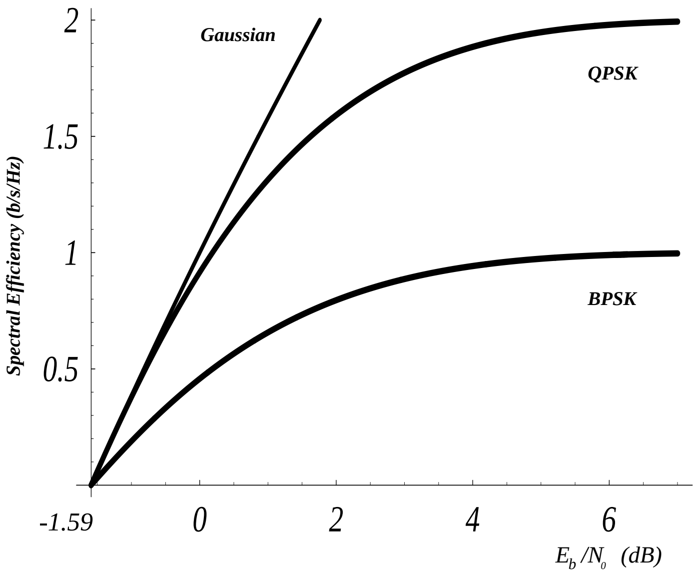

Fig. 3. Spectral efficiencies achieved by complex Gaussian inputs, QPSK and BPSK in the additive white Gaussian noise channel.

It is interesting to reexamine the interplay between power, bandwidth and rate by defining the rates achievable with a given bandwidth and power:

$$R_{BPSK}(B, P) = B\, C_{BPSK}\left(\frac{P}{B N_0}\right) \tag{21}$$

$$R_{QPSK}(B,P) = B\,C_{QPSK}\left(\frac{P}{BN_0}\right) = 2B\,C_{BPSK}\left(\frac{P}{2BN_0}\right) \quad (22)$$

where we have used (15). It follows from (21) and (22) that

$$R_{BPSK}(B,P) = R_{QPSK}(B/2,P) \quad (23)$$

for any P. Once again, we see that if we dictate the same power P and the same rate R for a QPSK and a BPSK system, the QPSK system requires half the bandwidth of the BPSK system, not only in the wideband regime but for arbitrary power and bandwidth.

If we dictate the same bandwidth and the same rate, and the BPSK system requires power P, then the QPSK requires reduced power $\alpha_B(P)P$ determined as the solution to the equation:

$$R_{BPSK}(B,P) = R_{QPSK}(B,\alpha_B(P)P) \quad (24)$$

Since in this case the value of the spectral efficiency is identical, the $\frac{E_b}{N_0}$ required by QPSK is lower. As $P \to 0$ the $\frac{E_b}{N_0}$ required by BPSK behaves as $(1-P/(BN_0))\log_e 2$, whereas the $\frac{E_b}{N_0}$ required by QPSK behaves as $(1-P/(2BN_0))\log_e 2$. Thus,

$$\lim_{P\to 0} \alpha_B(P) = \frac{1}{2}. \quad (25)$$

Let us suppose now that we dictate the same power and the same bandwidth for both systems. Then, the ratio of the achievable rates is

$$\frac{R_{QPSK}(B,P)}{R_{BPSK}(B,P)} = 2\frac{C_{BPSK}\left(\frac{P}{2BN_0}\right)}{C_{BPSK}\left(\frac{P}{BN_0}\right)} \quad (26)$$

which is a monotonically increasing function of $\frac{P}{BN_0}$ tending to 2 in the high power limit and converging to 1 as $\frac{P}{BN_0} \to 0$. Paradoxically, we now get no advantage for QPSK over BPSK in the asymptotically low-power regime. Again, the reason is that by focusing on the first-order behavior of capacity as a function of SNR, we lose any connection between rate and bandwidth. Indeed for any $\eta > 0$

$$\lim_{P\to 0} \frac{R_{BPSK}(B,P)}{R_{BPSK}(\eta B,P)} = 1 \quad (27)$$

and

$$\lim_{B\to \infty} \frac{R_{BPSK}(B,P)}{R_{BPSK}(\eta B,P)} = 1 \quad (28)$$

for any $B > 0$ and $P > 0$ respectively, and analogously for QPSK or for the (Gaussian-input) capacity.

III. Multidimensional Fading Channel Model

Consider the following discrete-time channel with m complex dimensions:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (29)$$

where the real and imaginary parts of the noise components are independent and satisfy

$$E[\|\mathbf{n}\|^2] = mN_0, \quad (30)$$

$\mathbf{H}$ is an $m \times n$ complex matrix whose random coefficients have finite second moments, independent real and imaginary parts, and are independent of $\mathbf{x}$ and $\mathbf{n}$.

Model (29) encompasses a variety of channels of interest in communications subject to crosstalk (e.g. digital subscriber lines), multiantenna wireless, frequency selective fading, and multiaccess channels. We assume that the variation of the channel matrix from symbol to symbol is ergodic, so that averaging capacity expressions over $\mathbf{H}$ has operational significance.

The realization of $\mathbf{H}$ at each symbol is known at the receiver. In contrast, knowledge of the channel at the transmitter does not go beyond knowledge of the maximal-eigenvalue eigenspaces of realizations of $\mathbf{H}^\dagger\mathbf{H}$.

Reference [10] shows the following result (also in the absence of channel information at the receiver)

Theorem 2: The required received energy per bit for reliable communication through (29) satisfies

$$\frac{E_b^{\mathrm{r}}}{N_0}_{\min} = \log_e 2 = -1.59 \text{ dB}. \quad (31)$$

Furthermore, if the transmitter knows the maximal-eigenvalue eigenspace of $\mathbf{H}^\dagger\mathbf{H}$ but not its maximal eigenvalue then the required transmitted energy per bit is equal to

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{E[\sigma^2_{\max}(\mathbf{H})]} \quad (32)$$

with $\sigma_{\max}$ denoting the largest singular value; if the transmitter only knows the a priori statistics of $\mathbf{H}$, then the required transmitted energy per bit is equal to

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\lambda_{\max}(E[\mathbf{H}^\dagger\mathbf{H}])} \quad (33)$$

with $\lambda_{\max}$ denoting the largest eigenvalue. If the transmitted components of $\mathbf{x}$ are constrained to be independent (e.g. in multiaccess communications or in pragmatic multiantenna designs) the transmitted energy per bit is equal to

$$\frac{E_b}{N_0}_{\min} = \frac{n\log_e 2}{E[\mathrm{trace}\{\mathbf{H}^\dagger\mathbf{H}\}]} \quad (34)$$

Regarding the wideband slope we have the following results [10].

Theorem 3: Suppose that the transmitter knows the maximal eigenvalue eigenspace of $\mathbf{H}^\dagger\mathbf{H}$ but not the maximal eigenvalue. Then

$$\mathcal{S}_0 = \frac{2\ell}{m\,\kappa(\sigma_{\max}(\mathbf{H}))} \quad (35)$$

with the *kurtosis*[1] of a random variable Z defined as

$$\kappa(Z) = \frac{E[Z^4]}{E^2[Z^2]}, \quad (36)$$

[1] The "amount of fading" defined in [11] is equal to the kurtosis minus 1. See [11], [12] for tables of standard fading distributions.

$\sigma_{\max}(\mathbf{H})$ denotes the maximal singular value of $\mathbf{H}$, and ℓ is equal to the multiplicity of $\sigma_{\max}(\mathbf{H})$. Furthermore, the optimum wideband slope is achieved by QPSK-modulating with equal power the ℓ orthogonal dimensions of the maximal-eigenvalue eigenspace of $\mathbf{H}^\dagger\mathbf{H}$.

The special case $m = n = 1$ of Theorem 3 was given in [12]. Kurtosis is a measure of the randomness of a random variable; its minimum value is 1, achieved uniquely by a deterministic variable. The fading penalty on capacity is due to the concavity of the $\log(1+x)$ function. The larger the "spread" of the fading distribution, the larger is the penalty. Theorem 3 states that in the low spectral efficiency region, the required bandwidth is proportional to the kurtosis of the maximal singular value of the channel. If the number of rows and columns of $\mathbf{H}$ grow, while keeping a constant ratio, and its coefficients are independent identically distributed, then the maximal singular value converges to a deterministic constant [13], and its multiplicity goes to 1. Accordingly, if m represents the number of receiving antennas and the transmitter knows the eigenstructure of $\mathbf{H}$, in the limit of many antennas at both transmitter and receiver the slope is 2 b/s/Hz/(3 dB), i.e., the same value obtained with one antenna but without fading.

But in the multiantenna literature it is much more common to assume that the transmitter has no knowledge whatsoever of $\mathbf{H}$. The following result applies to that case.

Theorem 4: Suppose that the transmitted components are constrained to be independent and to have the same power. Then

$$\mathcal{S}_0 = \frac{2}{m}\frac{\left(\mathrm{trace}(E[\mathbf{H}^\dagger\mathbf{H}])\right)^2}{\mathrm{trace}(E[(\mathbf{H}^\dagger\mathbf{H})^2])}. \qquad (37)$$

In the multiantenna literature it is common to model the entries of $\mathbf{H}$ as independent zero-mean Gaussian random variables. In that case, it can be checked that (37) reduces to twice the harmonic mean of the number of transmit and receive antennas:

$$\mathcal{S}_0 = \frac{2nm}{m+n}\text{b/s/Hz/(3 dB)}. \qquad (38)$$

Theorem 5: Under the conditions of Theorem 4, equal-power QPSK on each component is second-order optimal.

Proof: For ease of notation and without loss of generality we assume in the proof that $N_0 = 1$. The input signaling is

$$\mathbf{x} = \sqrt{\frac{\mathsf{SNR}\; m}{n}}\begin{bmatrix} e^{j\phi_1} \\ \vdots \\ e^{j\phi_n}\end{bmatrix} \qquad (39)$$

where the phases ϕ_i are independent and equally likely to take the values $\{\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}\}$. Thus,

$$E[\|\mathbf{x}\|^2] = m\;\mathsf{SNR} \qquad (40)$$

and

$$E[\mathbf{x}\mathbf{x}^\dagger] = \frac{m\;\mathsf{SNR}}{n}\mathbf{I}. \qquad (41)$$

This input distribution attains the following mutual information

$$\begin{aligned}\frac{1}{m}I(\mathbf{x};\mathbf{Hx}+\mathbf{n}) &= \frac{1}{m}D(P_{Y|X}\|P_{Y|X=0}|P_X) \\ &- \frac{1}{m}D(P_Y\|P_{Y|X=0}) \qquad (42)\\ &= \frac{1}{m}E[\|\mathbf{Hx}\|^2] - \frac{1}{m}D(P_Y\|P_{Y|X=0}) \qquad (43)\end{aligned}$$

Furthermore, it follows from (41) that

$$E[\|\mathbf{Hx}\|^2] = \frac{m\;\mathsf{SNR}}{n}\mathrm{trace}(E[\mathbf{H}^\dagger\mathbf{H}]). \qquad (44)$$

Therefore, from the result obtained in (37), the desired second-order optimality is equivalent to showing that

$$\lim_{\mathsf{SNR}\to 0}\frac{D(P_Y\|P_{Y|X=0})}{m\;\mathsf{SNR}^2} = \frac{m}{2n^2}\mathrm{trace}(E[(\mathbf{H}^\dagger\mathbf{H})^2]). \qquad (45)$$

To accomplish this, note that the divergence in (45) can be expressed as

$$\begin{aligned}D(P_Y\|P_{Y|X=0}) &= E[Z(\mathbf{n},\mathbf{H})\log Z(\mathbf{n},\mathbf{H})] \\ &= E[Z(\mathbf{n},\mathbf{H})-1] \\ &+ \frac{1}{2}E[(Z(\mathbf{n},\mathbf{H})-1)^2] \\ &+ o(\mathsf{SNR}^2) \qquad (46)\end{aligned}$$

where the expectation is with respect to the complex vector $\mathbf{n}$ distributed according to $P_{Y|X=0} = \mathcal{N}(0,\mathbf{I})$ and with respect to $\mathbf{H}$; and we have defined the likelihood ratio

$$\frac{dP_Y}{dP_{Y|X=0}} = Z(\mathbf{n},\mathbf{H}) = E[\exp(-\|\mathbf{Hx}-\mathbf{n}\|^2 + \|\mathbf{n}\|^2)\mid \mathbf{H},\mathbf{n}] \qquad (47)$$

where the expectation is with respect to $\mathbf{x}$. Note that

$$E[Z(\mathbf{n},\mathbf{H})-1\mid\mathbf{H}] = -1 + \int\frac{dP_Y}{dP_{Y|X=0}}dP_{Y|X=0} = 0 \qquad (48)$$

A Taylor series expansion of the exponential in (47) together with the fact that $E[\mathbf{n}] = 0$ results in

$$\begin{aligned}Z(\mathbf{n},\mathbf{H}) - 1 &= -E[\|\mathbf{Hx}\|^2\mid\mathbf{H}] \\ &+ \frac{1}{2}E[(\mathbf{n}^\dagger\mathbf{Hx}+\mathbf{x}^\dagger\mathbf{H}^\dagger\mathbf{n} - \|\mathbf{Hx}\|^2)^2\mid\mathbf{H},\mathbf{n}] \\ &+ o(\mathsf{SNR}) \qquad (49)\end{aligned}$$

where all the expectations are with respect to $\mathbf{x}$. Let us consider each term in the right side of (49) individually. Using (41) we get

$$E[\|\mathbf{Hx}\|^2\mid\mathbf{H}] = \frac{m\;\mathsf{SNR}}{n}\mathrm{trace}(\mathbf{H}^\dagger\mathbf{H}) \qquad (50)$$

and

$$\begin{aligned}&E[(\mathbf{n}^\dagger\mathbf{Hx}+\mathbf{x}^\dagger\mathbf{H}^\dagger\mathbf{n} - \|\mathbf{Hx}\|^2)^2\mid\mathbf{H},\mathbf{n}] \\ =\;& E[\|\mathbf{Hx}\|^4\mid\mathbf{H}] + E[(\mathbf{n}^\dagger\mathbf{Hx}+\mathbf{x}^\dagger\mathbf{H}^\dagger\mathbf{n})^2\mid\mathbf{H},\mathbf{n}]. \qquad (51)\end{aligned}$$

The second term in the right side of (51) is equal to

$$E[(\mathbf{n}^\dagger \mathbf{H}\mathbf{x} + \mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{n})^2 \mid \mathbf{H}, \mathbf{n}] = \frac{2m \, \mathsf{SNR}}{n} \mathbf{n}^\dagger \mathbf{H}\mathbf{H}^\dagger \mathbf{n} \quad (52)$$

because

$$E[(\mathbf{n}^\dagger \mathbf{H}\mathbf{x})^2 \mid \mathbf{H}, \mathbf{n}] = E[(\mathbf{x}^\dagger \mathbf{H}^\dagger \mathbf{n})^2 \mid \mathbf{H}, \mathbf{n}] = 0 \quad (53)$$

as can be seen using the QPSK nature of the independent components of $\mathbf{x}$. Using (49), (50), (51) and (52) we can write

$$\begin{aligned} & \frac{n^2 E[(Z(\mathbf{n}, \mathbf{H}) - 1)^2 \mid \mathbf{H}]}{m^2 \mathsf{SNR}^2} \\ = \; & E\left[\left(\mathrm{trace}(\mathbf{H}^\dagger \mathbf{H}) - \mathbf{n}^\dagger \mathbf{H}\mathbf{H}^\dagger \mathbf{n}\right)^2 \mid \mathbf{H}\right] \\ + \; & o(\mathsf{SNR}^2) \\ = \; & E\left[\left(\mathbf{n}^\dagger \mathbf{H}\mathbf{H}^\dagger \mathbf{n}\right)^2 \mid \mathbf{H}\right] - \mathrm{trace}^2(\mathbf{H}^\dagger \mathbf{H}) \\ + \; & o(\mathsf{SNR}^2) \end{aligned} \quad (54)$$

In order to compute the expectation with respect to $\mathbf{n}$ fix and arbitrary hermitian matrix $\mathbf{M}$. Then, using the fact that for all components

$$E[|\mathsf{n}_i|^4] = 2 \quad (55)$$

it is fairly straightforward to check that

$$E[(\mathbf{n}^\dagger \mathbf{M}\mathbf{n})^2] = \mathrm{trace}^2(\mathbf{M}) + \mathrm{trace}(\mathbf{M}^2) \quad (56)$$

Applying formula (56) to (54) with $\mathbf{M} = \mathbf{H}\mathbf{H}^\dagger$, the desired result (45) follows. □

IV. Ricean Channel

As a special case of the wideband slope formulas found above we can analyze the Ricean channel in the wideband regime assuming that the receiver knows both the specular and the Rayleigh component.

Theorem 6: Consider the $m = n = 1$ Ricean fading channel

$$\mathsf{y} = (\bar{\mathsf{h}} + \mathsf{g})\mathsf{x} + \mathsf{n} \quad (57)$$

where $\bar{\mathsf{h}}$ is deterministic, g is zero-mean complex Gaussian with variance γ^2, and the additive noise is Gaussian. If the receiver (but not the transmitter) knows the Rayleigh channel coefficients, then the wideband slope is equal to

$$\mathcal{S}_0 = \frac{1}{1 - \frac{1}{2}\left(1 + \frac{\gamma^2}{|\bar{\mathsf{h}}|^2}\right)^{-2}}. \quad (58)$$

Proof: When the receiver knows the channel coefficients we just need to specialize Theorems 3 or 4 (in the scalar case) to the case $m = n = 1$ in which they lead to the same result. Formula (58) follows from the kurtosis of the Rician distribution which is equal to

$$\kappa(|\bar{\mathsf{h}} + \mathsf{g}|) = 2 - \frac{|\bar{\mathsf{h}}|^4}{(|\bar{\mathsf{h}}|^2 + \gamma^2)^2}. \quad (59)$$

□

It is shown in [10] that if the receiver does not know the Rayleigh component the wideband slope is zero, regardless of the relative strength of the specular and Rayleigh components. Thus, in noncoherent communication (or in general, in the absence of perfect receiver knowledge) it is very demanding in terms of bandwidth to achieve $\frac{E_b^r}{N_0}_{\min}$ close to -1.59dB in the Ricean channel, regardless of the relative strengths of the specular and Rayleigh components. In many practical cases in which the specular component is not negligible, QPSK is an attractive suboptimal alternative as the following result shows.

Theorem 7: Consider the Ricean channel (57) with $\bar{\mathsf{h}} \neq 0$ and a receiver that does not know the Rayleigh coefficients. Then QPSK achieves

$$\frac{E_b}{N_0}_{\min} = \frac{\log_2 e}{|\bar{\mathsf{h}}|^2} \quad (60)$$

and

$$\mathcal{S}_0 = 2, \quad (61)$$

i.e., the same wideband performance as if the Rayleigh component were absent.

Proof: With the Ricean channel,

$$P_{Y|X=\mathsf{x}_0} = \mathcal{N}(\bar{\mathsf{h}}\mathsf{x}_0, N_0 + \gamma^2 |\mathsf{x}_0|^2), \quad (62)$$

and the QPSK input distribution

$$P_{\mathsf{x}} = \frac{1}{4}\delta_{\mathsf{A}(1+j)} + \frac{1}{4}\delta_{\mathsf{A}(1-j)} + \frac{1}{4}\delta_{\mathsf{A}(-1+j)} + \frac{1}{4}\delta_{\mathsf{A}(-1-j)} \quad (63)$$

we have

$$\mathsf{SNR} = \frac{2|\mathsf{A}|^2}{N_0} \quad (64)$$

and we obtain (in nats)

$$\begin{aligned} D(P_{Y|X} \| P_{Y|X=0} | P_X) &= (\gamma^2 + |\bar{\mathsf{h}}|^2)\mathsf{SNR} - \log_e(1 + \gamma^2 \mathsf{SNR}) \\ &= |\bar{\mathsf{h}}|^2 \mathsf{SNR} + \frac{\gamma^4}{2}\mathsf{SNR}^2 + o(\mathsf{SNR}^2) \end{aligned} \quad (65)$$

From the formulas for $\frac{E_b}{N_0}_{\min}$, and $\mathcal{S}_0$, both (60) and (61) will follow upon showing that

$$\lim_{\mathsf{SNR} \to 0} \frac{D(P_Y \| P_{Y|X=0})}{\mathsf{SNR}^2} = \frac{1}{2}(\gamma^4 + |\bar{\mathsf{h}}|^4) \quad (66)$$

where

$$\begin{aligned} P_Y = \; & \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(1+j), \alpha^2 N_0) + \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(1-j), \alpha^2 N_0) \\ + \; & \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(-1+j), \alpha^2 N_0) + \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(-1-j), \alpha^2 N_0), \end{aligned} \quad (67)$$

and

$$\alpha^2 = 1 + \mathsf{SNR}\gamma^2 \quad (68)$$

The desired divergence is

$$D(P_Y \| P_{Y|X=0}) = E[\log q(\mathsf{y})] \tag{69}$$

where the expectation is with respect to (67) and

$$\begin{aligned} q(\mathsf{y}) &\stackrel{\text{def}}{=} \frac{dP_Y}{dP_{Y|X=0}}(\mathsf{y}) \\ &= \frac{1}{\alpha^2}\exp\left(-\frac{2|\bar{\mathsf{h}}|^2|\mathsf{A}|^2}{\alpha^2 N_0}\right)\exp\left(-\frac{|\mathsf{y}|^2}{N_0}\left(\alpha^{-2}-1\right)\right) \\ &\quad \cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Re \mathsf{y}\right)\cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Im \mathsf{y}\right). \end{aligned} \tag{70}$$

where for convenience we have assumed $\bar{\mathsf{h}}\mathsf{A}$ to be real without affecting the result. Taking the expectation of the logarithm of (70) we obtain

$$\begin{aligned} D(P_Y \| P_{Y|X=0}) &= \gamma^2 \mathsf{SNR} - \log_e(1+\gamma^2\mathsf{SNR}) - \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{1+\gamma^2\mathsf{SNR}} \\ &+ E\left[\log_e \cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Re \mathsf{y}\right)\right] \\ &+ E\left[\log_e \cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Im \mathsf{y}\right)\right] \end{aligned} \tag{71}$$

The random variables $\Re\mathsf{y}$ and $\Im\mathsf{y}$ are Gaussian with mean $|\bar{\mathsf{h}}||\mathsf{A}|$ and variance

$$N_0/2 + |\mathsf{A}|^2\gamma^2 = \frac{\alpha^2 N_0}{2} \tag{72}$$

Thus, the second and fourth moments of the random variables in the argument of the hyperbolic cosines in (71) are equal to

$$4\frac{|\bar{\mathsf{h}}|^2|\mathsf{A}|^2}{\alpha^4 N_0^2}\left(\frac{\alpha^2 N_0}{2} + |\bar{\mathsf{h}}|^2|\mathsf{A}|^2\right) = \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\left(1+\frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\right), \tag{73}$$

and

$$\frac{16|\bar{\mathsf{h}}|^4|\mathsf{A}|^4}{\alpha^8 N_0^4}\frac{3\alpha^4 N_0^2}{4} + o(\mathsf{SNR}^2) = \frac{3|\bar{\mathsf{h}}|^4\mathsf{SNR}^2}{\alpha^4} + o(\mathsf{SNR}^2) \tag{74}$$

Using (73) and (74), and

$$\log_e(\cosh(x)) = \frac{x^2}{2} - \frac{x^4}{12} + o(x^4) \tag{75}$$

each of the expectations in (71) satisfy

$$\begin{aligned} E\left[\log_e \cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Re \mathsf{y}\right)\right] &= \frac{1}{2}\frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\left(1+\frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{2\alpha^2}\right) \\ &+ o(\mathsf{SNR}^2) \end{aligned} \tag{76}$$

Thus,

$$\begin{aligned} D(P_Y \| P_{Y|X=0}) &= \gamma^2\mathsf{SNR} - \log_e(1+\gamma^2\mathsf{SNR}) - \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2} \\ &+ \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\left(1+\frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{2\alpha^2}\right) + o(\mathsf{SNR}^2) \\ &= \frac{\gamma^4\mathsf{SNR}^2}{2} + \frac{|\bar{\mathsf{h}}|^4\mathsf{SNR}^2}{2} + o(\mathsf{SNR}^2), \end{aligned} \tag{77}$$

thereby establishing (66). □

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, Jul.-Oct. 1948.

[2] J.G. Smith, "The information capacity of amplitude- and variance-constrained scalar gaussian channels," *Information and Control*, vol. 18, pp. 203–219, 1971.

[3] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Transactions on Information Theory*, pp. 1060–1071, July 1995.

[4] G. Färber, "Die kanalkapazität allgemeiner übertragungskanänale bei begrenztem signalwertbereich, beliebigen signalübertragungszeiten sowie beliebiger stürung," *Arch. Elektr. Übertr.*, vol. 21, H.11, H.12, pp. 565–574, 653–660, 1967.

[5] G. Färber, "Bestimmung der kanalkapazität signalwertbegrentzer analog kanäle durch nichtlineare programming," *Nachr. Techn. Z.*, vol. 21, pp. 341–348, 1968.

[6] S. Shamai (Shitz), "On the capacity of a gaussian channel with peak power and bandlimited input signals," *AEÜ Archiv für Electronik und Übertragungstechnik*, pp. 340–346, 1988.

[7] M. J. E. Golay, "Note on the theoretical efficiency of information reception with PPM," *Proc. IRE*, vol. 37, pp. 1031, Sep. 1949.

[8] J.L. Massey, "All signal sets centered about the origin are optimal at low energy-to-noise ratios on the AWGN channel," *Proc. 1976 IEEE Int. Symp. Information Theory*, pp. 80–81, June 21-24 1976.

[9] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side-information" be?," *IEEE Trans. Information Theory*, 2001.

[10] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Information Theory*, to appear, 2002.

[11] M. K. Simon and M. S. Alouini, *Digital Commincation over Fading Channels: A Unified Approach to Performance Analysis*, Wiley Interscience, New York, 2000.

[12] S. Shamai (Shitz) and S. Verdú, "The impact of flat-fading on the spectral efficiency of CDMA," *IEEE Trans. on Information Theory*, vol. 47, pp. 1302–1327, May 2001.

[13] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, pp. 175–192, 1995.