Spectral Efficiency in the Wideband Regime

Sergio Verdú, *Fellow, IEEE*

Invited Paper

***Abstract*—The tradeoff of spectral efficiency (b/s/Hz) versus energy-per-information bit is the key measure of channel capacity in the wideband power-limited regime. This paper finds the fundamental bandwidth–power tradeoff of a general class of channels in the wideband regime characterized by low, but nonzero, spectral efficiency and energy per bit close to the minimum value required for reliable communication. A new criterion for optimality of signaling in the wideband regime is proposed, which, in contrast to the traditional criterion, is meaningful for finite-bandwidth communication.**

***Index Terms*—Antenna arrays, channel capacity, fading channels, noisy channels, spectral efficiency, wideband regime.**

I. Introduction

SHORTLY after "A Mathematical Theory of Communication," Claude Shannon [1] pointed out that as the bandwidth tends to infinity, the channel capacity of an ideal bandlimited additive white Gaussian noise (AWGN) channel approaches

$$\lim_{B\to\infty} B\log_2\left(1+\frac{P}{BN_0}\right)=\frac{P}{N_0}\log_2 e \quad \text{(b/s)} \tag{1}$$

where P is the received power and N_0 is the one-sided noise spectral level. Since capacity is monotonically increasing with bandwidth B, the right-hand side of (1) is the maximum rate achievable with power P. Moreover, communicating at rate R, the received signal energy per information bit is equal to

$$E_b^{\rm r}=\frac{P}{R} \tag{2}$$

and since the maximum value of R is the right-hand side of (1), the minimum received signal energy per information bit required for reliable communication satisfies

$$\frac{E_b^{\rm r}}{N_0}_{\min}=\frac{P}{N_0\frac{P}{N_0}\log_2 e}=\log_e 2=-1.59\text{ dB}. \tag{3}$$

Gaussian inputs are not mandatory to achieve (1). In 1948, Shannon [2] had already noticed that for low signal-to-noise ratios, binary antipodal inputs are as good as Gaussian inputs in the sense that the ratio of mutual information to capacity approaches unity. Since then, this criterion (which can be rephrased as the input attaining the derivative of capacity at zero signal-to-noise ratio) has traditionally been adopted as a synonym of asymptotic optimality in the low-power regime. Using this criterion, Golay [3] showed that (1) can be approached by on–off keying (pulse position modulation) with very low duty cycle, a signaling strategy whose error probability was analyzed in [4].

Enter fading. Jacobs [5] and Pierce [6] noticed not only that (1) is achieved if all the energy is concentrated in one message-dependent frequency slot, but also that the limiting rate in (1) is unexpectedly robust: it is achievable even if the orthogonal signaling undergoes fading which is unknown to the receiver (a result popularized by Kennedy [7]). Since only one frequency (or time) slot carries energy, this type of orthogonal signaling not only is extremely "peaky" but requires that the number of slots grows exponentially with the number of information bits.

Doppler spread or a limitation in the peakiness of the orthogonal signaling can be modeled by letting the signal-dependent waveform at the receiver have a given power spectral density $S(f)$, shifted in frequency by an amount that depends on the message, with different shifts sufficiently far apart to maintain orthogonality. In this case, the infinite-bandwidth achievable rate was obtained by Viterbi [8]

$$\int_{-\infty}^{\infty}\frac{S(f)}{N_0}\,df\log_2 e-\int_{-\infty}^{\infty}\log_2\left(1+\frac{S(f)}{N_0}\right)df.$$

As we saw in (3), determining the infinite-bandwidth capacity is equivalent to finding the minimum energy per bit required to transmit information reliably. To obtain this quantity, we can choose to maximize the information per unit energy in contrast to the standard Shannon setting in which the information per degree of freedom is maximized. Motivated by the optimality of on–off signaling in the infinite-bandwidth limit, Gallager [9] found the exact reliability function in the setting of a binary-input channel where information is normalized, not to blocklength, but to the number of "1's" contained in the codeword. More generally, we can pose the "capacity per unit cost" problem where an arbitrary cost function is defined on the input alphabet [10]. An important class of cost functions are those which, like energy, assign a zero cost to one of the input symbols. For those cost functions, the capacity per unit cost not only is equal to the derivative at zero cost of the Shannon capacity but admits a simple formula [10]. Even in this more general setting, capacity per unit cost is achieved by on–off signaling with vanishing duty cycle.

A wide variety of digital communication systems (particularly in wireless, satellite, deep-space, and sensor networks) operate in the power-limited region where both spectral efficiency (rate in bits per second divided by bandwidth in

Manuscript received April 18, 2001; revised January 22, 2002.

The author is with the Department Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: Verdu@Princeton.edu).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04027-0.

hertz) and energy-per-bit are relatively low. The wideband regime is an attractive choice because of power savings, ease of multiaccess, ability of overlay with other systems, and diversity against frequency-selective fading. The information-theoretic analysis of those channels, in addition to leading to the most efficient bandwidth utilization, reveals design insights on good signaling strategies.

From the existing results we could draw the following conclusions about signaling and capacity in the wideband regime.

- On–off signaling approaches capacity as the duty cycle vanishes.
- The derivative at zero signal-to-noise ratio of the Shannon capacity determines the wideband fundamental limits.
- Capacity is not affected by fading.
- Receiver knowledge of channel fade coefficients is useless.
- An input whose mutual information achieves the derivative of capacity at zero signal-to-noise ratio is wideband optimal.

These conclusions and design guidelines have been drawn in the literature in the natural belief that infinite-bandwidth limits are representative of the large (but finite) bandwidth regime of interest in practice. However, in this paper, we show that those conclusions are misguided as long as the allowed bandwidth is finite, regardless of how large it is. Indeed, operation in the regime of low spectral efficiency does not imply disregard for the bandwidth required by the system. Thus, we will see that design guidelines obtained by infinite-bandwidth analyses need not carry over to the wideband regime.

It follows from (1) that to achieve a given rate R bits per second in the AWGN channel (or any other channel that attains the same limiting capacity (1)), we require power

$$P = N_0 R \log_e 2. \tag{4}$$

However, this minimum power is sufficient only provided that infinite bandwidth is available. Thus, in addition to transmitter/receiver complexity, attaining (1) or (3) entails zero spectral efficiency. If we are willing to spend more power than (4), then the required bandwidth is finite. However, even in the wideband regime neither the limits in (1), (3) nor the derivative of the Shannon capacity at zero signal-to-noise ratio determine the bandwidth-versus-power tradeoff. Of course, the solution can be found from the full Shannon capacity function for arbitrary signal-to-noise ratios. Unfortunately, the capacity function and the inputs that attain it are unknown for many channels of interest, particularly in the presence of fading (cf. [11]). Even for channels whose capacity is known for all signal-to-noise ratios, no method is available to establish the bandwidth–power tradeoff in the wideband regime. In this paper, we show that it is possible to obtain analytically the fundamental limits of a general class of additive-noise channels in the wideband regime in which the spectral efficiency is low but nonzero. These results offer engineering guidance on the fundamental bandwidth–power tradeoff and on signaling strategies that attain it in the wideband limit.

The tradeoff bandwidth versus power is mirrored in the tradeoff of the information-theoretic quantities: spectral efficiency and E_b/N_0 (energy per bit normalized to background noise spectral level). Our approach for the wideband regime is to approximate spectral efficiency as an affine function of E_b/N_0 (decibels). Thus, we are interested in obtaining not only $\frac{E_b}{N_0}_{\min}$ but the *wideband slope* of the spectral efficiency–$\frac{E_b}{N_0}$ curve in bits per second per hertz per 3 dB (b/s/Hz/(3 dB)) at $\frac{E_b}{N_0}_{\min}$. Spectral efficiency in the wideband regime turns out to be determined by both the first and second derivatives of the channel capacity at zero signal-to-noise ratio.

Section II sets up the general fading channel model we consider in this work. Section III gives the basic relationships between E_b/N_0, signal-to-noise ratio, capacity, and spectral efficiency. Section IV is devoted to the problem of finding the $\frac{E_b}{N_0}_{\min}$ required for reliable communication. It establishes, in considerably wider generality than was previously known, that the received energy per bit normalized to noise spectral level in a Gaussian channel subject to fading is -1.59 dB, regardless of side information at the transmitter and/or receiver. In Section IV, we also show that if the channel is known at the receiver almost any input signaling achieves $\frac{E_b}{N_0}_{\min}$. Otherwise, it is necessary (and sufficient) to use a generalized form of on–off signaling with unbounded amplitude, which we refer to as flash signaling. Whereas the required transmit $\frac{E_b}{N_0}_{\min}$ depends on the transmitter side information but not on the receiver side information, Section V shows that the wideband slope depends crucially on the receiver side information. We show in Section V that the traditional optimality criterion (attaining the first derivative) is not strong enough to withstand the test of spectrally efficient finite-bandwidth communication. We propose a new asymptotic optimality criterion whereby both the first and second derivatives at zero signal-to-noise ratio are required to be optimal. Under this optimality criterion, the spectral efficiency is maximized in the wideband regime. If the receiver knows the channel, quadrature phase shift keying (QPSK) is shown to be wideband optimal. Receiver knowledge of fading coefficients is shown to have a deep impact on both the required bandwidth and the optimal signaling strategies. When the channel has an unknown component, approaching $\frac{E_b}{N_0}_{\min}$ turns out to be very demanding both in bandwidth and in the peak-to-average ratio of the transmitted signals. The kurtosis of the fading distribution plays a key role in determining the wideband capacity of the fading channel when the receiver is able to track the channel coefficients. Several new results on the spectral efficiency of multiple-access and multiantenna channels are also given in Sections IV and V.

II. Channel Model

In this paper, we deal with additive-noise channels in a general setting which allows other random channel impairments such as fading. Consider the following discrete-time channel with m complex dimensions:

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{5}$$

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

$$E[\|\boldsymbol{n}\|^2] = mN_0. \tag{6}$$

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

Model (5) encompasses channels that incorporate features such as multielement antennas, frequency-selective fading, multiaccess, and crosstalk. In each of those cases, m and n take different meanings. For example, they may be the number of receive and transmit antennas, or the spreading gain and the number of users, or they may both represent time (or frequency) slots. When several of those features are present at the same time, it may be more convenient to use matrices or higher dimensional objects to represent the input and output quantities. Furthermore, it is often sensible to decompose wideband channels into parallel noninteracting channels. For the sake of simplicity, we restrict ourselves to memoryless channels in this paper; the extension to channels with memory should follow well-known methods (e.g., [12]), in which the number of dimensions grows. Furthermore, in those cases where $\boldsymbol{H}$ is random, we assume that its variation from symbol to symbol is ergodic, so that averaging capacity expressions over $\boldsymbol{H}$ has operational significance.

We will consider a variety of special cases of (5) depending on the transmitter/receiver knowledge of $\boldsymbol{H}$, its statistics, and the statistics of the background noise. When transmitter and/or receiver do not know $\boldsymbol{H}$, we assume that they know its prior distribution. We also evaluate the penalty incurred by not using this prior information at the transmitter. By "knowledge of the channel at the receiver" we imply that the realization of $\boldsymbol{H}$ at each symbol is known at the receiver. An interesting generalization, which is not treated in this paper, is to let both transmitter and receiver have access to noisy observations of $\boldsymbol{H}$.

The discrete-time channel (5) arises, for example, from the analysis of the continuous-time channel

$$y(t) = \tilde{x}(t) + n(t) \tag{7}$$

where $n(t)$ is white noise with power spectral density $\frac{N_0}{2}$, and the information-bearing received signal $\tilde{x}(t)$ is a channel-distorted version of the transmitted signal. If the effective duration and bandwidth of $\tilde{x}(t)$ are T and B, passing $y(t)$ through an orthonormal transformation with $m = TB$ complex dimensions is sufficient to preserve all the information (asymptotically). Thus, if we denote the number of bits encoded in $\tilde{x}(t)$ by b, the number of bits per second per hertz is equal to the number of bits per dimension b/m.

III. SPECTRAL EFFICIENCY VERSUS E_b/N_0

Let E_b denote the transmitted energy per information bit, in the same units as N_0, which we can take to be Joules for the sake of concreteness. The key design quantities *bandwidth* B (Hz), *transmitted power* P (W), and *data rate* R (b/s) satisfy the relationships

$$R\frac{E_b}{N_0} = \frac{P}{N_0} \tag{8}$$

and

$$\frac{R}{B} = \mathsf{C}\left(\frac{E_b}{N_0}\right) \tag{9}$$

where we have denoted

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

Since the required bandwidth to transmit data rate R with power P is given by

$$B = \frac{R}{\mathsf{C}\left(\frac{1}{R}\frac{P}{N_0}\right)} \tag{10}$$

the maximization of the spectral efficiency function is a central goal.

Whereas the spectral efficiency as a function of $\frac{E_b}{N_0}$ is defined for any system with given R, P, N_0, B simply through (8) and (9), this paper focuses on the maximum achievable spectral efficiency (under various constraints of transmitter/receiver knowledge and input signaling). Since each second $\times$ hertz requires one complex dimension, the (maximum achievable) spectral efficiency is equal to the conventional channel capacity measured in bits per channel use.[1] Usually, to obtain channel capacity, it is convenient to place a constraint, not on E_b directly, but on the energy transmitted per symbol vector $E[\|\boldsymbol{x}\|^2]$, or equivalently on its normalized version

$$\begin{aligned} \mathsf{SNR} &= \frac{E[\|\boldsymbol{x}\|^2]}{E[\|\boldsymbol{n}\|^2]} && (11)\\ &= \frac{E[\|\boldsymbol{x}\|^2]}{mN_0} && (12)\\ &= \frac{E_b}{N_0}\frac{b}{m} && (13)\\ &= \frac{E_b}{N_0}C(\mathsf{SNR}) && (14)\end{aligned}$$

where b is equal to the number of bits encoded in $\boldsymbol{x}$ with a capacity-achieving system and Shannon's capacity function $C(\mathsf{SNR})$ (bits/dimension) of the discrete-time channel (5) gives the maximum number of bits per complex dimension achievable under the constraint of arbitrarily reliable communication (vanishing block error probability). It follows that the spectral efficiency–$\frac{E_b}{N_0}$ function can be obtained from Shannon's capacity via[2]

$$\mathsf{C}\left(\frac{E_b}{N_0}\right) = C(\mathsf{SNR}) \tag{15}$$

where SNR is the solution to

$$\frac{E_b}{N_0}C(\mathsf{SNR}) = \mathsf{SNR}. \tag{16}$$

Equivalently, from (15) and (16), the $\frac{E_b}{N_0}$ required to achieve spectral efficiency equal to C is equal to

$$\frac{E_b}{N_0}(\mathsf{C}) = \frac{C^{-1}(\mathsf{C})}{\mathsf{C}} \tag{17}$$

where C^{-1} denotes the inverse function of $C(\mathsf{SNR})$.

[1]The spectral efficiency achieved by a nonideal practical signaling scheme in which each complex dimension occupies α seconds $\times$ hertz is equal to capacity divided α.

[2]The choice of C and C avoids the abuse of notation that assigns the same symbol to capacity functions of SNR and $\frac{E_b}{N_0}$, while at the same time it is advisable not to depart from common usage and denote both functions with the initial of capacity.

The spectral efficiency function in (15) represents b/s/Hz, as long as each complex dimension in (5) takes 1 s $\times$ hertz. However, this is not always the case. For example, if (5) models the signal received at an m-element antenna array (e.g., [13]), the whole m-dimensional vector occupies 1 s $\times$ hertz, in which case, the units of C are b/s/Hz/(antenna element).

Analytically, the $\frac{E_b}{N_0}$ versus spectral efficiency characteristic is of primary importance in the study of the behavior of required power in the wideband limit (where the spectral efficiency R/B is small). As we will see, in the very noisy regime, a first-order analysis of the capacity versus the SNR function is good enough to provide $\frac{E_b}{N_0}_{\min}$, but fails to reveal the first-order growth of the spectral efficiency versus $\frac{E_b}{N_0}$.

Even though the goal, as stated above, is to find the best tradeoff between *transmitted* energy per information bit and spectral efficiency, it is also useful for the sake of comparing results obtained for different channels to represent the fundamental limits in terms of *received* energy per information bit

$$\frac{E_b^{\mathrm{r}}}{N_0} = \frac{E_b}{N_0}\,\frac{E[||\boldsymbol{Hx}||^2]}{E[||\boldsymbol{x}||^2]}. \tag{18}$$

Note that, in general, the channel gain $E[||\boldsymbol{Hx}||^2]/E[||\boldsymbol{x}||^2]$ depends on the input distribution.

Since the explicit solution of (16) is not always feasible, explicit expressions for the spectral efficiency versus E_b/N_0 function are relatively rare. Fortunately, as we will see, in the low-spectral-efficiency regime, it is possible to sidestep not only the solution of the nonlinear equation (16) but even the computation of $C(\mathsf{SNR})$. A notable exception that admits an explicit expression for E_b/N_0 versus spectral efficiency is the deterministic Gaussian channel with either white or nonwhite noise:

Additive White Gaussian Noise (AWGN) Channel.

$$\boldsymbol{H} = \mathsf{A}\boldsymbol{I} \tag{19}$$

$$E[\boldsymbol{nn}^\dagger] = N_0\boldsymbol{I} \tag{20}$$

where A is an arbitrary deterministic complex scalar known at the receiver and $\boldsymbol{n}$ is Gaussian. In this case, the capacity per dimension

$$C(\mathsf{SNR}) = \log_2(1 + |\mathsf{A}|^2\mathsf{SNR}) \tag{21}$$

has a straightforward inverse function. Thus, the transmitted $\frac{E_b}{N_0}$ required to achieve a given spectral efficiency (17) is equal to

$$\frac{E_b}{N_0}(\mathsf{C}) = \frac{2^{\mathsf{C}} - 1}{|\mathsf{A}|^2\mathsf{C}} \tag{22}$$

which in terms of received energy per bit is

$$\frac{E_b^{\mathrm{r}}}{N_0}(\mathsf{C}) = \frac{2^{\mathsf{C}} - 1}{\mathsf{C}}. \tag{23}$$

Fig. 1 compares the capacity (21) (achieved with Gaussian inputs) with the mutual information achieved by binary phase shift keying (BPSK) and by QPSK whose input distributions are

$$P_{\mathsf{x}} = \frac{1}{2}\,\delta_A + \frac{1}{2}\delta_{-A} \tag{24}$$

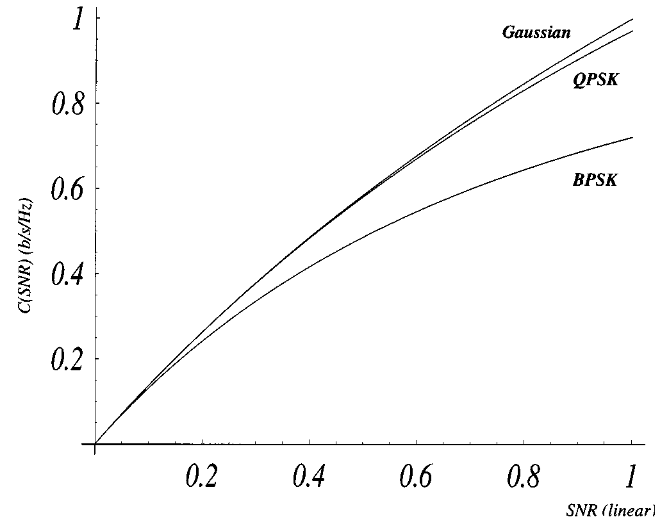

Fig. 1. Capacity achieved by complex Gaussian inputs, QPSK, and BPSK in the AWGN channel.

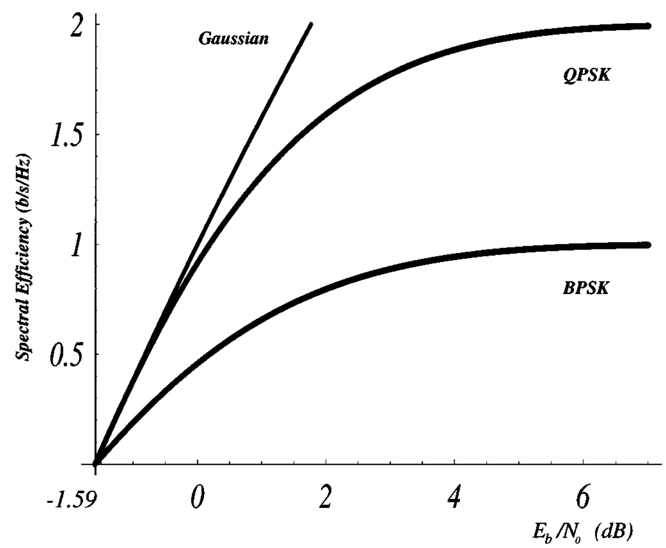

Fig. 2. Spectral efficiencies achieved by complex Gaussian inputs, QPSK, and BPSK in the AWGN channel.

and

$$P_{\mathsf{x}} = \frac{1}{4}\,\delta_{Ae^{j\pi/4}} + \frac{1}{4}\,\delta_{Ae^{j3\pi/4}} + \frac{1}{4}\,\delta_{Ae^{j5\pi/4}} + \frac{1}{4}\,\delta_{Ae^{j7\pi/4}} \tag{25}$$

respectively. Observe from Fig. 1 that both BPSK and QPSK satisfy the traditional wideband optimality criterion, namely, they achieve the same derivative at $\mathsf{SNR} = 0$ as the capacity. Unfortunately, this criterion does not withstand the test of finite bandwidth analysis, since in the complex-valued channel BPSK requires twice the bandwidth of QPSK for any given power and rate, as can be seen in Fig. 2.[3]

Deterministic Channel With Colored Gaussian Noise.

If the channel matrix $\boldsymbol{H}$ is constant over time, known at both transmitter and receiver, and the noise covariance is

$$E[\boldsymbol{nn}^\dagger] = N_0\boldsymbol{\Sigma} \tag{26}$$

[3]Note that BPSK is as efficient for real-valued channels (arising in baseband or single-sideband communication), as QPSK is for complex-valued channels.

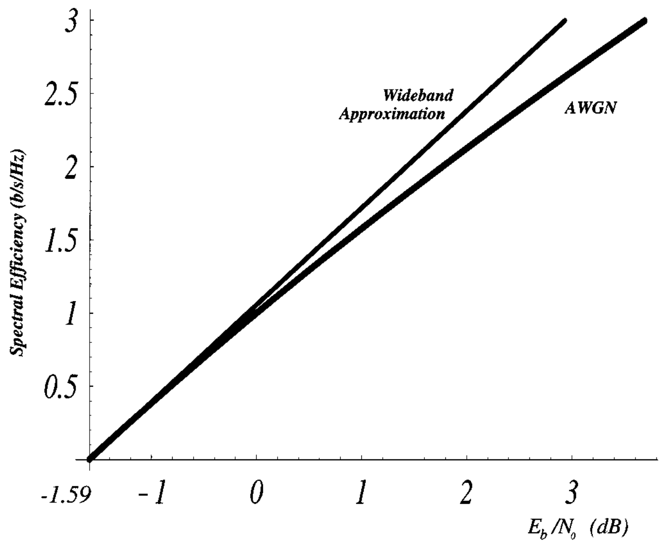

Fig. 3. Spectral efficiency of the AWGN channel and its wideband approximation.

then the water-filling representation for $C(\mathsf{SNR})$ (e.g., [14]) leads to the explicit expression

$$\frac{E_b}{N_0}(\mathsf{C}) = \frac{1}{m\mathsf{C}} \min_{1\le \bar{m} \le \bar{\bar{m}}} \bar{m}\, 2^{m\mathsf{C}/\bar{m}} \prod_{j=1}^{\bar{m}} \rho_j^{1/\bar{m}} - \sum_{j=1}^{\bar{m}} \rho_j \quad (27)$$

where $\rho_1, \ldots, \rho_{\bar{\bar{m}}}$ are the ordered version of the reciprocals of the nonzero eigenvalues of the matrix

$$\boldsymbol{H}^{\dagger}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}.$$

To conclude this section, we formalize the performance measures of interest in this paper.

Our approach is to analyze the first-order behavior of the spectral efficiency versus $\frac{E_b}{N_0}$ function in the wideband limit

$$\begin{aligned} 10\log_{10}\frac{E_b}{N_0}(\mathsf{C}) = &\, 10\log_{10}\frac{E_b}{N_0}_{\min} \\ &+ \frac{\mathsf{C}}{\mathcal{S}_0}10\log_{10} 2 \\ &+ o(\mathsf{C}), \qquad \mathsf{C}\to 0 \end{aligned} \quad (28)$$

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 $\frac{E_b}{N_0}_{\min}$

$$\mathcal{S}_0 \stackrel{\text{def}}{=} \lim_{\frac{E_b}{N_0}\downarrow \frac{E_b}{N_0}_{\min}} \frac{\mathsf{C}(\frac{E_b}{N_0})}{10\log_{10}\frac{E_b}{N_0} - 10\log_{10}\frac{E_b}{N_0}_{\min}} 10\log_{10} 2. \quad (29)$$

The rationale for this first-order analysis is illustrated in Fig. 3, which compares the exact spectral efficiency with its wideband approximation for the AWGN channel. The approximation is an excellent one well beyond the wideband regime. For example, at spectral efficiency equal to 1 b/s/Hz, $\frac{E_b}{N_0} = 0$ dB, whereas the wideband approximation gives $\frac{E_b}{N_0} = -0.09$ dB.[4]

[4]The wideband approximation is optimistic for the AWGN channel, and pessimistic for many fading channels of interest.

Simply using (8), we can generalize (4) to write the minimum transmit power required (with infinite bandwidth) to sustain data rate R as

$$P_{\min} = N_0 R \frac{E_b}{N_0}_{\min}. \quad (30)$$

The bandwidth required to sustain a given rate R with transmitted power $P > P_{\min}$, can be approximated in the wideband regime by the formula

$$\begin{aligned} B &\approx \frac{R}{\mathcal{S}_0}\frac{10\log_{10} 2}{10\log_{10}\frac{P}{P_{\min}}} && (31)\\ &\approx \frac{R}{\mathcal{S}_0}\frac{P_{\min}\log_e 2}{P - P_{\min}} && (32)\\ &= \frac{R^2}{P-P_{\min}}\frac{N_0\log_e 2}{\mathcal{S}_0}\frac{E_b}{N_0}_{\min} && (33)\end{aligned}$$

where (31) follows from the affine approximation to spectral efficiency, (32) follows from the linear approximation to $\log(1+x)$, and (33) follows from (30).

Recent results on the slope b/s/Hz/(3 dB) of various fading channels in the region of high spectral efficiency can be found in [13], [15], and [16].

IV. Minimum $\frac{E_b}{N_0}$

A. Background

Since $C(\mathsf{SNR})$ is a monotonically increasing concave function, (16) results in

$$\begin{aligned} \frac{E_b}{N_0}_{\min} &= \lim_{\mathsf{SNR}\to 0}\frac{\mathsf{SNR}}{C(\mathsf{SNR})} && (34)\\ &= \frac{\log_e 2}{\dot{C}(0)} && (35)\end{aligned}$$

where

$\dot{C}(0) =$ derivative at 0 of $C(\mathsf{SNR})$ computed in nats/dimension.[5]

Several tools to analyze $\dot{C}(0)$ were developed in [10] in the general context of memoryless channels. In particular, it was shown in [10] that if $\mathcal{A}$ is the input symbol alphabet, and the cost function c: $\mathcal{A}\to\Re^+$ is such that there is a zero-cost symbol (denoted by "0"), i.e., $c(0)=0$, then the capacity per unit cost is equal to

$$\sup_{x\in\mathcal{A}}\frac{D(P_{Y|X=x}\|P_{Y|X=0})}{c(x)} = \sup_{x\in\mathcal{A}}\frac{E_{P_{Y|X=x}}\log\left(\frac{dP_{Y|X=x}}{dP_{Y|X=0}}\right)}{c(x)} \quad (36)$$

where $D(P\|Q)$ denotes the divergence between distributions P and Q, and $P_{Y|X=x}$ denotes the conditional output distribution given input x. The foregoing optimization problem has several appealing attributes.

- It is often easier to compute than Shannon capacity.
- Divergence is between conditional output distributions instead of a conditional divergence (mutual information).

[5]In this paper, in addition to logarithms in base e, base 2, and base 10, we use logarithms with arbitrary base. When no base is indicated the logarithms on both sides of the equation have identical bases.

- Optimization is over the input alphabet instead of over the set of distributions defined on it.
- Low duty cycle on–off signaling is asymptotically (as the duty cycle vanishes with the allowed SNR) optimum with on-level at the argument that maximizes (36).
- It shows a connection between information theory and estimation theory. Suppose $\mathcal{A} = \Re$ and $c(x) = x^2$

$$\sup_{x\in\Re} \frac{D(P_{Y|X=x} \| P_{Y|X=0})}{x^2} \geq \lim_{x\to 0} \frac{D(P_{Y|X=x} \| P_{Y|X=0})}{x^2} \quad (37)$$

$$= \frac{\log e}{2} I_0(P_{Y|X}) \quad (38)$$

with Fisher's information

$$I_0(P_{Y|X}) = E_{P_{Y|X=0}}\left[\left(\frac{\partial}{\partial x}\frac{dP_{Y|X=x}}{dP_{Y|X=0}}\right)^2_{x=0}\right].$$

The Cramer–Rao bound together with (38) implies that the minimum energy necessary to transmit 0.5 nats $= 0.721$ b cannot exceed the minimum conditional variance of an estimate of the input from the output given that the input is 0.

Except for the different normalization, the sought-after $\dot{C}(0)$ is given by a formula akin to (26). It is shown in [10] that

$$\lim_{\mathsf{SNR}\to 0} \frac{C(\mathsf{SNR})}{\mathsf{SNR}} = \lim_{\mathsf{SNR}\to 0} \sup_{P_{\boldsymbol{x}}} \frac{I(\boldsymbol{x}; \boldsymbol{Hx}+\boldsymbol{n})}{m\,\mathsf{SNR}} \quad (39)$$

$$= N_0 \lim_{\mathsf{SNR}\to 0} \sup_{P_{\boldsymbol{x}}} \frac{I(\boldsymbol{x}; \boldsymbol{Hx}+\boldsymbol{n})}{E[\|\boldsymbol{x}\|^2]} \quad (40)$$

$$= N_0 \sup_{P_{\boldsymbol{x}}} \frac{D(P_{Y|X=\boldsymbol{x}} \| P_{Y|X=0} | P_{\boldsymbol{x}})}{E[\|\boldsymbol{x}\|^2]} \quad (41)$$

$$= N_0 \sup_{\boldsymbol{x}_0} \frac{D(P_{Y|X=\boldsymbol{x}_0} \| P_{Y|X=0})}{\|\boldsymbol{x}_0\|^2} \quad (42)$$

where $P_{Y|X=\boldsymbol{x}}$ stands for the distribution of $\boldsymbol{Hx}+\boldsymbol{n}$ conditioned on $\boldsymbol{x}$, and in particular $P_{Y|X=0}$ stands for the distribution of $\boldsymbol{n}$. Equation (41) is nontrivial and follows from the approach taken in [10]: no loss of optimality is incurred (as far as the first derivative at 0 is concerned) by restricting $\boldsymbol{x}$ to have the following on–off structure. Fix $\boldsymbol{x}_0$ and let $\boldsymbol{x}$ be $\boldsymbol{x}_0$ with probability δ and 0 with probability $1-\delta$, with δ vanishing proportionally to SNR according to

$$\delta = \mathsf{SNR}\, \frac{N_0 m}{\|\boldsymbol{x}_0\|^2}. \quad (43)$$

We note that a more careful notation would reserve $I(\boldsymbol{x}; \boldsymbol{Hx}+\boldsymbol{n})$ to the case where the receiver does not know the channel, and would use $I(\boldsymbol{x}; \boldsymbol{Hx}+\boldsymbol{n}|\boldsymbol{H})$ to denote the mutual information when the channel is known at the receiver. Although in (40) and the sequel we have chosen not to make the distinction explicit, we should keep in mind that when the channel is known at the receiver, mutual informations and divergences are obtained by averaging conditional expressions over the channel statistics.

B. AWGN Channel

As an example, let us compute (42) for the AWGN channel. Since the noise components in (20) are independent and identically distributed, it suffices to take a scalar channel model $m = 1$. The following well-known formula for the divergence between the distributions of complex-Gaussian random variables will be useful. Let $\mathcal{N}(\mathsf{m}, \sigma^2)$ denote the distribution of a complex Gaussian random variable with mean m and independent real and imaginary components each with variance $\sigma^2/2$, i.e.,

$$p(\mathsf{x}) = \frac{1}{\pi\sigma^2} \exp\left(-|\mathsf{x}-\mathsf{m}|^2/\sigma^2\right).$$

Then

$$D\left(\mathcal{N}\left(\mathsf{m}_1, \sigma_1^2\right) \| \mathcal{N}\left(\mathsf{m}_0, \sigma_0^2\right)\right) = \log\frac{\sigma_0^2}{\sigma_1^2} + \left(\frac{|\mathsf{m}_1-\mathsf{m}_0|^2}{\sigma_0^2} + \frac{\sigma_1^2}{\sigma_0^2} - 1\right)\log e. \quad (44)$$

Using (44) with

$$P_{Y|X=\mathsf{x}} = \mathcal{N}(\mathsf{Ax}, N_0)$$

we obtain

$$D\left(P_{Y|X=\mathsf{x}} \| P_{Y|X=0}\right) = \frac{|\mathsf{Ax}|^2}{N_0} \log e. \quad (45)$$

Therefore, for the AWGN channel the ratio on the right-hand side of (42) is independent of $\boldsymbol{x}_0$. Evaluating the divergence in nats we get

$$\dot{C}(0) = N_0 \sup_{\boldsymbol{x}_0} \frac{D\left(P_{Y|X=\boldsymbol{x}_0} \| P_{Y|X=0}\right)}{\|\boldsymbol{x}_0\|^2} = |\mathsf{A}|^2 \quad (46)$$

and (via (35))

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{|\mathsf{A}|^2} \quad (47)$$

implying the well-known result (cf. (3))

$$\frac{E_b^{\mathsf{r}}}{N_0}_{\min} = -1.59 \text{ dB} \quad (48)$$

which can be obtained directly by letting $\mathsf{C} \to 0$ in the explicit formula (22).

C. Fading Channels With Additive Gaussian Noise

In this subsection, we show that (48) is the required received $\frac{E_b^{\mathsf{r}}}{N_0}_{\min}$ of a very wide class of fading channels with background Gaussian noise.

We saw in (18) that the received and transmitted energies per bit are related by the channel gain. The maximum channel gain

$$\mathsf{G} = \sup_{P_{\boldsymbol{x}}} \frac{E[\|\boldsymbol{Hx}\|^2]}{E[\|\boldsymbol{x}\|^2]} \quad (49)$$

achievable over all choices of the input depends on the knowledge available at the transmitter. We highlight a few special cases of interest.

- The transmitter knows $\boldsymbol{H}$. Then

$$\mathsf{G} = \sup_{\boldsymbol{H}} \sigma^2_{\max}(\boldsymbol{H}) \quad (50)$$

with $\sigma_{\max}$ denoting the largest singular value, and the (essential) supremum on the right-hand side is over all realizations of $\boldsymbol{H}$. Note that to obtain this channel gain, "power control" is required, whereby the instantaneous norm of $\boldsymbol{x}$ is allowed to depend on the channel realization. In effect, the transmitter only puts energy in the maximal-eigenvalue eigenspace of $\boldsymbol{H}^\dagger \boldsymbol{H}$ and only when the maximal eigenvalue is the best possible over all realizations. Note that $\mathsf{G} = \infty$ in the special case of the flat-fading scalar channel where the fading distribution has infinite support (such as Rayleigh).

- The transmitter knows the maximal-eigenvalue eigenspace of $\boldsymbol{H}^\dagger \boldsymbol{H}$ but not the maximal eigenvalue. (Equivalently, the transmitter knows the channel but it is not allowed to employ power control.) Then

$$\mathsf{G} = E[\sigma^2_{\max}(\boldsymbol{H})]. \tag{51}$$

- The transmitter does not know $\boldsymbol{H}$ but it knows its distribution. Then, the input distribution is not allowed to depend on $\boldsymbol{H}$, and the channel gain is

$$\mathsf{G} = \lambda_{\max}(E[\boldsymbol{H}^\dagger \boldsymbol{H}]) \tag{52}$$

with $\lambda_{\max}$ denoting the largest eigenvalue.

- The components of the input are constrained to be independent with equal power. This constraint is common in multiantenna systems and in multiaccess channels. Then

$$\mathsf{G} = \frac{1}{n} E[\mathrm{trace}\{\boldsymbol{H}^\dagger \boldsymbol{H}\}]. \tag{53}$$

Theorem 1: Consider the m-dimensional complex channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{54}$$

where the complex Gaussian vector $\boldsymbol{n}$ has independent and identically distributed components and satisfies (6). Then, regardless of whether $\boldsymbol{H}$ is known at the transmitter and/or receiver, the required received and transmitted energy per bit for reliable communication satisfy

$$\frac{E_b^{\mathrm{r}}}{N_0}_{\min} = \log_e 2 = -1.59 \text{ dB} \tag{55}$$

and

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\mathsf{G}} \tag{56}$$

respectively, where G is the maximum channel gain.

Proof: According to (18) and (35), our goal is to show that

$$\dot{C}(0) = \mathsf{G} \tag{57}$$

and that there is an input distribution that achieves both $\dot{C}(0)$ and the maximum channel gain.

Immaterial for the result, the nature of the receiver knowledge is crucial for the proof. First consider the case where $\boldsymbol{H}$ is known to the receiver. Although in this case we can write the input–output mutual information in closed form, we will follow the general approach outlined in Section IV-A. When the receiver knows the channel, the conditional output distribution is Gaussian

$$P_{Y|X=\boldsymbol{x}_0} = \mathcal{N}(\boldsymbol{H}\boldsymbol{x}_0, N_0\boldsymbol{I}). \tag{58}$$

To compute (42) we need to use the generalization of (44) to multidimensional proper[6] complex Gaussian distributions with independent components

$$\begin{aligned} D\left(\mathcal{N}(\mathsf{m}_1, \boldsymbol{\Sigma}_1) || \mathcal{N}(\mathsf{m}_0, \boldsymbol{\Sigma}_0)\right) &= \log\det \boldsymbol{\Sigma_0} - \log\det \boldsymbol{\Sigma_1} \\ &\quad + (\mathsf{m}_1 - \mathsf{m}_0)^\dagger \boldsymbol{\Sigma_0}^{-1} (\mathsf{m}_1 - \mathsf{m}_0) \log e \\ &\quad + \mathrm{trace}\left(\boldsymbol{\Sigma_0}^{-1}\boldsymbol{\Sigma_1} - \boldsymbol{I}\right) \log e. \end{aligned} \tag{59}$$

Applying (59) to (58), we get that for all $\boldsymbol{x}_0$ and $\boldsymbol{H}$

$$D\left(P_{Y|X=\boldsymbol{x}_0} || P_{Y|X=0}\right) = \frac{||\boldsymbol{H}\boldsymbol{x}_0||^2}{N_0} \log e. \tag{60}$$

Following (41) we get (in nats)

$$\begin{aligned} \dot{C}(0) &= N_0 \sup_{P_{\boldsymbol{x}}} \frac{D\left(P_{Y|X=\boldsymbol{x}} || P_{Y|X=0} | P_{\boldsymbol{x}}\right)}{E[||\boldsymbol{x}||^2]} \\ &= \sup_{P_{\boldsymbol{x}}} \frac{E[||\boldsymbol{H}\boldsymbol{x}||^2]}{E[||\boldsymbol{x}||^2]} \\ &= \mathsf{G}. \end{aligned} \tag{61}$$

As we saw before, the actual value of G will depend on the available knowledge at the transmitter. Note that in the case where the transmitter knows the channel realization and it is allowed to do power control, we do not enforce a constant energy constraint in order to maximize mutual information. This requires a slight generalization of the conventional setting whereby the energy constraint is an auxiliary random variable dependent on the channel realization.

With receiver knowledge, the input distribution that achieves both $\dot{C}(0)$ and G is a Gaussian distribution with covariance dictated by the solution of the maximal channel gain problem. In general, the desired input distribution can be constructed by on–off signaling $P_X = (1-\delta)P_0 + \delta P_{\boldsymbol{x}_0}$ where $P_{\boldsymbol{x}_0}$ is obtained by solving the maximal channel gain problem.

Now, we turn to the case where $\boldsymbol{H}$ is unknown to the receiver. Instead of the full generality of the theorem statement, we assume, for now, that the coefficients of $\boldsymbol{H}$ are jointly Gaussian. In this case, the conditional output distribution is still Gaussian

$$P_{Y|X=\boldsymbol{x}_0} = \mathcal{N}\left(\overline{\boldsymbol{H}}\boldsymbol{x}_0, \mathrm{cov}(\boldsymbol{H}\boldsymbol{x}_0|\boldsymbol{x}_0) + N_0\boldsymbol{I}\right) \tag{62}$$

where

$$E[\boldsymbol{H}] = \overline{\boldsymbol{H}}$$

and we have denoted the conditional covariance

$$\mathrm{cov}(\boldsymbol{H}\boldsymbol{x}_0|\boldsymbol{x}_0) = E\left[\left(\boldsymbol{H} - \overline{\boldsymbol{H}}\right)\boldsymbol{x}_0\boldsymbol{x}_0^\dagger \left(\boldsymbol{H} - \overline{\boldsymbol{H}}\right)^\dagger \Big| \boldsymbol{x}_0\right]. \tag{63}$$

[6]In the sense of [17].

With (62) and (59), we get

$$D(P_{Y|X=\boldsymbol{x}_0}||P_{Y|X=0}) = \frac{E[||\boldsymbol{H}\boldsymbol{x}_0||^2]}{N_0}\log e - \log\det\left(\boldsymbol{I} + \frac{1}{N_0}\mathrm{cov}(\boldsymbol{H}\boldsymbol{x}_0|\boldsymbol{x}_0)\right). \tag{64}$$

Note that the second term on the right-hand side of (64) is equal to $I(\boldsymbol{H}; \boldsymbol{H}\boldsymbol{x}_0 + \boldsymbol{n})$.

Proceeding as before we get

$$\begin{aligned}\dot{C}(0) &= \sup_{P_{\boldsymbol{x}}}\left\{\frac{E[||\boldsymbol{H}\boldsymbol{x}||^2]}{E[||\boldsymbol{x}||^2]} - \frac{N_0 E\left[\log_e \det\left(\boldsymbol{I} + \frac{1}{N_0}\mathrm{cov}(\boldsymbol{H}\boldsymbol{x}|\boldsymbol{x})\right)\right]}{E[||\boldsymbol{x}||^2]}\right\} \\ &= \sup_{P_{\boldsymbol{x}}}\frac{E[||\boldsymbol{H}\boldsymbol{x}||^2]}{E[||\boldsymbol{x}||^2]} = \mathsf{G}\end{aligned} \tag{65}$$

as we wanted to show. To write (65) we have used the fact that the solution to the maximal gain problem is scale invariant. So we can amplify $||\boldsymbol{x}||$ by a factor large enough to render the nuisance term on the left-hand side of (65) as small as desired because of the logarithmic increase of its numerator.

Finally, we lift the restriction that the channel coefficients (unknown to the receiver) are Gaussian, which was, in fact, a worst case restriction. If the "true" conditional output distribution is denoted by $P_{Y|X=\boldsymbol{x}}$, let us denote the Gaussian distribution with identical mean and covariance by $\Phi_{Y|X=\boldsymbol{x}}$. In particular, note that

$$P_{Y|X=0} = \Phi_{Y|X=0}.$$

We can write the divergence of interest as

$$\begin{aligned}&D\left(P_{Y|X=\boldsymbol{x}_0}\middle\|P_{Y|X=0}\right) \\ &= D\left(\Phi_{Y|X=\boldsymbol{x}_0}\middle\|\Phi_{Y|X=0}\right) + D\left(P_{Y|X=\boldsymbol{x}_0}\middle\|\Phi_{Y|X=\boldsymbol{x}_0}\right)\end{aligned} \tag{66}$$

because the expectation of the log-likelihood ratio in the first divergence of the right-hand side only involves the second-order statistics. Since the second divergence in (66) is nonnegative, the case of Gaussian channel coefficients unknown to the receiver is indeed the least favorable. But even in that worst case we were able to achieve the same expressions for $\frac{E_b}{N_0}_{\min}$ as in the (best) case where the receiver knows the channel perfectly. Thus, Theorem 1 follows. $\square$

The special case $m = n = 1$ of Theorem 1 merits particular attention

$$\mathsf{y} = (\bar{\mathsf{h}} + \mathsf{g})\mathsf{x} + \mathsf{n} \tag{67}$$

where the deterministic component is denoted by $\bar{\mathsf{h}}$ and the zero-mean random component g has variance γ^2. In this case, regardless of transmitter/receiver knowledge of channel coefficients and the distribution of those coefficients, we get

$$\frac{E_b}{N_0} = \frac{\log_e 2}{|\bar{\mathsf{h}}|^2 + \gamma^2}. \tag{68}$$

Special cases of Theorem 1 dealing with the scalar channel have appeared in the literature. In addition to the references cited in Section I, the case $n = m = 1$, with fading known at the receiver, (55) was explicitly stated in [15], [18]. The scalar real-valued channel, with fading unknown at both transmitter and receiver, appears in [10] in the context of capacity per unit cost. The derivative of capacity of the multiantenna channel with independent data feeding different antenna elements was shown to be given by (53) in [13, eq. (11)].

The fact that with full channel knowledge at the transmitter and power control, the transmitted $\frac{E_b}{N_0}$ can be made as small as desired in the infinite-support scalar flat-fading channel was shown in [15]. In fact, [15] shows that property even if the fading coefficients can only be tracked coarsely. This is another illustration of a case where the conclusions derived from the analysis of spectral efficiency versus $\frac{E_b}{N_0}$ are quite different from those drawn from a cursory analysis of $C(\mathsf{SNR})$.

The restriction in Theorem 1 to independent and identically distributed noise components is not a critical one, of course, as the observation basis can be changed to force that condition. Naturally, care should be taken interpreting the results, as N_0 in the traditional meaning of $\frac{E_b}{N_0}$ measures the total received noise power (cf. (6)); but if the noise components do not have equal strength, the $\frac{E_b}{N_0}_{\min}$ is actually dictated by the lowest strength component. To illustrate this point, it is best to consider the special case $\boldsymbol{H} = \boldsymbol{I}$, and independent noise components. The following result is a straightforward application of the above techniques.

Theorem 2: Consider the m-dimensional complex channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} \tag{69}$$

where the complex Gaussian vector $\boldsymbol{n}$ has independent components with variances

$$E[|n_j|^2] = \sigma_j^2$$

such that

$$\frac{1}{m}\sum_{j=1}^{m}\sigma_j^2 = N_0.$$

The required $\frac{E_b}{N_0}$ for reliable communication is

$$\frac{E_b}{N_0}_{\min} = \frac{\min_j \sigma_j^2}{N_0}\log_e 2. \tag{70}$$

The fact that we no longer get $\log_e 2$ is just an artifact of the meaning of N_0. In fact, the received energy per bit divided by the noise energy restricted to the dimensions used by the optimal signaling, continues to be −1.59 dB.

D. Additive Non-Gaussian Noise

Theorem 1 shows that $\frac{E_b}{N_0}_{\min} = -1.59$ dB is an extremely robust feature of channels where the additive noise is Gaussian. In fact, it can even be generalized to certain nonlinear channels. Even though no explicit results for $C(\mathsf{SNR})$ exist if the noise is not Gaussian, it is indeed possible to find $\frac{E_b}{N_0}_{\min}$ for specific examples. Consider the following result that applies to Laplacian noise.

Theorem 3—Laplacian Noise Channel: Consider the m-dimensional complex channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} \tag{71}$$

where the vector $\boldsymbol{n}$ satisfies (6) and has independent complex components each with distribution[7]

$$p(\mathsf{x}) = \frac{1}{\mathsf{N}_0} \exp\left(-\frac{2}{\sqrt{\mathsf{N}_0}}\left(|\Re \mathsf{x}| + |\Im \mathsf{x}|\right)\right). \tag{72}$$

Then

$$\frac{E_b^r}{N_0}_{\min} = \frac{1}{2}\log_e 2 = -4.60 \text{ dB}. \tag{73}$$

Proof: The result is obviously independent of m, so we can take $m = 1$. A straightforward computation yields (in nats)

$$\begin{aligned} D\left(P_{Y|X=\mathsf{x}_0} \| P_{Y|X=0}\right) = &-2 + \frac{2}{\sqrt{N_0}}\left(|\Re \mathsf{x}_0| + |\Im \mathsf{x}_0|\right) \\ &+ \exp\left(-\frac{2|\Re \mathsf{x}_0|}{\sqrt{N_0}}\right) \\ &+ \exp\left(-\frac{2|\Im \mathsf{x}_0|}{\sqrt{N_0}}\right). \end{aligned} \tag{74}$$

When the result of (74) is divided by $|\mathsf{x}_0|^2$, it is clear that the ratio is maximized as $|\mathsf{x}_0| \to 0$. Using a Taylor series expansion of the exponential in (74), (42) becomes

$$\dot{C}(0) = 2 \tag{75}$$

as we wanted to show. □

Incidentally, note that the bound in (37) is satisfied with equality in the Laplacian case, a property which also holds for the Gaussian channel as long as the channel fading is known at the receiver.

E. Signaling

For several of the scenarios considered above, the capacity-achieving distributions are either unknown as a function of the SNR, or they can be found only numerically. A relevant byproduct of the analysis in this section is the identification of simple distributions that, although do not achieve $C(\mathsf{SNR})$, achieve $\dot{C}(0)$, and, consequently, are good enough to achieve $\frac{E_b}{N_0}_{\min}$. This fact and the results in the next section, motivate the following formalization of the traditional optimality criterion.

Definition 1: An input distribution parametrized by SNR, $\boldsymbol{x}_{\mathsf{SNR}}$ is *first-order optimal* if it satisfies the SNR constraint (12) and it achieves $\frac{E_b}{N_0}_{\min}$, namely,

$$\lim_{\mathsf{SNR}\to 0} \frac{I\left(\boldsymbol{x}_{\mathsf{SNR}}; \boldsymbol{y}\right)}{m\,\mathsf{SNR}} = \dot{C}(0). \tag{76}$$

Trivially, a capacity-achieving distribution is first-order optimal. For the AWGN channel, in addition to the Gaussian distribution, we have seen that the one-dimensional on–off distribution

$$\mathsf{x} = \begin{cases} 0, & \text{with probability } 1-\delta \\ \mathsf{c}, & \text{with probability } \delta \end{cases} \tag{77}$$

with

$$\delta = \mathsf{SNR}\,\frac{N_0}{|\mathsf{c}|^2} \tag{78}$$

is first-order optimal for any nonzero c. More generally, for channel (54), we saw in the proof of Theorem 1 that the following n-dimensional distribution is first-order optimal:

$$\boldsymbol{x} = \begin{cases} 0, & \text{with probability } 1-\delta \\ \rho(\mathsf{SNR})\boldsymbol{x}_0, & \text{with probability } \delta \end{cases} \tag{79}$$

with

$$\delta = \frac{N_0 \mathsf{SNR}}{\rho^2(\mathsf{SNR})\|\boldsymbol{x}_0\|^2} \tag{80}$$

and $\boldsymbol{x}_0$ a maximal-eigenvalue eigenvector of $\boldsymbol{H}^\dagger\boldsymbol{H}$ if $\boldsymbol{H}$ is known, or of the matrix $E[\boldsymbol{H}^\dagger\boldsymbol{H}]$ otherwise. Any $\rho(\mathsf{SNR})$ which lets $\delta \to 0$ can be chosen if $\boldsymbol{H}$ is perfectly known, otherwise, $\rho(\mathsf{SNR}) \to \infty$ as $\mathsf{SNR} \to 0$. In the latter case, the first-order optimal distribution suggests signals with extremely low duty cycle. But note that even in the case of a known channel, the peak-to-average ratio of the input distribution (79) goes to infinity as $\mathsf{SNR} \to 0$. Of course, the peak-to-average ratio of the capacity-achieving Gaussian input is no better. Fortunately, if the channel is known at the receiver, unit peak-to-average ratio[8] signaling strategies such as BPSK and QPSK are first-order optimal. In fact, it does not take much for the input to be first-order optimal.

Theorem 4: Assume that the receiver knows $\boldsymbol{H}$. An input distribution $\boldsymbol{x}_{\mathsf{SNR}}$ which satisfies the SNR constraint (12) is *first-order optimal* if and only if

$$\lim_{\mathsf{SNR}\to 0} \frac{\|E[\boldsymbol{x}_{\mathsf{SNR}}]\|^2}{E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2\right]} = 0 \tag{81}$$

and

$$\lim_{\mathsf{SNR}\to 0} \frac{E[\|\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}\|^2]}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} = \mathsf{G}. \tag{82}$$

Proof: To show sufficiency, first note that when the receiver knows the channel, the mutual information does not depend on the mean of the input. Then, we make use of the canonical decomposition of mutual information as

$$I(X;Y) = D(P_{Y|X=\mathsf{x}}\|P_{Y|X=0}|P_{\mathsf{x}}) - D(P_Y\|P_{Y|X=0}) \tag{83}$$

with $Y = \boldsymbol{H}(\boldsymbol{x}_{\mathsf{SNR}} - E[\boldsymbol{x}_{\mathsf{SNR}}]) + \boldsymbol{n}$, $X = \boldsymbol{x}_{\mathsf{SNR}} - E[\boldsymbol{x}_{\mathsf{SNR}}]$. According to the proof of Theorem 1, the first term on the right-hand side of (83) achieves the target in (76) provided (82) is satisfied. Therefore, it will be sufficient to show that the second term on the right-hand side of (83) is $o(\mathsf{SNR})$ for any fixed nonzero $\boldsymbol{H}$. We can cast this in a more general setting in which $Y = \boldsymbol{v}_{\mathsf{SNR}} + \boldsymbol{n}$, $P_{Y|X=0} = \Phi_{\boldsymbol{n}}$ is a zero-mean complex Gaussian variate with covariance $N_0\boldsymbol{I}$, $\boldsymbol{v}_{\mathsf{SNR}}$ has zero mean, covariance matrix $\boldsymbol{\Sigma}_{\mathsf{SNR}}$, and is independent of $\boldsymbol{n}$, and

$$E[\|\boldsymbol{v}_{\mathsf{SNR}}\|^2] = \mathrm{trace}\left(\boldsymbol{\Sigma}_{\mathsf{SNR}}\right) = \alpha\mathsf{SNR}$$

[7]The real and imaginary parts are denoted by $\mathsf{x} = \Re\mathsf{x} + j\Im\mathsf{x}$.

[8]Once the discrete-time information sequences are mapped to continuous-time waveforms the peak-to-average ratio will no longer be unity.

for some finite constant α. The divergence of interest can be written as

$$D(P_Y \| \Phi_{\boldsymbol{n}}) = D(P_Y \| \Phi_Y) + D(\Phi_Y \| \Phi_{\boldsymbol{n}}) \tag{84}$$

where Φ_Y is a Gaussian variate with the same mean and covariance matrix as Y. The first term on the right-hand side of (84) is the "non-Gaussianness" [19] of the vector $\boldsymbol{v}_{\mathsf{SNR}} + \boldsymbol{n}$. Using the results in [19] it follows that $D(P_Y \| \Phi_Y) = o(\mathsf{SNR})$. The second term on the right-hand side of (84) is equal to

$$\begin{aligned} D(\Phi_Y \| \Phi_{\boldsymbol{n}}) &= D(\mathcal{N}(0, N_0 \boldsymbol{I} + \boldsymbol{\Sigma}_{\mathsf{SNR}}) \| \mathcal{N}(0, N_0 \boldsymbol{I})) \\ &= \operatorname{trace}\left(N_0^{-1} \boldsymbol{\Sigma}_{\mathsf{SNR}}\right) \log e \\ &\quad - \log \det\left(\boldsymbol{I} + N_0^{-1}\boldsymbol{\Sigma}_{\mathsf{SNR}}\right) \\ &\le \frac{1}{2N_0^2} \operatorname{trace}\left(\boldsymbol{\Sigma}_{\mathsf{SNR}}^2\right) \log e \qquad (85) \\ &\le \frac{1}{2N_0^2} \operatorname{trace}^2\left(\boldsymbol{\Sigma}_{\mathsf{SNR}}\right) \log e \\ &= o(\mathsf{SNR}) \qquad (86) \end{aligned}$$

where (85) follows from the fact that for any nonnegative matrix $\boldsymbol{A}$

$$\log_e \det(\boldsymbol{I} + \boldsymbol{A}) \ge \operatorname{trace}(\boldsymbol{A}) - \frac{1}{2}\operatorname{trace}(\boldsymbol{A}^2). \tag{87}$$

The necessity of (81) is clear from the fact that since the mutual information is invariant to the mean of the input we can only hope to achieve (76) if any power wasted in the mean is negligible

$$\lim_{\mathsf{SNR}\to 0} \frac{E[\|\boldsymbol{x}_{\mathsf{SNR}} - E[\boldsymbol{x}_{\mathsf{SNR}}]\|^2]}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} = 1. \tag{88}$$

Furthermore, the necessity of achieving the maximal channel gain (82) is evident from the proof of Theorem 1. □

In the setting of the scalar AWGN channel, [20] and [21] showed that any zero-mean input constellation asymptotically maximizes cutoff rate and channel capacity, respectively, in the traditional sense of meeting the first derivative.

The rank-one signaling strategy whereby all the energy is transmitted along one eigenmode of $\boldsymbol{H}^\dagger \boldsymbol{H}$ (or of its expectation if the matrix is unknown) is sometimes referred to as *beamforming*. In the special case of a single receive antenna, beamforming was shown to be asymptotically optimal for vanishing SNR in [22] (see also [23]) using the traditional (first-order) optimality criterion. In fact, if the largest eigenvalue eigenspace has dimension one, then beamforming is optimum up to a certain nonzero SNR, explicitly computed in [24]. However, if the largest eigenvalue eigenspace has dimension larger than one, then rank-one signaling is wasteful of bandwidth as will be seen in Section V-D.

Turning our attention to the case where the channel is unknown at the receiver, it is convenient to define the following class of input signals.

Definition 2: An input distribution $\boldsymbol{x}_{\mathsf{SNR}}$ is said to be *flash signaling* if it satisfies the SNR constraint (12)

$$E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2] = m N_0 \mathsf{SNR} \tag{89}$$

and for all $\nu > 0$

$$\lim_{\mathsf{SNR}\to 0} \frac{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\| > \nu\}]}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} = 1. \tag{90}$$

The simplest form of flash signaling is the on–off signaling we saw in (79) with the on-level $\rho(\mathsf{SNR}) \to \infty$ as $\mathsf{SNR} \to 0$. The first appearance of this specific form of flash signaling as a capacity-maximizing strategy is apparently due to [25] in the context of the flat Rayleigh channel with fading coefficients unknown at the receiver. In general, flash signaling is the mixture of a probability distribution that asymptotically concentrates all its mass at 0 and a probability distribution that migrates to infinity; the weight of the latter vanishes sufficiently fast to satisfy the vanishing power constraint.

Theorem 5: Flash signaling that satisfies

$$\lim_{\mathsf{SNR}\to 0} \frac{E[\|\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}\|^2]}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} = \mathsf{G} \tag{91}$$

is first-order optimal.

Proof: Choose an arbitrary $\epsilon > 0$. Define

$$\begin{aligned} \boldsymbol{v}^+ &= E[\boldsymbol{x}_{\mathsf{SNR}} 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon\}] &(92)\\ \boldsymbol{v}^- &= E[\boldsymbol{x}_{\mathsf{SNR}} 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\| \le \epsilon\}] &(93)\\ \zeta^+ &= E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon\}] &(94)\\ \zeta^- &= E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\| \le \epsilon\}]. &(95) \end{aligned}$$

Then

$$\begin{aligned} \frac{\|E[\boldsymbol{x}_{\mathsf{SNR}}]\|^2}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} &= \frac{\|\boldsymbol{v}^+ + \boldsymbol{v}^-\|^2}{\zeta^+ + \zeta^-} &(96)\\ &\le \frac{2\|\boldsymbol{v}^+\|^2 + 2\|\boldsymbol{v}^-\|^2}{\zeta^+ + \zeta^-} &(97)\\ &\le \frac{2\|\boldsymbol{v}^+\|^2}{\zeta^+} + \frac{2\|\boldsymbol{v}^-\|^2}{\zeta^-}\frac{\zeta^-}{\zeta^+ + \zeta^-} &(98)\\ &\le \frac{2\|\boldsymbol{v}^+\|^2}{\zeta^+} + 2\frac{\zeta^-}{\zeta^+ + \zeta^-} &(99)\\ &\le \frac{2\|\boldsymbol{v}^+\|^2}{\zeta^+} + o(1) &(100)\\ &\le 2P[\|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon] \frac{\|E[\boldsymbol{x}_{\mathsf{SNR}} | \|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon]\|^2}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 | \|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon]} + o(1) &(101)\\ &\le 2P[\|\boldsymbol{x}_{\mathsf{SNR}}\| > \epsilon] + o(1) &(102)\\ &\le o(1) &(103) \end{aligned}$$

where (99) and (102) follow from the fact that the norm squared of the mean is less than or equal to the mean of the norm squared, and both (100) and (103) follow from (90). Thus, condition (81) is satisfied. Using Theorem 4, first-order optimality is ensured provided that the receiver knows $\boldsymbol{H}$.

Now let us assume that the receiver does not know $\boldsymbol{H}$. First-order optimality is equivalent to

$$\frac{I(\boldsymbol{x}_{\mathsf{SNR}};\boldsymbol{y})}{m\mathsf{SNR}} = \frac{N_0}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]}\left(E\left[D\left(P_{Y|X=\boldsymbol{x}_{\mathsf{SNR}}}\middle\| P_{Y|X=0}\right)\right] - D\left(P_Y \| P_{Y|X=0}\right)\right) \tag{104}$$

$$\rightarrow \mathsf{G} \tag{105}$$

where the conditional divergence (in nats) is equal to [cf. (64)]

$$\frac{E[||\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}||^2]}{N_0} - E\left[\log_e \det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right)\right].$$

Using the same argument followed in the proof of Theorem 4 we can show that the unconditional divergence in (104) vanishes faster than SNR. In view of (91), it remains to show that

$$\lim_{\mathsf{SNR}\to 0} \frac{E\left[\log_e \det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right)\right]}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} = 0. \tag{106}$$

Recalling the notation in (63) and using the Hadamard inequality, we can write for any realization of $\boldsymbol{x}_{\mathsf{SNR}}$

$$\frac{1}{m}\log\det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right) \le \frac{1}{m}\sum_{i=1}^{m}\log\left(1 + \frac{1}{N_0}E[|(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}})_i|^2|\boldsymbol{x}_{\mathsf{SNR}}]\right) \tag{107}$$

$$\le \log\left(1 + \frac{1}{mN_0}E[||\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}||^2|\boldsymbol{x}_{\mathsf{SNR}}]\right) \tag{108}$$

$$\le \log\left(1 + \frac{\lambda_{\max}}{mN_0}||\boldsymbol{x}_{\mathsf{SNR}}||^2\right) \tag{109}$$

where (108) follows from Jensen's inequality and we have used the shorthand notation $\lambda_{\max}$ for the maximal eigenvalue of $E[\boldsymbol{H}^\dagger\boldsymbol{H}]$. We will upper-bound the expectation of the right-hand side of (109) with respect to $\boldsymbol{x}_{\mathsf{SNR}}$ by splitting it according to whether $||\boldsymbol{x}_{\mathsf{SNR}}|| < \nu$ for an arbitrary $\nu > 0$. Using $\log_e(1+x) \le x$

$$E\left[\log_e\left(1 + \frac{\lambda_{\max}}{mN_0}||\boldsymbol{x}_{\mathsf{SNR}}||^2\right)1\{||\boldsymbol{x}_{\mathsf{SNR}}|| < \nu\}\right] \le \frac{\lambda_{\max}}{mN_0}E\left[||\boldsymbol{x}_{\mathsf{SNR}}||^2 1\{||\boldsymbol{x}_{\mathsf{SNR}}|| < \nu\}\right] \tag{110}$$

which vanishes faster than $E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]$ because of (90).

On the other hand

$$E\left[\log_e\left(1 + \frac{\lambda_{\max}}{mN_0}||\boldsymbol{x}_{\mathsf{SNR}}||^2\right)1\{||\boldsymbol{x}_{\mathsf{SNR}}|| > \nu\}\right] \le \frac{1}{\nu^2}\log_e\left(1 + \frac{\nu^2\lambda_{\max}}{mN_0}\right)E\left[||\boldsymbol{x}_{\mathsf{SNR}}||^2 1\{||\boldsymbol{x}_{\mathsf{SNR}}|| > \nu\}\right] \tag{111}$$

where (111) follows because $\frac{1}{x}\log(1+x)$ is monotonically decreasing. In view of (90), the ratio of (111) to $E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]$ converges to a constant that can be made as small as desired by choosing ν to be sufficiently large. Thus, the proof of (106) is complete. □

Theorem 6: Assume that neither the transmitter nor the receiver know $\boldsymbol{H}$, and

$$\lambda_{\max}(E[\boldsymbol{H}]^\dagger E[\boldsymbol{H}]) = \lambda_{\max}(E[\boldsymbol{H}^\dagger\boldsymbol{H}]). \tag{112}$$

Then, any input that is first-order optimal for a deterministic channel with matrix $E[\boldsymbol{H}]$ is first-order optimal for $\boldsymbol{H}$.

Proof: We can proceed as in the proof of Theorem 5 until we reach the point where we wish to show that the following ratio vanishes:

$$\frac{E\left[\log_e\det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right)\right]}{mE[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} \le \frac{\log_e\left(1 + \frac{1}{mN_0}(E[||\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}||^2] - ||\overline{\boldsymbol{H}}\boldsymbol{x}_{\mathsf{SNR}}||^2)\right)}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} \tag{113}$$

where we denoted $E[\boldsymbol{H}] = \overline{\boldsymbol{H}}$, and we used the Hadamard and Jensen inequalities. The right-hand side of (113) vanishes as $\mathsf{SNR} \to 0$ because the first-order optimality for the deterministic channel $\overline{\boldsymbol{H}}$ implies that

$$\lim_{\mathsf{SNR}\to 0}\frac{E[||\overline{\boldsymbol{H}}\boldsymbol{x}_{\mathsf{SNR}}||^2]}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} = \lambda_{\max}\left(\overline{\boldsymbol{H}}^\dagger\overline{\boldsymbol{H}}\right) \tag{114}$$

$$= \lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right) \tag{115}$$

$$\ge \frac{E[||\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}||^2]}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} \tag{116}$$

for any SNR. □

Theorem 7: Assume that neither the receiver nor the transmitter know $\boldsymbol{H}$. If

$$\lambda_{\max}\left(E[\boldsymbol{H}]^\dagger E[\boldsymbol{H}]\right) < \lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right) \tag{117}$$

then $\boldsymbol{x}_{\mathsf{SNR}}$ is first-order optimal only if it is flash signaling and (91) is satisfied.

Proof: The upper bound (cf. (104))

$$\frac{I(\boldsymbol{x}_{\mathsf{SNR}};\boldsymbol{y})}{m\mathsf{SNR}} \le \frac{E[||\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}||^2]}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} - \frac{N_0 E\left[\log_e\det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right)\right]}{E[||\boldsymbol{x}_{\mathsf{SNR}}||^2]} \tag{118}$$

immediately implies the necessity of (91). The proof will be complete upon showing that the second term on the right-hand side of (118) is bounded away from 0 if the input is not flash signaling. Note that if $\boldsymbol{A}$ is an $m \times m$ nonnegative definite matrix, then

$$\det(\boldsymbol{I} + \boldsymbol{A}) \ge 1 + \lambda_{\max}(\boldsymbol{A}) \ge 1 + \frac{1}{m}\operatorname{trace}\boldsymbol{A} \tag{119}$$

which when particularized to the determinant in (118) yields

$$\det\left(\boldsymbol{I} + \frac{1}{N_0}\operatorname{cov}(\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}|\boldsymbol{x}_{\mathsf{SNR}})\right) \ge 1 + \frac{1}{mN_0}\boldsymbol{x}_{\mathsf{SNR}}^\dagger\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right] - E\left[\boldsymbol{H}^\dagger\right]E[\boldsymbol{H}]\right)\boldsymbol{x}_{\mathsf{SNR}}$$

$$\ge 1 + \frac{1}{mN_0}\left(\boldsymbol{x}_{\mathsf{SNR}}^\dagger E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\boldsymbol{x}_{\mathsf{SNR}} - \lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger\right]E[\boldsymbol{H}]\right)||\boldsymbol{x}_{\mathsf{SNR}}||^2\right)$$

$$\ge 1 + \frac{\gamma}{2mN_0}||\boldsymbol{x}_{\mathsf{SNR}}||^2 \tag{120}$$

where we have denoted the positive quantity

$$\gamma = \lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger \boldsymbol{H}\right]\right) - \lambda_{\max}\left(E[\boldsymbol{H}]^\dagger E[\boldsymbol{H}]\right) \tag{121}$$

and (120) holds outside the set of realizations

$$\mathcal{B}_{\mathsf{SNR}} = \Big\{\boldsymbol{x}_{\mathsf{SNR}}\colon \boldsymbol{x}_{\mathsf{SNR}}^\dagger E\left[\boldsymbol{H}^\dagger \boldsymbol{H}\right]\boldsymbol{x}_{\mathsf{SNR}} < \left(\lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger \boldsymbol{H}\right]\right) - \gamma/2\right)\|\boldsymbol{x}_{\mathsf{SNR}}\|^2\Big\}. \tag{122}$$

Accordingly, the expectation in the numerator of the second term on the right-hand side of (118) is lower-bounded by

$$\begin{aligned}
&E\left[\log\left(1+\frac{\gamma}{2mN_0}\|\boldsymbol{x}_{\mathsf{SNR}}\|^2\right)1\{\boldsymbol{x}_{\mathsf{SNR}}\notin\mathcal{B}_{\mathsf{SNR}}\}\right]\\
&\geq E\Big[\log\left(1+\frac{\gamma}{2mN_0}\|\boldsymbol{x}_{\mathsf{SNR}}\|^2\right)1\{\|\boldsymbol{x}_{\mathsf{SNR}}\|<\nu\}\\
&\qquad\times 1\{\boldsymbol{x}_{\mathsf{SNR}}\notin\mathcal{B}_{\mathsf{SNR}}\}\Big]\\
&\geq \frac{1}{\nu^2}\log\left(1+\frac{\gamma\nu^2}{2mN_0}\right)E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\|<\nu\}\right.\\
&\qquad\left.\times 1\{\boldsymbol{x}_{\mathsf{SNR}}\notin\mathcal{B}_{\mathsf{SNR}}\}\right]\\
&\geq \frac{1}{\nu^2}\log\left(1+\frac{\gamma\nu^2}{2mN_0}\right)E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\|\boldsymbol{x}_{\mathsf{SNR}}\|<\nu\}]\\
&\qquad-\frac{1}{\nu^2}\log\left(1+\frac{\gamma\nu^2}{2mN_0}\right)\\
&\qquad\times E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\boldsymbol{x}_{\mathsf{SNR}}\in\mathcal{B}_{\mathsf{SNR}}\}\right]
\end{aligned} \tag{123}$$

where we have chosen an arbitrary $\nu > 0$. From the definition of $\mathcal{B}_{\mathsf{SNR}}$, it follows that we can upper-bound

$$\begin{aligned}
&E[\|\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}\|^2]\\
&\leq \lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right)E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\boldsymbol{x}_{\mathsf{SNR}}\notin\mathcal{B}_{\mathsf{SNR}}\}\right]\\
&\quad+\left(\lambda_{\max}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right)-\gamma/2\right)\\
&\quad\times E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\boldsymbol{x}_{\mathsf{SNR}}\in\mathcal{B}_{\mathsf{SNR}}\}\right].
\end{aligned} \tag{124}$$

Upon dividing both sides of (124) by $E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]$, we see that condition (91) requires that

$$\lim_{\mathsf{SNR}\to 0}\frac{E\left[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2 1\{\boldsymbol{x}_{\mathsf{SNR}}\in\mathcal{B}_{\mathsf{SNR}}\}\right]}{E[\|\boldsymbol{x}_{\mathsf{SNR}}\|^2]} = 0. \tag{125}$$

Applying this fact to (123) we reach the conclusion that unless the condition for flash signaling (90) is satisfied, the second term on the right-hand side of (118) is bounded away from 0. □

F. Multiaccess Channels

Another important scenario captured by the setting of this paper is the analysis of the total throughput of additive-noise multiple-access channels. Suppose that the receiver observes

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \tag{126}$$

For consistency with the standard notation in the multiuser literature [26], we denote the dimensionality of (126) by N, which is the number of degrees of freedom per transmitted symbol. The input vector $\boldsymbol{x}$ is K-dimensional where K is the number of users, and the $N \times K$ matrix $\boldsymbol{H}$ can be used to model fading, multiple-access interference, etc. For example, in a code-division multiple-access (CDMA) channel subject to flat fading [15], N is the spreading gain, or number of chips per symbol

$$\boldsymbol{H} = \boldsymbol{S}\boldsymbol{A}$$

where the columns of the matrix $\boldsymbol{S}$ are the spreading vectors and $\boldsymbol{A}$ is a $K \times K$ diagonal matrix of the channel fading coefficients seen by each user. Note that the symbol emanating from each user in this memoryless multiple-access channel is a complex scalar. Naturally, it is possible to generalize (126) by having each user transmit a symbol vector (thereby encompassing multiantenna arrays at the transmitters). This multiaccess setup is different from the one considered in the single-user case in several respects. First, the encoders operate autonomously and are fed independent messages. Thus, the components of $\boldsymbol{x}$ must be statistically independent. Second, it is more natural to define SNR on a per-user basis rather than on a per-received dimension basis as we had done before. Let SNR be the transmitted energy per user per input dimension ($E[|\mathsf{x}_k|^2]$) divided by N_0, and let $C(\mathsf{SNR}) = b/N$ as before, where b is the total number of reliable information bits transmitted by all K users in one channel use. Note that although SNR is common for all users, the individual rates need not be equal. The resulting "system" energy per bit can be seen to be equal to the harmonic mean of the individual energies per bit. Because of the different definition of SNR, instead of (16) it is easy to see that we now have (cf. [27], [15])

$$\frac{E_b}{N_0}C(\mathsf{SNR}) = \beta\mathsf{SNR} \tag{127}$$

where the ratio of users to dimensions is denoted by

$$\beta = \frac{K}{N}. \tag{128}$$

Accordingly, in this case

$$\frac{E_b}{N_0}_{\min} = \lim_{\mathsf{SNR}\to 0}\frac{\beta\mathsf{SNR}}{C(\mathsf{SNR})}. \tag{129}$$

Despite those differences, $\frac{E_b^{\mathrm{r}}}{N_0}_{\min}$ turns out to be the same as before.

Theorem 8: Consider the K-user multiple-access channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{130}$$

where the N-dimensional complex Gaussian vector $\boldsymbol{n}$ has independent and identically distributed components, each with variance N_0. The required $\frac{E_b}{N_0}$ for reliable communication is

$$\frac{E_b}{N_0}_{\min} = \frac{K\log_e 2}{E\left[\mathrm{trace}\left\{\boldsymbol{H}^\dagger\boldsymbol{H}\right\}\right]} \tag{131}$$

which at the receiver is

$$\frac{E_b^{\mathrm{r}}}{N_0}_{\min} = \log_e 2 = -1.59\text{ dB} \tag{132}$$

regardless of whether $\boldsymbol{H}$ is known at the transmitters and/or receiver. Furthermore, if $\boldsymbol{H}$ is known at the receiver, a bank of single-user receivers that approximate multiple-access interference as independent Gaussian noise achieves (132).

Proof: We first consider the case of known $\boldsymbol{H}$. With an overall transmitted energy in one K-vector symbol equal to $K\mathsf{SNR}N_0$, the received signal energy is

$$\mathsf{SNR}N_0\sum_{k=1}^{K}\|\boldsymbol{h}_k\|^2$$

where $\boldsymbol{h}_k$ is the kth column of the matrix $\boldsymbol{H}$. Therefore, the overall channel gain is equal to

$$\frac{E[||\boldsymbol{Hx}||^2]}{E[||\boldsymbol{x}||^2]} = \frac{1}{K}\sum_{k=1}^{K}||\boldsymbol{h}_k||^2 = \frac{1}{K}\,\mathrm{trace}\left\{\boldsymbol{H}^\dagger\boldsymbol{H}\right\} \quad (133)$$

and (132) follows from (131). Now according to (129), in order to prove (131), we need to show that

$$\dot{C}(0) = \beta\frac{1}{K}\sum_{k=1}^{K}||\boldsymbol{h}_k||^2. \quad (134)$$

We first upper-bound $\dot{C}(0)$ by analyzing a receiver consisting of K autonomous receivers each informed of the transmitted codewords of the other users. In such a hypothetical genie-aided setup, the capacity is given by

$$\overline{C}(\mathsf{SNR}) = \frac{1}{N}\sum_{k=1}^{K}\log_2(1+||\boldsymbol{h}_k||^2\mathsf{SNR}) \quad (135)$$

whose derivative at $\mathsf{SNR} = 0$ (in nats) is equal to the sought-after expression on the right-hand side of (134).

To lower-bound $\dot{C}(0)$, we consider a suboptimal receiver consisting of K autonomous receivers each of which treats multiuser interference as Gaussian noise. The capacity achieved by each of those (matched-filter) receivers is lower-bounded by the capacity of an AWGN channel

$$\boldsymbol{y} = \boldsymbol{h}_k\mathsf{x}_k + \tilde{\boldsymbol{n}}_k \quad (136)$$

where $\tilde{\boldsymbol{n}}_k$ has covariance matrix equal to

$$N_0(1+\alpha_k\mathsf{SNR})\boldsymbol{I}$$

with α_k equal to the largest eigenvalue of the covariance matrix of the vector

$$\sum_{j\neq k}\boldsymbol{h}_j\mathsf{x}_j.$$

The capacity lower bound is then

$$\tilde{C}(\mathsf{SNR}) = \frac{1}{N}\sum_{k=1}^{K}\log_2\left(1+\frac{||\boldsymbol{h}_k||^2\mathsf{SNR}}{1+\alpha_k\mathsf{SNR}}\right); \quad (137)$$

a function whose derivative at $\mathsf{SNR} = 0$ is identical to that of the upper bound (135).

To deal with the more general case of unknown $\boldsymbol{H}$, we invoke the multiaccess results shown in [10]. In particular, condition [10, eq. (24)] is satisfied in this setting, namely, the zero-energy input is the most favorable from the viewpoint of other users. In such case, the optimum capacity per unit energy is achieved by time-sharing: at most one user is allowed to transmit at any given time, and a bank of single-user receivers is optimum. The resulting capacity is equal to

$$C(\mathsf{SNR}) = \frac{1}{N}\sum_{k=1}^{K}C_k(\mathsf{SNR}) \quad (138)$$

where $C_k(\mathsf{SNR})$ is the capacity of the single-user channel

$$\boldsymbol{y} = \boldsymbol{h}_k\mathsf{x}_k + \boldsymbol{n}_k \quad (139)$$

which corresponds to the special case $n = 1$ of the channel considered in Theorem 1. Proceeding as in the proof of that theorem, it is easy to show that $\dot{C}(0)$ is given by β times $E[||\boldsymbol{h}_k||^2]$. □

The main practical lesson we can learn from this section is that if the background noise is Gaussian, finding the minimum transmitted energy per bit required for reliable communication is simply a matter of subtracting the channel loss from −1.59 dB. The fact that channels with vastly different information-carrying capabilities have the same $\frac{E_b}{N_0}_{\min}$ points out the essential limitation of this performance measure: it is not intended to give any indication whatsoever about required bandwidth. In order to assess the interplay between bandwidth and power for a given wideband channel, we must resort to the analysis propounded in the next section.

V. Wideband Slope b/s/Hz/(3 dB)

A. The Role of the Second Derivative

In this section, we study the growth of spectral efficiency with $\frac{E_b}{N_0}$ in the wideband regime. Specifically, we focus on the performance measure $\mathcal{S}_0$, defined in (29) as the increase of bits per second per hertz per 3 dB (b/s/Hz/(3 dB)) of E_b achieved at $\frac{E_b}{N_0}_{\min}$. The following general result is a straightforward consequence of (16).

Theorem 9: At $\frac{E_b}{N_0}_{\min}$, the slope of the spectral efficiency versus $\frac{E_b}{N_0}$ in b/s/Hz/(3 dB) is given by

$$\mathcal{S}_0 = \frac{2\left[\dot{C}(0)\right]^2}{-\ddot{C}(0)} \quad (140)$$

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

Proof: By definition of $\dot{C}$ and $\ddot{C}$, the function $C(\mathsf{SNR})$ (in bits per dimension) admits the following Taylor series for small SNR:

$$C(\mathsf{SNR}) = \dot{C}(0)\mathsf{SNR}\log_2 e + \tfrac{1}{2}\ddot{C}(0)\mathsf{SNR}^2\log_2 e + o(\mathsf{SNR}^2). \quad (141)$$

Equations (16) and (141) enable us to write

$$\frac{E_b}{N_0} = \left(\dot{C}(0)\log_2 e + \frac{1}{2}\ddot{C}(0)\mathsf{SNR}\log_2 e + o(\mathsf{SNR})\right)^{-1} \quad (142)$$

and from (35) and (142) we get

$$\frac{\frac{E_b}{N_0}_{\min}}{\frac{E_b}{N_0}} = 1 + \frac{\ddot{C}(0)}{2\dot{C}(0)}\mathsf{SNR} + o(\mathsf{SNR}). \quad (143)$$

The target ratio in (29) is a function of $\frac{E_b}{N_0}$ which can be put as a parametric expression in terms of the value of SNR that solves (16)

$$\begin{aligned}\frac{\mathsf{C}\left(\frac{E_b}{N_0}\right)}{10\log_{10}\frac{E_b}{N_0}\Big/\frac{E_b}{N_0}_{\min}} &= \frac{\dot{C}(0)\mathsf{SNR}\log_2 e + o(\mathsf{SNR})}{-10\log_{10}\left(1+\frac{\ddot{C}(0)}{2\dot{C}(0)}\mathsf{SNR}+o(\mathsf{SNR})\right)}\\ &= \frac{\dot{C}(0)\mathsf{SNR}\log_2 e + o(\mathsf{SNR})}{-10\frac{\ddot{C}(0)}{2\dot{C}(0)}\mathsf{SNR}\log_{10} e + o(\mathsf{SNR})}\\ &= \frac{2\left[\dot{C}(0)\right]^2}{-\ddot{C}(0)}\,\frac{1}{10\log_{10}2} + o(\mathsf{SNR}).\end{aligned} \quad (144)$$

Letting $\frac{E_b}{N_0} \to \frac{E_b}{N_0}_{\min}$, or equivalently, $\mathsf{SNR} \to 0$ we get (140).

So far, we have assumed that $\ddot{C}(0)$ is finite. If $\dot{C}(0) < \infty$ and $\ddot{C}(0) = -\infty$, then it is easy to check that $\mathcal{S}_0 = 0$, for in that case

$$C(\mathsf{SNR}) = \dot{C}(0)\mathsf{SNR}\log_2 e + o(\mathsf{SNR}) \tag{145}$$

while

$$\frac{\frac{E_b}{N_0}_{\min}}{\frac{E_b}{N_0}} - 1 = \frac{C(\mathsf{SNR})\log_e 2}{\dot{C}(0)\mathsf{SNR}} - 1 \tag{146}$$

goes to $-\infty$ when divided by vanishing SNR. □

Note that by definition, the wideband slope $\mathcal{S}_0$ is invariant to channel gain. Thus, it is not necessary to distinguish between transmitted and received $\mathcal{S}_0$. This property would be lost if the slope were defined with respect to $\frac{E_b}{N_0}$ in linear scale. Although such a slope carries much less engineering insight, it admits a formula similar to (140), replacing the square by a cube. Note that to make a fair comparison between the wideband slopes achieved by different receiver/transmitter designs (and, thus, between the bandwidth requirements for given rate and power), it is essential that the transmitted $\frac{E_b}{N_0}_{\min}$ be the same.

In the analysis of wideband channels it is sometimes useful to represent the channel as a bank of ℓ parallel independent subchannels. If the transmitter devotes the same power to all subchannels and each subchannel has the same number of dimensions, then the capacity of the overall channel is

$$C(\mathsf{SNR}) = \frac{1}{\ell}\sum_{i=1}^{\ell} C_i(\mathsf{SNR}). \tag{147}$$

Formula (35) implies that the transmit $\frac{E_b}{N_0}_{\min}$ (in linear scale) is equal to the harmonic mean of the individual transmit $\frac{E_b}{N_0}_{\min}$. In most cases of interest, the receive $\frac{E_b^{\mathrm{r}}}{N_0}_{\min}$ is the same for all subchannels, and, thus, the same for the overall channel. From (35) and (140), it is easy to verify that the product of the squared transmit $\frac{E_b}{N_0}_{\min}$ and the slope of the overall channel is equal to the harmonic mean of such products corresponding to the subchannels. In particular, if the individual transmit $\frac{E_b}{N_0}_{\min}$ and the individual slope of all subchannels are equal, then the overall $\frac{E_b}{N_0}_{\min}$ and slope are identical to those of the subchannels.

In parallel with the definition we made in Section IV we have the following.

Definition 3: An input distribution parametrized by SNR, $\boldsymbol{x}_{\mathsf{SNR}}$ is *second-order optimal* if it is first-order optimal and it achieves $\mathcal{S}_0$. Equivalently, it achieves both the first and second derivatives of capacity

$$\frac{1}{m}\frac{d}{d\mathsf{SNR}} I(\boldsymbol{x}_{\mathsf{SNR}};\boldsymbol{y})|_{\mathsf{SNR}=0} = \dot{C}(0) \tag{148}$$

and

$$\frac{1}{m}\frac{d^2}{d\mathsf{SNR}^2} I(\boldsymbol{x}_{\mathsf{SNR}};\boldsymbol{y})|_{\mathsf{SNR}=0} = \ddot{C}(0). \tag{149}$$

B. On–Off Signaling

In view of the universal first-order optimality of on–off signaling, we study its capabilities in the region of small but nonzero spectral efficiency. We assume a generalized form of on–off signaling where the input vector has a $(1-\delta)$-mass at the all-zero vector. The input distribution conditioned on the input being nonzero is denoted by $\overline{P}_X$, and a unit mass at the all-zero vector is denoted by P_0. Thus,

$$P_X = (1-\delta)P_0 + \delta\overline{P}_X \tag{150}$$

with δ chosen so that the SNR constraint is satisfied, namely,

$$\delta = \frac{\mathsf{SNR}N_0 m}{E[\|\boldsymbol{x}\|^2|\boldsymbol{x}\neq 0]}. \tag{151}$$

The output distribution corresponding to $\overline{P}_X$ is denoted by

$$\overline{P}_Y = \int P_{Y|X=\boldsymbol{x}}\, d\overline{P}_X(\boldsymbol{x}). \tag{152}$$

Theorem 10: Denote Pearson's $\mathcal{X}$-divergence by

$$\Delta(Q\|P) \stackrel{\mathrm{def}}{=} E_P\left(\frac{dQ}{dP} - 1\right)^2. \tag{153}$$

The $\frac{E_b}{N_0}_{\min}$ and $\mathcal{S}_0$ achieved by generalized on–off signaling (150) are given by

$$\frac{E_b}{N_0}_{\min}\left(\overline{P}_X\right) = \frac{E[\|\boldsymbol{x}\|^2|\boldsymbol{x}\neq 0]\log_e 2}{N_0 D\left(P_{Y|X}\,\|P_{Y|X=0}\,|\overline{P}_X\right)} \tag{154}$$

and

$$\mathcal{S}_0\left(\overline{P}_X\right) = \frac{2}{m}\frac{\left(D\left(P_{Y|X}\,\|P_{Y|X=0}\,|\overline{P}_X\right)\right)^2}{\Delta\left(\overline{P}_Y\|P_{Y|X=0}\right)} \tag{155}$$

respectively, where divergence is measured in nats in both (154) and (155).

Proof: In this case, the role played by $C(\mathsf{SNR})$ is taken by the function

$$\frac{1}{m} I(X;Y)$$

where the input is the distribution (150). The first and second derivatives of this function (in nats) with respect to SNR will be denoted by $\dot{\mathcal{C}}(0)$ and $\ddot{\mathcal{C}}(0)$, respectively. From the results of [10], already used in Section IV, we obtain (154). To obtain (155), we use

$$\mathcal{S}_0\left(\overline{P}_X\right) = \frac{2\left[\dot{\mathcal{C}}(0)\right]^2}{-\ddot{\mathcal{C}}(0)} \tag{156}$$

with

$$\dot{\mathcal{C}}(0) = \frac{N_0 D\left(P_{Y|X}\,\|P_{Y|X=0}\,|\overline{P}_X\right)}{E[\|\boldsymbol{x}\|^2|\boldsymbol{x}\neq 0]} \tag{157}$$

and

$$\ddot{\mathcal{C}}(0) = 2\lim_{\mathsf{SNR}\to 0}\frac{\frac{1}{m}I(X;Y) - \dot{\mathcal{C}}(0)\mathsf{SNR}}{\mathsf{SNR}^2}. \tag{158}$$

Decomposing the input–output mutual information as

$$I(X;Y) = D\left(P_{Y|X=\mathsf{x}}\,\|P_{Y|X=0}\,|\,P_{\mathsf{x}}\right) - D\left(P_Y\,\|P_{Y|X=0}\right) \tag{159}$$

with

$$P_Y = (1-\delta)P_{Y|X=0} + \delta\overline{P}_Y \tag{160}$$

and using (151), the right-hand side of (158) becomes

$$\ddot{C}(0) = 2 \lim_{\mathsf{SNR}\to 0} \frac{-D\left(P_Y \,\|P_{Y|X=0}\right)}{m\,\mathsf{SNR}^2} = -\frac{2mN_0^2}{(E[\|\boldsymbol{x}\|^2|\boldsymbol{x}\neq 0])^2} \lim_{\delta\to 0} \frac{1}{\delta^2} D\left(P_Y \,\|P_{Y|X=0}\right). \tag{161}$$

The desired expression (155) follows by applying the following general result to the case $P = P_{Y|X=0}$ and $Q = \overline{P}_Y$:

$$\lim_{\delta\to 0} \frac{D((1-\delta)P+\delta Q\|P)}{\delta^2} = \frac{1}{2}\Delta(P\|Q)\log e. \tag{162}$$

To verify (162) assume natural logarithms without loss of generality and note that by definition of divergence

$$D((1-\delta)P+\delta Q\|P) = E[(1+\delta q(Y))\log_e(1+\delta q(Y))] = E[r(\delta q(Y))] \tag{163}$$

where the expectation is with respect to Y distributed according to P, and we have defined

$$q(y) = \frac{dQ}{dP}(y) - 1 \tag{164}$$

and

$$r(x) = (1+x)\log_e(1+x) - x. \tag{165}$$

Note that in order to write (163) we used

$$E[q(Y)] = 0. \tag{166}$$

Furthermore

$$\Delta(P\|Q) = E[q^2(Y)]. \tag{167}$$

Since $r(x) \geq 0$ for all $x > -1$, and $r(x) = \frac{x^2}{2} + o(x^3)$. Fatou's lemma leads to

$$\lim_{\delta\to 0} \frac{D((1-\delta)P+\delta Q\|P)}{\delta^2} \geq \frac{1}{2}E[q^2(Y)]. \tag{168}$$

On the other hand, for any $\gamma > 0$, there exists a sufficiently small $\delta_0 > 0$ such that

$$r(x) \leq (1+\gamma)\frac{x^2}{2} \tag{169}$$

whenever $x > -\delta_0$. Since $q(y) \geq -1$

$$\lim_{\delta\to 0} \frac{D((1-\delta)P+\delta Q\|P)}{\delta^2} \leq \frac{1+\gamma}{2}E[q^2(Y)]. \tag{170}$$

But γ can be made arbitrarily small. Thus, (162) follows.[9] □

C. AWGN Channel

Before we proceed to study the wideband slope of fading channels it is instructive to deal with the scalar AWGN channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}. \tag{171}$$

Directly from (21) and (140) we obtain that the wideband slope for the AWGN channel is

$$\mathcal{S}_0 = 2 \text{ b/s/Hz/(3 dB)} \tag{172}$$

[9]The same technical argument can be used elsewhere in the paper regarding expectations of Taylor series.

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.

In the case of binary quantization of both the real and imaginary components of the output of the AWGN channel, the capacity is [28], [29]

$$C(\mathsf{SNR}) = 2\log 2 - 2h(Q(\sqrt{\mathsf{SNR}})) \tag{173}$$

with

$$h(x) = -x\log x - (1-x)\log(1-x)$$

and

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-t^2/2)\,dt.$$

The first and second derivatives (in nats) of (173) are

$$\dot{C}(0) = \frac{2}{\pi} \tag{174}$$

$$\ddot{C}(0) = \frac{4}{3\pi}\left(\frac{1}{\pi} - 1\right) \tag{175}$$

which result in

$$\frac{E_b}{N_0}_{\min} = 0.37 \text{ dB} \tag{176}$$

$$\mathcal{S}_0 = \frac{6}{\pi - 1} = 2.8 \text{ b/s/Hz/(3 dB)}. \tag{177}$$

Returning to the unquantized channel, as a simple exercise we apply the formulas obtained in Theorem 10 to the Gaussian channel. Suppose we use on–off signaling with on-level $\mathsf{x} = \mathsf{x}_0$. Using (44) and

$$\Delta(\mathcal{N}(\mathsf{m},\sigma^2)\|\mathcal{N}(0,\sigma^2)) = \exp\left(\frac{2|\mathsf{m}|^2}{\sigma^2}\right) - 1 \tag{178}$$

we get

$$\frac{E_b}{N_0}_{\min}\left(\frac{|\mathsf{x}_0|^2}{N_0}\right) = \log_e 2 \tag{179}$$

and

$$\mathcal{S}_0\left(\frac{|\mathsf{x}_0|^2}{N_0}\right) = \frac{2}{N_0^2}\,\frac{|\mathsf{x}_0|^4}{\exp\left(\frac{2|\mathsf{x}_0|^2}{N_0}\right) - 1} \tag{180}$$

$$\leq 0.3238 \text{ b/s/Hz/(3 dB)} \tag{181}$$

where the upper bound is achieved at

$$\frac{|\mathsf{x}_0|^2}{N_0} = 0.7968. \tag{182}$$

Thus, even though the analysis in Section IV showed that any on-level is equally good as far as achieving $\frac{E_b}{N_0}_{\min}$ for the Gaussian channel, the finer analysis in this section shows otherwise. More importantly, unless bandwidth is infinite, on–off signaling is decidedly inefficient for the AWGN channel. By comparing (172) and (181), we see that on–off signaling requires 618% the minimum bandwidth.

The simple on–off signaling strategy above can be modified so that information is encoded not only at the times in which energy is sent but also in the phase. To that end, let us assume that $\overline{P}_X$ is uniformly distributed on $\{|\mathsf{x}| = A\}$ for some fixed real $A > 0$. Applying Theorem 10 to this input we get

$$\frac{E_b}{N_0}_{\min}\left(\frac{A^2}{N_0}\right) = \log_e 2 \tag{183}$$

and

$$\mathcal{S}_0\left(\frac{A^2}{N_0}\right) = \frac{2A^4}{N_0^2}\,\frac{1}{I_0\left(\frac{2A^2}{N_0}\right)-1} \leq 2 \tag{184}$$

where the upper bound is tight for $A^2/N_0 \to 0$. Thus, this type of generalized on–off signaling which uses phase modulation can achieve wideband slope as close as desired to the optimal one.

However, among all second-order optimal signaling strategies, QPSK is the best practical choice in the wideband coherent regime (when the receiver knows the channel). Before dealing with the general case, we show the second-order optimality of QPSK in the simple setting of the AWGN.

Theorem 11: Consider the scalar AWGN channel

$$\mathsf{y} = \mathsf{x} + \mathsf{n}. \tag{185}$$

The wideband slopes achieved by various first-order optimal distributions are as follows.

1) BPSK achieves wideband slope equal to 1 b/s/Hz/(3 dB).
2) QPSK achieves wideband slope equal to 2 b/s/Hz/(3 dB).
3) Any signaling distribution that can be written as a mixture of (rotated and scaled) QPSK constellations achieves wideband slope equal to 2 b/s/Hz/(3 dB).

Proof: First-order optimality of the input distributions in the statement of the result follows from the fact that they have zero mean (Theorem 4). 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

$$C_{\mathrm{QPSK}}(\mathsf{SNR}) = 2\,C_{\mathrm{BPSK}}(\mathsf{SNR}/2). \tag{186}$$

Moreover, it follows from (186) and (16) that the spectral efficiencies are related by

$$\mathsf{C}_{\mathrm{QPSK}}\left(\frac{E_b}{N_0}\right) = 2\,\mathsf{C}_{\mathrm{BPSK}}\left(\frac{E_b}{N_0}\right). \tag{187}$$

Thus, QPSK achieves twice the spectral efficiency of BPSK at any $\frac{E_b}{N_0}$, and, consequently, twice the wideband slope.

The BPSK distribution is

$$P_{\mathsf{x}} = \frac{1}{2}\,\delta_A + \frac{1}{2}\,\delta_{-A} \tag{188}$$

where A is (without loss of generality) a real scalar that satisfies

$$\mathsf{SNR} = A^2. \tag{189}$$

Note that for convenience and without impacting the results we have chosen $N_0 = 1$. From (44) and (188) we get

$$D\left(P_{Y|X=\mathsf{x}}\,\middle\|\,P_{Y|X=0}\middle|\,P_{\mathsf{x}}\right) = \mathsf{SNR}\log e. \tag{190}$$

The unconditional output distribution is now

$$P_Y = \tfrac{1}{2}\,\mathcal{N}(A,1) + \tfrac{1}{2}\,\mathcal{N}(-A,1) \tag{191}$$

and

$$D\left(P_Y\,\middle\|\,P_{Y|X=0}\right) = E[h(\mathsf{y})\log h(\mathsf{y})] \tag{192}$$

where the expectation is with respect to $\mathcal{N}(0,1)$ and

$$h(\mathsf{y}) = \exp(-A^2)\cosh(2A\Re\mathsf{y}). \tag{193}$$

Using the fact that $2A\Re\mathsf{y}$ is a real Gaussian random variable with variance $2\mathsf{SNR}$, it is straightforward to show that

$$E[h(\mathsf{y})\log h(\mathsf{y})] = \mathsf{SNR}^2\log e + o(\mathsf{SNR}^2). \tag{194}$$

Thus, the mutual information achieved by BPSK satisfies (in nats)

$$I(X;Y) = \mathsf{SNR} - \mathsf{SNR}^2 + o(\mathsf{SNR}^2) \tag{195}$$

or equivalently

$$\dot{C}_{\mathrm{BPSK}}(0) = 1 \tag{196}$$

and

$$\ddot{C}_{\mathrm{BPSK}}(0) = -2. \tag{197}$$

Substituting these values into (140) we conclude that the wideband slope of BPSK is 1 b/s/Hz/(3 dB), and thus, the wideband slope of QPSK is 2 b/s/Hz/(3 dB).

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. □

D. Perfect Receiver Side Information

Theorem 12: Consider the m-dimensional complex channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{198}$$

where the complex Gaussian vector $\boldsymbol{n}$ has independent and identically distributed components. Suppose that the receiver knows $\boldsymbol{H}$, and that the transmitter knows $\boldsymbol{H}$ but has no ability to do power control (or, equivalently, it knows the maximal eigenvalue eigenspace of $\boldsymbol{H}^\dagger\boldsymbol{H}$ but not the maximal eigenvalue). Then

$$\mathcal{S}_0 = \frac{2\ell}{m\,\kappa(\sigma_{\max}(\boldsymbol{H}))} \tag{199}$$

with the *kurtosis*[10] of a random variable Z defined as

$$\kappa(Z) = \frac{E[Z^4]}{E^2[Z^2]} \tag{200}$$

$\sigma_{\max}(\boldsymbol{H})$ denotes the maximal singular value of $\boldsymbol{H}$, and ℓ is equal to the multiplicity of $\sigma_{\max}(\boldsymbol{H})$.

Furthermore, the optimum wideband slope (199) can be achieved by QPSK modulating with equal power the ℓ orthogonal dimensions of the maximal-eigenvalue eigenspace of $\boldsymbol{H}^\dagger\boldsymbol{H}$.

[10]The "amount of fading" defined in [30] is equal to the kurtosis minus 1. See [30], [15] for tables of standard fading distributions.

Proof: In the absence of power control, the water-filling formula implies that if SNR is so small that the water level only covers the deepest level, capacity is given by

$$C(\mathsf{SNR}) = \frac{\ell}{m} E\left[\log_2\left(1 + \frac{m}{\ell}\lambda_{\max}(\boldsymbol{H}^\dagger\boldsymbol{H})\mathsf{SNR}\right)\right] \tag{201}$$

whose derivatives in nats are equal to

$$\dot{C}(0) = E\left[\lambda_{\max}\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)\right] \tag{202}$$

and

$$\ddot{C}(0) = -\frac{m}{\ell} E\left[\lambda^2_{\max}\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)\right]. \tag{203}$$

Using (140) and $\lambda_{\max}(\boldsymbol{H}^\dagger\boldsymbol{H}) = \sigma^2_{\max}(\boldsymbol{H})$ we get (199).

To achieve (201), the input must be a zero-mean Gaussian vector whose components have equal variance along the orthogonal directions of the maximal-eigenvalue eigenspace of $\boldsymbol{H}^\dagger\boldsymbol{H}$, and zero otherwise. The fact that QPSK-modulating each of those dimensions is also second-order optimal follows immediately from Theorem 11, since those dimensions are orthogonal. □

The special case $m = n = 1$ of Theorem 12 was given in [15]. 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 12 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 $\boldsymbol{H}$ grows, while keeping a constant ratio, and its coefficients are independent and identically distributed with variance ζ^2, then the maximal singular value converges to a deterministic constant [31]

$$\sigma^2_{\max}(m^{-1/2}\boldsymbol{H}) \to \left(1 + \sqrt{\frac{n}{m}}\right)^2 \zeta^2 \tag{204}$$

and its multiplicity goes to 1. Accordingly, if m and n represent the number of receive and transmit antennas, respectively,[11] and the transmitter knows the eigenstructure of $\boldsymbol{H}$, 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. This slope is obtained at a value of $\frac{E_b}{N_0}_{\min}$ that decreases with the number of antennas as

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{E[\sigma^2_{\max}(\boldsymbol{H})]} \approx \frac{\log_e 2}{\zeta^2\left(\sqrt{m}+\sqrt{n}\right)^2}. \tag{205}$$

Together with the above result on the asymptotic insensitivity of the slope to the number of antennas, (205) implies that with many transmit and receive antennas, doubling the number of transmit and receive antennas halves the required power for fixed rate and bandwidth, provided the channel is known at the transmitter (cf. [32]).

However, in the multiantenna literature it is much more common to assume that the transmitter has no knowledge whatsoever of $\boldsymbol{H}$ and that transmit antennas are fed by independent equal-power streams. In this case, we saw in Section IV that

$$\frac{E_b}{N_0}_{\min} = \frac{n\log_e 2}{\operatorname{trace}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right)}. \tag{206}$$

The corresponding wideband slope is given by the following result.

Theorem 13: Consider the m-dimensional complex channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{207}$$

where the complex Gaussian vector $\boldsymbol{n}$ has independent and identically distributed components. Suppose that the receiver knows the $m \times n$ matrix $\boldsymbol{H}$, but the transmitter has no knowledge of the channel matrix (or its statistics). Then

$$\mathcal{S}_0 = \frac{2}{m}\frac{\left(\operatorname{trace}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right)\right)^2}{\operatorname{trace}\left(E\left[\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)^2\right]\right)}. \tag{208}$$

Proof: In this case, the optimum input distribution is an n-dimensional Gaussian vector with independent and identically distributed components achieving capacity [33], [13]

$$C(\mathsf{SNR}) = \frac{1}{m} E\left[\log\det\left[\boldsymbol{I} + \frac{m}{n}\boldsymbol{H}^\dagger\boldsymbol{H}\mathsf{SNR}\right]\right]. \tag{209}$$

Formula (208) follows from (140) upon taking the first and second derivatives of (209). To that end, the following formulas can be derived easily from the definition of determinant. If $\boldsymbol{A}$ is an $n \times n$ matrix, then

$$\frac{d}{du}\log\det[\boldsymbol{I} + u\boldsymbol{A}]|_{u=0} = \operatorname{trace}(\boldsymbol{A})\log e \tag{210}$$

$$\frac{d^2}{du^2}\log\det[\boldsymbol{I} + u\boldsymbol{A}]|_{u=0} = -\operatorname{trace}(\boldsymbol{A}^2)\log e. \tag{211}$$

□

Apropos of the expression in (208), note that (aside from the factor $2/m$) the numerator is the square of the expected Frobenius (or Euclidean) norm squared of $\boldsymbol{H}$, whereas the denominator is equal to the expected Frobenius norm squared of $\boldsymbol{H}^\dagger\boldsymbol{H}$

$$\operatorname{trace}\left(E\left[\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)^2\right]\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\left|\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)_{ij}\right|^2\right]. \tag{212}$$

If the entries of $\boldsymbol{H}$ are independent zero-mean random variables with variance ζ^2, then it can be checked that (206) and (208) reduce to

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{m\zeta^2} \tag{213}$$

and

$$\mathcal{S}_0 = \frac{2n}{\kappa(|H_{ij}|) + m + n - 2} \quad \text{b/s/Hz/(3 dB)/receive antenna} \tag{214}$$

where m and n play the role of the number of receive and transmit antennas, respectively.[12] Furthermore, if the channel coefficients are complex Gaussian random variables, $|H_{ij}|$ follows the Rayleigh distribution whose kurtosis is equal to 2.

[11] Caution: "M transmit and N receive antennas" is common notation in the contemporary literature on the capacity of multielement arrays.

[12] The effect of antenna correlation on the required bandwidth is explored in [34].

Under those assumptions, in the wideband regime, the spectral efficiency is a multiple of the harmonic mean of the number of receive and transmit antennas

$$\mathcal{S}_0 = \frac{2nm}{m+n} \quad \text{b/s/Hz/(3 dB)}. \tag{215}$$

The low-SNR slope (215) is greater than or equal to the high-SNR slope, $\min\{m, n\}$ [13], with equality if and only if $m = n$. If $\min\{m, n\}$ is held fixed and $\max\{m, n\} \to \infty$, then (215) goes to $2\min\{m, n\}$. The result in (215) should be contrasted with the misconception that in the low-SNR regime capacity is not affected by the number of transmit antennas (e.g., [16]). While $\frac{E_b}{N_0}_{\min}$ (213) does not depend on the number of transmit antennas, suppose that we fix the number of receive antennas m and the power and the data rate, then a system with one transmit antenna requires $(m+1)/2$ times the bandwidth of a system with m transmit antennas, which in turn requires twice the bandwidth of a system with a very large number of transmit antennas.

Note that since the slope of the spectral efficiency versus $\frac{E_b}{N_0}$ curve is rather constant for a fairly wide range of $\frac{E_b}{N_0}$, approximating spectral efficiency by

$$\mathsf{C} \approx \frac{2}{3.01}\,\frac{nm}{n+m}\left(10\log_{10}\frac{E_b^{\mathrm{r}}}{N_0} + 1.59\right) \quad \text{b/s/Hz} \tag{216}$$

would be pretty accurate even for ambitious b/s/Hz values provided the harmonic mean of the number of receive and transmit antennas is large enough. Note that (215) obtained without knowing the eigenstructure of the channel does not contradict the limiting result we obtained above (2 b/s/Hz/(3 dB)) with knowledge of the structure of the channel, because the transmitted power required to achieve -1.59 dB at the receiver is lower in the latter case. Indeed, taking the ratio of (205) and (213), we arrive at the conclusion that knowing the eigenstructure of the channel at the transmitter implies an asymptotic power reduction factor of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_i^2(\boldsymbol{H})}{\sigma_{\max}^2(\boldsymbol{H})} \to \frac{m}{(\sqrt{m}+\sqrt{n})^2} \tag{217}$$

where the right-hand side holds in the many-antenna limit. Note that with the same number of antennas at transmitter and receiver, knowledge of the channel at the transmitter gives a gain of 6 dB at zero spectral efficiency, or equivalently, a factor of 4 in rate (b/s) for the same power in the infinite bandwidth limit [35], or a factor of 4 in the required number of antennas.

Theorem 14: Under the conditions of Theorem 13, 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

$$\boldsymbol{x} = \sqrt{\frac{\mathsf{SNR}m}{n}}\begin{bmatrix} e^{j\phi_1} \\ \vdots \\ e^{j\phi_n}\end{bmatrix} \tag{218}$$

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,

$$E[\|\boldsymbol{x}\|^2] = m\mathsf{SNR} \tag{219}$$

and

$$E\left[\boldsymbol{x}\boldsymbol{x}^\dagger\right] = \frac{m\mathsf{SNR}}{n}\boldsymbol{I}. \tag{220}$$

This input distribution attains the following mutual information:

$$\begin{aligned}\frac{1}{m}I(\boldsymbol{x};\, \boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}) &= \frac{1}{m}D\left(P_{Y|X}\,\|P_{Y|X=0}\,|\,P_X\right) - \frac{1}{m}D\left(P_Y\,\|P_{Y|X=0}\right) \qquad (221)\\ &= \frac{1}{m}E[\|\boldsymbol{H}\boldsymbol{x}\|^2] - \frac{1}{m}D\left(P_Y\,\|P_{Y|X=0}\right). \qquad (222)\end{aligned}$$

Furthermore, it follows from (220) that

$$E[\|\boldsymbol{H}\boldsymbol{x}\|^2] = \frac{m\mathsf{SNR}}{n}\mathrm{trace}\left(E\left[\boldsymbol{H}^\dagger\boldsymbol{H}\right]\right). \tag{223}$$

Therefore, from the result obtained in (208), the desired second-order optimality is equivalent to showing that

$$\lim_{\mathsf{SNR}\to 0}\frac{D\left(P_Y\,\|P_{Y|X=0}\right)}{m\mathsf{SNR}^2} = \frac{m}{2n^2}\,\mathrm{trace}\left(E\left[\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right)^2\right]\right). \tag{224}$$

To accomplish this, note that the divergence in (224) can be expressed as

$$\begin{aligned}D\left(P_Y\,\|P_{Y|X=0}\right) &= E[Z(\boldsymbol{n}, \boldsymbol{H})\log Z(\boldsymbol{n}, \boldsymbol{H})]\\ &= E[Z(\boldsymbol{n}, \boldsymbol{H}) - 1]\\ &\quad + \frac{1}{2}E[(Z(\boldsymbol{n}, \boldsymbol{H}) - 1)^2]\\ &\quad + o(\mathsf{SNR}^2)\end{aligned} \tag{225}$$

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

$$\frac{dP_Y}{dP_{Y|X=0}} = Z(\boldsymbol{n}, \boldsymbol{H}) = E[\exp(-\|\boldsymbol{H}\boldsymbol{x}-\boldsymbol{n}\|^2 + \|\boldsymbol{n}\|^2)|\boldsymbol{H}, \boldsymbol{n}] \tag{226}$$

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

$$E[Z(\boldsymbol{n}, \boldsymbol{H}) - 1|\boldsymbol{H}] = -1 + \int \frac{dP_Y}{dP_{Y|X=0}}\,dP_{Y|X=0} = 0. \tag{227}$$

A Taylor series expansion of the exponential in (226) together with the fact that $E[\boldsymbol{n}] = 0$ results in

$$\begin{aligned}Z(\boldsymbol{n}, \boldsymbol{H}) - 1 = &-E[\|\boldsymbol{H}\boldsymbol{x}\|^2|\boldsymbol{H}]\\ &+ \frac{1}{2}E\left[\left(\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^\dagger\boldsymbol{H}^\dagger\boldsymbol{n} - \|\boldsymbol{H}\boldsymbol{x}\|^2\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right]\\ &+ o(\mathsf{SNR})\end{aligned} \tag{228}$$

where all the expectations are with respect to $\boldsymbol{x}$. Let us consider each term on the right-hand side of (228) individually. Using (220) we get

$$E[\|\boldsymbol{H}\boldsymbol{x}\|^2|\boldsymbol{H}] = \frac{m\mathsf{SNR}}{n}\,\mathrm{trace}\left(\boldsymbol{H}^\dagger\boldsymbol{H}\right) \tag{229}$$

and

$$\begin{aligned}&E\left[\left(\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^\dagger\boldsymbol{H}^\dagger\boldsymbol{n} - \|\boldsymbol{H}\boldsymbol{x}\|^2\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right]\\ &\quad = E[\|\boldsymbol{H}\boldsymbol{x}\|^4|\boldsymbol{H}] + E\left[\left(\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^\dagger\boldsymbol{H}^\dagger\boldsymbol{n}\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right].\end{aligned} \tag{230}$$

The second term on the right-hand side of (230) is equal to

$$E\left[\left(\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{x} + \boldsymbol{x}^\dagger\boldsymbol{H}^\dagger\boldsymbol{n}\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right] = \frac{2m\mathsf{SNR}}{n}\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{H}^\dagger\boldsymbol{n} \tag{231}$$

because

$$E\left[\left(\boldsymbol{n}^\dagger\boldsymbol{H}\boldsymbol{x}\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right] = E\left[\left(\boldsymbol{x}^\dagger\boldsymbol{H}^\dagger\boldsymbol{n}\right)^2\middle|\,\boldsymbol{H}, \boldsymbol{n}\right] = 0 \tag{232}$$

as can be seen using the QPSK nature of the independent components of $\boldsymbol{x}$. Using (228)–(231) we can write

$$\begin{aligned}\frac{n^2 E[(Z(\boldsymbol{n}, \boldsymbol{H}) - 1)^2|\boldsymbol{H}]}{m^2 \mathsf{SNR}^2} &= E\left[\left(\operatorname{trace}\left(\boldsymbol{H}^\dagger \boldsymbol{H}\right) - \boldsymbol{n}^\dagger \boldsymbol{H}\boldsymbol{H}^\dagger \boldsymbol{n}\right)^2 \middle| \boldsymbol{H}\right] + o(\mathsf{SNR}^2)\\ &= E\left[\left(\boldsymbol{n}^\dagger \boldsymbol{H}\boldsymbol{H}^\dagger \boldsymbol{n}\right)^2 \middle| \boldsymbol{H}\right] - \operatorname{trace}^2\left(\boldsymbol{H}^\dagger \boldsymbol{H}\right) + o(\mathsf{SNR}^2).\end{aligned} \tag{233}$$

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

$$E[|\mathsf{n}_i|^4] = 2 \tag{234}$$

it is fairly straightforward to check that

$$E\left[\left(\boldsymbol{n}^\dagger \boldsymbol{M}\boldsymbol{n}\right)^2\right] = \operatorname{trace}^2(\boldsymbol{M}) + \operatorname{trace}(\boldsymbol{M}^2). \tag{235}$$

Applying formula (235) to (233) with $\boldsymbol{M} = \boldsymbol{H}\boldsymbol{H}^\dagger$, the desired result (224) follows. □

Note that the only property of QPSK used in the proof of Theorem 14 is its proper complex (i.e., rotationally invariant) nature.

If in contrast to Theorems 12 and 13, the transmitter knows $E[\boldsymbol{H}^\dagger \boldsymbol{H}]$, then it is easy to find the wideband slope using the foregoing methods. Suppose that $\boldsymbol{v}$ is the unique maximal-eigenvalue eigenvector of $E[\boldsymbol{H}^\dagger \boldsymbol{H}]$, then the wideband slope is

$$\mathcal{S}_0 = \frac{2}{m\kappa(||\boldsymbol{H}\boldsymbol{v}||)}. \tag{236}$$

More generally, if the rank of the maximal-eigenvalue eigenspace of $E[\boldsymbol{H}^\dagger \boldsymbol{H}]$ is $\ell > 1$, then the wideband slope can be found by applying the result in Theorem 13 to a modified channel matrix $\boldsymbol{H}\boldsymbol{V}$, where the $n \times \ell$ matrix $\boldsymbol{V}$ consists of the ℓ orthonormal eigenvectors of the maximal-eigenvalue eigenspace, since first-order optimality requires that the input vector restrict itself to that eigenspace. In the case of $\ell > 1$, contrary to the misconception obtained using the traditional optimality criterion [22], rank-one signaling, i.e., beamforming (see Section IV-E), is not wideband optimal. This is made evident by considering the special case of one receive antenna, n transmit antennas, and independent equal-variance Gaussian channel coefficients. Then, the bandwidth required by rank-one signaling is $2n/(1 + n)$ times the bandwidth required by optimum rank-n signaling (which transmits equal-power independent streams through each antenna).

Note that if $n = 1$, $\boldsymbol{H}^\dagger \boldsymbol{H}$ is a scalar, and Theorems 12 and 13 have the same assumptions. Then, (199), (208), and (236) boil down to

$$\mathcal{S}_0 = \frac{2}{m\kappa(||\boldsymbol{H}||)}. \tag{237}$$

As an application of (237) we have the following formula for the flat Ricean channel.

Theorem 15: Consider the $m = n = 1$ Ricean fading channel

$$\mathsf{y} = \left(\bar{\mathsf{h}} + \mathsf{g}\right)\mathsf{x} + \mathsf{n} \tag{238}$$

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

$$\mathcal{S}_0 = \frac{1}{1 - \frac{1}{2}\left(1 + \frac{\gamma^2}{|\bar{\mathsf{h}}|^2}\right)^{-2}}. \tag{239}$$

Proof: When the receiver knows the channel coefficients we just need to specialize Theorems 12 or 13 (in the scalar case) to the case $m = n = 1$ in which they lead to the same result. Formula (239) follows from the kurtosis of the Ricean distribution which is equal to

$$\kappa\left(\left|\bar{\mathsf{h}} + \mathsf{g}\right|\right) = 2 - \frac{\left|\bar{\mathsf{h}}\right|^4}{\left(\left|\bar{\mathsf{h}}\right|^2 + \gamma^2\right)^2}. \tag{240}$$

□

E. Imperfect Receiver Side Information

While receiver side information of the channel fading does not improve $\frac{E_b}{N_0}_{\min}$, it has a drastic effect on the required bandwidth in the wideband regime as shown in this subsection.

Theorem 16: First-order optimal flash signaling achieves $\mathcal{S}_0 = 0$ even if the receiver knows the channel.

Proof: According to Theorem 9, we need to show that flash signaling achieves $\ddot{C}(0) = -\infty$. Since $\dot{C}(0)$ is achieved, in order to show

$$\lim_{\mathsf{SNR}\to 0} \frac{I(X; Y) - \dot{C}(0)\mathsf{SNR}}{\mathsf{SNR}^2} = -\infty \tag{241}$$

we will show that

$$\lim_{\mathsf{SNR}\to 0} \frac{D(P_Y || P_{Y|X=0})}{\mathsf{SNR}^2} = \infty. \tag{242}$$

The unconditional output distribution in (242) is

$$P_Y = ((1 - \delta(\mathsf{SNR}))P_\delta + \delta(\mathsf{SNR})Q_\delta \tag{243}$$

where P_δ and Q_δ denote the convolution of

$$P_{Y|X=0} = \mathcal{N}(0, N_0 \boldsymbol{I})$$

with $P_{\boldsymbol{s}_\delta}$ and $Q_{\boldsymbol{s}_\delta}$, which denote the distributions of $\boldsymbol{H}\boldsymbol{x}_{\mathsf{SNR}}$ conditioned on $||\boldsymbol{x}_{\mathsf{SNR}}|| < \nu$ and $||\boldsymbol{x}_{\mathsf{SNR}}|| \geq \nu$, respectively, with arbitrarily small ν. Since the receiver knows the channel, it will suffice to show that (241) is satisfied for all nonzero deterministic $\boldsymbol{H}$.

First-order optimal flash signaling requires that $\delta(\mathsf{SNR}) = o(\mathsf{SNR})$ and

$$\lim_{\mathsf{SNR}\to 0} \frac{\delta(\mathsf{SNR})E[||\boldsymbol{s}_\delta||^2]}{\mathsf{SNR}} = \lambda > 0. \tag{244}$$

To show (242), note that we can write

$$\frac{D(P_Y || P_{Y|X=0})}{\mathsf{SNR}^2} = \frac{E[(1 + \mathsf{SNR}(W + V))\log(1 + \mathsf{SNR}(W + V))]}{\mathsf{SNR}^2} \tag{245}$$

where we have introduced the zero-mean random variable

$$V = \frac{\delta(\mathsf{SNR})}{\mathsf{SNR}} \left(\frac{dQ_\delta}{dP_{Y|X=0}}(\boldsymbol{n}) - 1 \right) \tag{246}$$

which satisfies the vanishing lower bound

$$V \geq -\frac{\delta(\mathsf{SNR})}{\mathsf{SNR}} \tag{247}$$

and the zero-mean random variable

$$W = \frac{1-\delta(\mathsf{SNR})}{\mathsf{SNR}} \left(\frac{dP_\delta}{dP_{Y|X=0}}(\boldsymbol{n}) - 1 \right). \tag{248}$$

In order to show that (245) goes to infinity, we will show that $E[(V+W)^2]$ diverges as $\mathsf{SNR} \to 0$. In order to streamline notation, we will take without loss of generality $N_0 = 1$ in the remainder of the proof.

The ratio of densities appearing in (246) can be written as

$$\frac{dQ_\delta}{dP_{Y|X=0}}(\boldsymbol{n}) = e^{\|\boldsymbol{n}\|^2} \int e^{-\|\boldsymbol{n}-\boldsymbol{s}\|^2} dQ_{\boldsymbol{s}_\delta} \tag{249}$$

a random variable whose second moment is equal to

$$\begin{aligned} &E\left[\left(\frac{dQ_\delta}{dP_{Y|X=0}}(\boldsymbol{n}) \right)^2 \right] \\ &= \frac{1}{\pi^m} \int e^{\|\boldsymbol{n}\|^2} \left(\int e^{-\|\boldsymbol{n}-\boldsymbol{s}\|^2} dQ_{\boldsymbol{s}_\delta} \right)^2 d\boldsymbol{n} \\ &= \frac{1}{\pi^m} \iiint e^{(-\|\boldsymbol{n}-\boldsymbol{s}-\hat{\boldsymbol{s}}\|^2 + \|\boldsymbol{s}+\hat{\boldsymbol{s}}\|^2 - \|\boldsymbol{s}\|^2 - \|\hat{\boldsymbol{s}}\|^2)} dQ_{\boldsymbol{s}_\delta}\, dQ_{\boldsymbol{s}_\delta}\, d\boldsymbol{n} \\ &= E\left[\exp\left(\boldsymbol{s}_\delta^\dagger \hat{\boldsymbol{s}}_\delta + \hat{\boldsymbol{s}}_\delta^\dagger \boldsymbol{s}_\delta \right) \right] \end{aligned} \tag{250}$$

where $\boldsymbol{s}_\delta$ and $\hat{\boldsymbol{s}}_\delta$ are independent and have identical distribution $Q_{\boldsymbol{s}_\delta}$. Using (244), we see that $E[V^2] \to \infty$ is equivalent to

$$\frac{E\left[\exp\left(\boldsymbol{s}_\delta^\dagger \hat{\boldsymbol{s}}_\delta + \hat{\boldsymbol{s}}_\delta^\dagger \boldsymbol{s}_\delta \right) \right]}{E^2[\|\boldsymbol{s}_\delta\|^2]} \to \infty. \tag{251}$$

To show (251), we quantize the unit ball of the m Cartesian product of the complex field into "phase bins" that are sufficiently fine so that if we fix a sufficiently small $\epsilon > 0$, we can find $\eta > 0$ such that for all sufficiently small δ, a phase bin $\mathcal{P}_\delta$ can be found so that

$$E[\|\boldsymbol{s}_\delta\|^2 | \boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta] > \epsilon E[\|\boldsymbol{s}_\delta\|^2] \tag{252}$$

$$P[\boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta] > \eta \tag{253}$$

and if $\boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta$, $\hat{\boldsymbol{s}}_\delta / \|\hat{\boldsymbol{s}}_\delta\| \in \mathcal{P}_\delta$, then

$$\boldsymbol{s}_\delta^\dagger \hat{\boldsymbol{s}}_\delta + \hat{\boldsymbol{s}}_\delta^\dagger \boldsymbol{s}_\delta > \eta \|\boldsymbol{s}_\delta\|\, \|\hat{\boldsymbol{s}}_\delta\|. \tag{254}$$

Using (252)–(254), the numerator in (251) is lower-bounded by

$$\begin{aligned} &\eta^2 E\left[\exp\left(\boldsymbol{s}_\delta^\dagger \hat{\boldsymbol{s}}_\delta + \hat{\boldsymbol{s}}_\delta^\dagger \boldsymbol{s}_\delta \right) | \boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta,\, \hat{\boldsymbol{s}}_\delta / \|\hat{\boldsymbol{s}}_\delta\| \in \mathcal{P}_\delta \right] \\ &\quad > \eta^2 E\left[\exp\left(\eta \|\boldsymbol{s}_\delta\|\, \|\hat{\boldsymbol{s}}_\delta\| \right) | \boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta,\, \hat{\boldsymbol{s}}_\delta / \|\hat{\boldsymbol{s}}_\delta\| \in \mathcal{P}_\delta \right] \\ &\quad > \frac{\eta^6}{4!} E^2\left[\|\boldsymbol{s}_\delta\|^4 \, | \boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta \right] \\ &\quad \geq \frac{\eta^6}{4!} E^4\left[\|\boldsymbol{s}_\delta\|^2 \, | \boldsymbol{s}_\delta / \|\boldsymbol{s}_\delta\| \in \mathcal{P}_\delta \right] \\ &\quad \geq \frac{\eta^6}{4!} \epsilon^4 E^4\left[\|\boldsymbol{s}_\delta\|^2 \right]. \end{aligned} \tag{255}$$

Dividing (255) by $E^2[\|\boldsymbol{s}_\delta\|^2]$, we obtain (251) in view of (244). By analyzing $E[W^2]$ and $E[VW]$ using similar methods it can be shown that $E[(V+W)^2]$ diverges and the proof of (242) is complete. □

Theorem 17: If neither the receiver nor the transmitter know $\boldsymbol{H}$ and

$$\lambda_{\max}\left(E[\boldsymbol{H}]^\dagger E[\boldsymbol{H}] \right) < \lambda_{\max}\left(E[\boldsymbol{H}^\dagger \boldsymbol{H}] \right) \tag{256}$$

then

$$\mathcal{S}_0 = 0.$$

Proof: According to Theorem 7 under condition (256), flash signaling satisfying (91) is necessary for first-order optimality, and hence for second-order optimality. But, according to Theorem 16, that kind of signaling achieves $\mathcal{S}_0 = 0$, regardless of whether the receiver knows the channel. □

Theorem 18: If neither the receiver nor the transmitter know $\boldsymbol{H}$ and

$$\lambda_{\max}\left(E[\boldsymbol{H}]^\dagger E[\boldsymbol{H}] \right) = \lambda_{\max}\left(E[\boldsymbol{H}^\dagger \boldsymbol{H}] \right) \tag{257}$$

then

$$\mathcal{S}_0 = \frac{2\ell}{m} \tag{258}$$

where ℓ is the multiplicity of the maximal singular value of $E[\boldsymbol{H}]$.

Proof: In view of Theorem 6, we can restrict the search for second-order optimum inputs to those inputs that are first-order optimal for a deterministic channel with matrix $E[\boldsymbol{H}]$. Then the problem becomes a special case of the one solved in Theorem 12 where the channel matrix therein is deterministic. □

The capacity-achieving distribution for the flat Rayleigh channel in which neither transmitter nor receiver know the fading coefficients is shown in [36] to be discrete with a finite number of masses, which depends on SNR. Furthermore, [36] shows that there exists $\epsilon > 0$ such that $\mathsf{SNR} < \epsilon$, implies that a two- mass distribution is optimal with one mass at zero, and the other at a point whose magnitude goes to infinity as $\mathsf{SNR} \to 0$ and whose phase is irrelevant. Note that the latter statement is consistent with the first-order and second-order optimality of flash signaling.

If the receiver does not have full knowledge of $\boldsymbol{H}$, then condition (256) is usually satisfied. For example, for the flat Ricean channel (238), the left- and right-hand sides of (256) are $|\bar{\mathsf{h}}|^2$ and $|\bar{\mathsf{h}}|^2 + \gamma^2$, respectively. Accordingly, if the receiver does not know the Rayleigh coefficients, the wideband slope is zero, no matter how small γ^2, in contrast to (239). Thus, it is very demanding in terms of bandwidth to achieve $\frac{E_b^{\mathsf{r}}}{N_0}_{\min}$ close to -1.59 dB in the Ricean channel, regardless of the relative strengths of the specular and Rayleigh components. To illustrate the burden of communicating in the wideband regime through channels with zero wideband slope, numerical results[13] indicate that to achieve spectral efficiency equal to 0.01 b/s/Hz we require $\frac{E_b^{\mathsf{r}}}{N_0} \approx$ 0.44 dB for a Ricean channel with $|\bar{\mathsf{h}}| = \gamma$. Fig. 4 shows the tremendous impact of noncoherence in the wideband regime for the special case of the Rayleigh channel, an impact that may not be apparent from a plot of the ratio of the capacities (with and without channel knowledge) as a function of SNR [36]. We see in Fig. 4 that the insensitivity of $\frac{E_b^{\mathsf{r}}}{N_0}_{\min}$ to lack of knowledge of the channel at the receiver is of little relevance to practice. Sometimes the statement that wideband capacity is not affected

[13]Obtained by M. Gursoy, private communication.

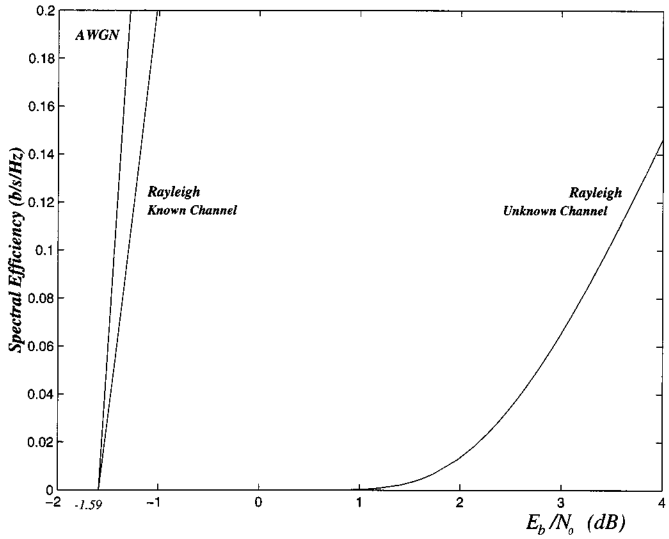

Fig. 4. Spectral efficiency of the AWGN channel and the Rayleigh flat fading channel with and without receiver knowledge of fading coefficients.

by knowledge of the channel is toned down by claiming that this insensitivity may hold for impractically large bandwidths. However, the bandwidth penalty due to lack of channel knowledge is equal to a factor of 1000 at $\frac{E_b^r}{N_0} = 1.25$ dB, and goes to infinity as we move closer to $\frac{E_b^r}{N_0}_{\min}$. Fig. 4 also illustrates (199): when the fading coefficients are known at the receiver, the bandwidth required by Rayleigh fading is twice the bandwidth required in the absence of fading.

Another case in which (256) is satisfied is the block-constant multiantenna model of [37], which is encompassed by the model in Section II by letting m be the number of received antennas times the block duration T and n be the number of transmit antennas times the block duration. Then, $\boldsymbol{H}$ becomes a block-diagonal matrix with all $T\ n \times m$ diagonal blocks being identical.

In many practical cases in which the specular component is not negligible, QPSK is an attractive suboptimal alternative as the following result shows.

Theorem 19: Consider the Ricean channel (238) with $\bar{\mathsf{h}} \neq 0$ and a receiver that does not know the Rayleigh coefficients. Then QPSK achieves

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{|\bar{\mathsf{h}}|^2} \tag{259}$$

and

$$\mathcal{S}_0 = \frac{2}{1 + 2\frac{\gamma^2}{|\bar{\mathsf{h}}|^2}}. \tag{260}$$

Proof: With the Ricean channel

$$P_{Y|X=\mathsf{x}_0} = \mathcal{N}(\bar{\mathsf{h}}\mathsf{x}_0,\, N_0 + \gamma^2|\mathsf{x}_0|^2) \tag{261}$$

and the QPSK input distribution

$$P_{\mathsf{x}} = \frac{1}{4}\delta_{\mathsf{A}(1+j)} + \frac{1}{4}\delta_{\mathsf{A}(1-j)} + \frac{1}{4}\delta_{\mathsf{A}(-1+j)} + \frac{1}{4}\delta_{\mathsf{A}(-1-j)} \tag{262}$$

we have

$$\mathsf{SNR} = \frac{2|\mathsf{A}|^2}{N_0} \tag{263}$$

and we obtain (in nats)

$$\begin{aligned} D(P_{Y|X}\|P_{Y|X=0}|P_X) &= (\gamma^2+|\bar{\mathsf{h}}|^2)\mathsf{SNR} - \log_e(1+\gamma^2\mathsf{SNR}) \\ &= |\bar{\mathsf{h}}|^2\mathsf{SNR} + \frac{\gamma^4}{2}\mathsf{SNR}^2 + o(\mathsf{SNR}^2). \end{aligned} \tag{264}$$

From the formulas for $\frac{E_b}{N_0}_{\min}$ and $\mathcal{S}_0$, both (259) and (260) will follow upon showing that

$$\lim_{\mathsf{SNR}\to 0} \frac{D(P_Y\|P_{Y|X=0})}{\mathsf{SNR}^2} = \frac{1}{2}(\gamma^2 + |\bar{\mathsf{h}}|^2)^2 \tag{265}$$

where

$$\begin{aligned} P_Y = &\frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(1+j),\, \alpha^2 N_0) + \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(1-j),\, \alpha^2 N_0) \\ &+ \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(-1+j),\, \alpha^2 N_0) \\ &+ \frac{1}{4}\mathcal{N}(\bar{\mathsf{h}}\mathsf{A}(-1-j),\, \alpha^2 N_0) \end{aligned} \tag{266}$$

and

$$\alpha^2 = 1 + \mathsf{SNR}\gamma^2. \tag{267}$$

The desired divergence is

$$D(P_Y\|P_{Y|X=0}) = E[\log q(\mathsf{y})] \tag{268}$$

where the expectation is with respect to (266) and

$$\begin{aligned} q(\mathsf{y}) &\stackrel{\text{def}}{=} \frac{dP_Y}{dP_{Y|X=0}}(\mathsf{y}) \\ &= \frac{1}{\alpha^2}\exp\left(-\frac{2|\bar{\mathsf{h}}|^2|\mathsf{A}|^2}{\alpha^2 N_0}\right)\exp\left(-\frac{|\mathsf{y}|^2}{N_0}(\alpha^{-2}-1)\right) \\ &\quad \cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Re\mathsf{y}\right)\cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Im\mathsf{y}\right) \end{aligned} \tag{269}$$

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 (269) we obtain

$$\begin{aligned} D(P_Y\|P_{Y|X=0}) = &(|\bar{\mathsf{h}}|^2 + \gamma^2)\mathsf{SNR} - \frac{2|\bar{\mathsf{h}}|^2\mathsf{SNR}}{1+\gamma^2\mathsf{SNR}} \\ &- \log_e(1+\gamma^2\mathsf{SNR}) \\ &+ E\left[\log_e\cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Re\mathsf{y}\right)\right] \\ &+ E\left[\log_e\cosh\left(2\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\Im\mathsf{y}\right)\right]. \end{aligned} \tag{270}$$

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

$$N_0/2 + |\mathsf{A}|^2\gamma^2 = \frac{\alpha^2 N_0}{2}. \tag{271}$$

Thus, the second and fourth moments of the random variables in the argument of the hyperbolic cosines in (270) are equal to

$$4\frac{|\bar{\mathsf{h}}|^2|\mathsf{A}|^2}{\alpha^4 N_0^2}\left(\frac{\alpha^2 N_0}{2} + |\bar{\mathsf{h}}|^2|\mathsf{A}|^2\right) = \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\left(1 + \frac{|\bar{\mathsf{h}}|^2\mathsf{SNR}}{\alpha^2}\right) \tag{272}$$

and

$$\frac{16|\bar{\mathsf{h}}|^4|\mathsf{A}|^4}{\alpha^8 N_0^4}\frac{3\alpha^4 N_0^2}{4} + o(\mathsf{SNR}^2) = \frac{3|\bar{\mathsf{h}}|^4\mathsf{SNR}^2}{\alpha^4} + o(\mathsf{SNR}^2). \tag{273}$$

Using (272) and (273), and

$$\log_e(\cosh(x)) = \frac{x^2}{2} - \frac{x^4}{12} + o(x^4) \quad (274)$$

each of the expectations in (270) satisfy

$$E\left[\log_e \cosh\left(2\,\frac{|\bar{\mathsf{h}}||\mathsf{A}|}{\alpha^2 N_0}\,\Re \mathsf{y}\right)\right] = \frac{1}{2}\,\frac{|\bar{\mathsf{h}}|^2 \mathsf{SNR}}{\alpha^2}\left(1 + \frac{|\bar{\mathsf{h}}|^2 \mathsf{SNR}}{2\alpha^2}\right) + o(\mathsf{SNR}^2). \quad (275)$$

Thus,

$$\begin{aligned} D(P_Y \| P_{Y|X=0}) &= |\bar{\mathsf{h}}|^2 \mathsf{SNR} + \frac{\gamma^4}{2}\mathsf{SNR}^2 - \frac{|\bar{\mathsf{h}}|^2 \mathsf{SNR}}{\alpha^2} + \frac{|\bar{\mathsf{h}}|^4 \mathsf{SNR}^2}{2\alpha^4} + o(\mathsf{SNR}^2) \\ &= (\gamma^2 + |\bar{\mathsf{h}}|^2)^2 \frac{\mathsf{SNR}^2}{2} + o(\mathsf{SNR}^2) \quad (276) \end{aligned}$$

thereby establishing (265). □

Regarding the wideband slope when the transmitter but not the receiver knows the channel, first note that when the channel is fully known at the transmitter and power control is allowed, the minimum energy per bit is zero in the usual fading models, and, thus, the wideband slope is zero (regardless of whether the receiver knows the channel). If the transmitter knows the maximal-eigenvalue eigenspace (but not the maximal eigenvalue) and the receiver does not know the channel, then first-order signaling requires flash signaling along the directions of the maximal-eigenvalue eigenspace (for simplicity, assume that $E[\boldsymbol{H}] = 0$). Once that signaling is used, the wideband slope is zero because even if the receiver had side information of the maximal-eigenvalue eigenspace, we would be in a situation equivalent to $\boldsymbol{H} = \mathsf{A}\boldsymbol{I}$ where A is a zero-mean random variable unknown to both receiver and transmitter: a channel which is encompassed by Theorem 17.

To conclude this subsection, we give a new general bound on mutual information for a given arbitrary input distribution, which holds for all signal-to-noise ratios. We give the bound in a general setting that encompasses the linear model treated in this paper as well as nonlinear channel models, which may be of interest in optical-fiber transmission and neurobiology.

Theorem 20: Suppose that $\boldsymbol{g}(\boldsymbol{x})$ denotes a complex random vector, which conditioned on $\boldsymbol{x}$ is complex Gaussian with mean

$$\bar{\boldsymbol{g}}(\boldsymbol{x}) = E[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x}] \quad (277)$$

and covariance matrix

$$\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x}) = E[(\boldsymbol{g}(\boldsymbol{x}) - \bar{\boldsymbol{g}}(\boldsymbol{x}))(\boldsymbol{g}(\boldsymbol{x}) - \bar{\boldsymbol{g}}(\boldsymbol{x}))^\dagger|\boldsymbol{x}]. \quad (278)$$

Denote the covariance matrix of the conditional mean vector by

$$\operatorname{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x})) = E[(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\bar{\boldsymbol{g}}(\boldsymbol{x})])(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\bar{\boldsymbol{g}}(\boldsymbol{x})])^\dagger]. \quad (279)$$

Let $\boldsymbol{n}$ be a complex Gaussian vector whose components are independent, with independent real and imaginary parts each with variance $N_0/2$. Then

$$I(\boldsymbol{x};\, \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}) \le \operatorname{trace}(\operatorname{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x})))\,\frac{\log e}{N_0} + E\left[\operatorname{trace}\left(\operatorname{cov}^2(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})\right)\right]\frac{\log e}{2N_0^2}. \quad (280)$$

Proof: To show this result, we need to use a more general form of the decomposition in (159). For any probability measure for which $P_Y \ll Q$

$$I(X;\,Y) = D(P_{Y|X=\mathsf{x}} \| Q | P_{\mathsf{x}}) - D(P_Y \| Q). \quad (281)$$

In the present case

$$P_{Y|X=\mathsf{x}} = \mathcal{N}\left(\bar{\boldsymbol{g}}(\boldsymbol{x});\, N_0\boldsymbol{I} + \operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})\right) \quad (282)$$

and we choose

$$Q = \mathcal{N}(E[\bar{\boldsymbol{g}}(\boldsymbol{x})];\, N_0\boldsymbol{I}). \quad (283)$$

Using (59), the conditional divergence in (281) evaluated with (282) and (283) becomes (in nats)

$$\begin{aligned} D(P_{Y|X=\mathsf{x}} \| Q | P_{\mathsf{x}}) &= \frac{1}{N_0} E[\|\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\bar{\boldsymbol{g}}(\boldsymbol{x})]\|^2] \\ &\quad - E[\log_e \det\left(\boldsymbol{I} + N_0^{-1}\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})\right) \\ &\quad + E[\operatorname{trace}\left(N_0^{-1}\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})\right) \quad (284) \\ &\le \frac{1}{N_0}\operatorname{trace}(\operatorname{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x}))) \\ &\quad + \frac{1}{2N_0^2} E\left[\operatorname{trace}\left(\operatorname{cov}^2(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})\right)\right] \quad (285) \end{aligned}$$

where (285) follows from (87). Since the second term on the right-hand side of (281) is nonnegative, (280) follows from (285). □

Particularizing Theorem 20 to the linear Ricean fading channel where

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x} \quad (286)$$

and $\boldsymbol{H}$ is Gaussian with mean $\overline{\boldsymbol{H}}$, independent of $\boldsymbol{x}$, we obtain the bound

$$I(\boldsymbol{x};\, \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) \le E\left[\|\bar{\boldsymbol{H}}(\boldsymbol{x} - E[\boldsymbol{x}])\|^2\right]\frac{\log e}{N_0} + E\left[\operatorname{trace}\left(\operatorname{cov}^2(\boldsymbol{H}\boldsymbol{x}|\boldsymbol{x})\right)\right]\frac{\log e}{2N_0^2} \quad (287)$$

which in the Rayleigh special case $\bar{\boldsymbol{H}} = 0$ has been found recently in [38], [39].

Although (287) holds for all SNR, it is not a tight bound. In fact, it is very coarse at all but small SNR and it does not give the exact asymptotic second-order behavior for vanishing SNR. Under quite general conditions on the input and the channel (including non-Gaussian channels), it is shown in [40] that

$$\begin{aligned} I(\boldsymbol{x};\, \boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}) &= E\left[\|\bar{\boldsymbol{H}}(\boldsymbol{x}-E[\boldsymbol{x}])\|^2\right]\frac{\log e}{N_0} \\ &\quad + E\left[\operatorname{trace}\left(\operatorname{cov}^2(\boldsymbol{H}\boldsymbol{x}|\boldsymbol{x})\right)\right]\frac{\log e}{2N_0^2} \\ &\quad - \operatorname{trace}\left(\operatorname{cov}^2(\boldsymbol{H}\boldsymbol{x})\right)\frac{\log e}{2N_0^2} + o(N_0^{-2}). \quad (288) \end{aligned}$$

Note that, unlike the case where the channel is known at the receiver, mutual information is, in general, dependent on the mean of the input.

An interesting case to which we can apply (287) is the Ricean block-fading channel where the specular and scattered components remain constant during blocks of length m. Although the scattered coefficient is not known at the receiver, the fact that it remains piecewise constant enables its estimation at the receiver. The channel matrix is now the $m \times m$ multiple of the identity matrix

$$\boldsymbol{H} = (\bar{\mathsf{h}} + \mathsf{g})\boldsymbol{I} \tag{289}$$

where $\bar{\mathsf{h}}$ is deterministic and g is zero-mean complex Gaussian with variance γ^2. The bound in (287) is the convex function of SNR (in nats/dimension)

$$\frac{1}{m} I(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) \le |\bar{\mathsf{h}}|^2 \mathsf{SNR} + \frac{m}{2} \kappa(\|\boldsymbol{x}\|)\gamma^4 \mathsf{SNR}^2 \tag{290}$$

where $\kappa(\|\boldsymbol{x}\|)$ is the kurtosis of the norm of the input vector. Thus, unless the input kurtosis grows without bound as $\mathsf{SNR} \to 0$ (and, consequently, the peak-to-average ratio), there is no hope to achieve the mutual information required for $\frac{E_b^{\mathsf{r}}}{N_0}_{\min} = -1.59$ dB, namely, $(|\bar{h}|^2 + \gamma^2)\mathsf{SNR} + o(\mathsf{SNR})$. Furthermore, doubling the period over which the fading remains stable has the same effect as doubling the input kurtosis.

In general, when the channel is unknown at the receiver, the maximum rate achievable under a fixed constraint on the input kurtosis is not a concave function of SNR. In the absence of concavity of the maximum achievable rate function, (34) need not hold and $\frac{E_b}{N_0}_{\min}$ may be achieved at a nonzero SNR, in which case, the curve $\frac{E_b}{N_0}(\mathsf{C})$ is bowl-shaped and achieves its minimum at a nonzero C^*. (For example, in the special case $\bar{\mathsf{h}} = 0$, $\frac{E_b}{N_0}(\mathsf{C}) \to \infty$ as $\mathsf{C} \to 0$.) Thus, for every $\frac{E_b}{N_0} > \frac{E_b}{N_0}_{\min}$ there are two spectral efficiencies $\mathsf{C}_1 < \mathsf{C}_2$ such that

$$\frac{E_b}{N_0} = \frac{E_b}{N_0}(\mathsf{C}_1) = \frac{E_b}{N_0}(\mathsf{C}_2).$$

Any sensible design will choose to operate at C_2, as we can maintain the same power and data rate achieved at C_1 but with smaller bandwidth. Therefore, under input kurtosis constraints, the region of small SNR (specifically, $0 < \mathsf{SNR} < C^{-1}(\mathsf{C}^*)$) is to be avoided. Provided this design principle is followed, the required $\frac{E_b}{N_0}$ is, as usual, an increasing function of the spectral efficiency. Inefficient communication as the bandwidth grows without bound can be averted only by letting the data rate and the power grow at least as fast as the bandwidth.

F. Multiaccess Channels

We now turn attention to the multiaccess channel with an optimum receiver that has perfect channel side information.

Theorem 21: Consider the K-user randomly spread CDMA channel subject to frequency flat fading

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{291}$$

where the dimensionality of $\boldsymbol{y}$ is equal to the spreading factor N, $\boldsymbol{A}$ is a diagonal $K \times K$ matrix whose diagonal is composed of the independent and identically distributed fading coefficients $\mathsf{A}_1 \dots \mathsf{A}_K$ experienced by the K users, $\boldsymbol{S}$ is the $N \times K$ spreading matrix with independent zero-mean coefficients, the noise components are independent, and the transmitters have no knowledge of either $\boldsymbol{S}$ or $\boldsymbol{A}$. Then, the slope of the spectral efficiency is equal to

$$\mathcal{S}_0 = \frac{2K}{N\kappa(|\mathsf{A}|) + K - 1} \tag{292}$$

where $\kappa(|\mathsf{A}|)$ denotes the kurtosis of the magnitude of the fading coefficients.

Proof: As we saw in Section IV, the differences between the multiaccess setup and the single-user setup are the enforced independence of the input components and the different normalization of SNR, which leads to (127). However, by definition, $\mathcal{S}_0$ is invariant to any factor multiplying SNR. Therefore, we can use the same formula (140) as in the single-user case. Furthermore, the setting of Theorem 13 is identical to the multiaccess setting of interest here and, consequently, the result (208) can be used in this case with

$$\boldsymbol{H} = \boldsymbol{S}\boldsymbol{A}. \tag{293}$$

Since the result in (208) is invariant to scaling of $\boldsymbol{H}$, we can assume, for convenience, that the entries of $\boldsymbol{S}$ have unit variance. Then

$$\mathrm{trace}(E[\boldsymbol{H}^\dagger \boldsymbol{H}]) = KNE[|\mathsf{A}_1|^2] \tag{294}$$

and

$$\begin{aligned}\sum_{i=1}^K \sum_{j=1}^K & E[|(\boldsymbol{H}^\dagger \boldsymbol{H})_{ij}|^2] \\ &= KE[|(\boldsymbol{H}^\dagger \boldsymbol{H})_{11}|^2] + K(K-1)E[|(\boldsymbol{H}^\dagger \boldsymbol{H})_{12}|^2] \\ &= K \sum_{j=1}^N \sum_{\ell=1}^N E[|H_{j1}|^2 |H_{\ell 1}|^2] \\ &\quad + K(K-1) \sum_{j=1}^N E[|H_{j1}|^2 |H_{j2}|^2] \\ &= KN^2 E[|\mathsf{A}_1|^4] + KN(K-1)E^2[|\mathsf{A}_1|^2].\end{aligned} \tag{295}$$

Substituting (294) and (295) into (208) $(m = N; n = K)$ we get

$$\mathcal{S}_0 = \frac{2K^2 N^2 E^2[|\mathsf{A}_k|^2]}{KN^3 E[|\mathsf{A}_k|^4] + N^2 K(K-1)E^2[|\mathsf{A}_k|^2]} \tag{296}$$

which is equal to the desired result (292). □

As $K \to \infty$ and $N \to \infty$ with $\beta = K/N$, (292) converges to

$$\mathcal{S}_0 = \frac{2\beta}{\kappa(|\mathsf{A}|) + \beta}; \tag{297}$$

a result which was obtained in [15] using substantially different methods. Actually, the asymptotic result holds even if the randomly chosen signature waveforms remain fixed from symbol to symbol, whereas the fixed-dimensional Theorem 21 has the narrower scope of "long CDMA codes" where the signatures change from symbol to symbol.

Using a bank of single-user matched filters which neglect multiaccess interference the $\frac{E_b}{N_0}_{\min}$ is the same as with multiuser detection as we had seen in Theorem 8. However, the wideband slope achieved by the matched filter bank turns out to be

$$\mathcal{S}_0 = \frac{2K}{N\kappa(|\mathsf{A}|) + 2(K-1)}. \tag{298}$$

Comparing (298) to (292) we see that even in the low-power scenario, where background noise rather than multiaccess interference is the major source of interference, the use of optimum multiuser detection can save up to 50% of the bandwidth (as the ratio K/N grows). Thus, contrary to what is sometimes claimed (e.g., [41]) multiuser detection can be quite effective for low signal-to-noise ratio communication with error control coding. The optimality of time division multiple access (TDMA) for multiaccess and broadcast channels in the low-power regime is another misconception unveiled using the wideband slope [42].

Another problem of interest is to find the best possible slope over all assignments of signature waveforms. To simplify the setup, we assume that there is no fading and that all users are received with the same power, which we can take equal to unity. Then, the matrix $\boldsymbol{H}^\dagger \boldsymbol{H}$ is equal to the normalized cross-correlation matrix $\boldsymbol{R}$ [26], and (208) becomes

$$\mathcal{S}_0 = \frac{2K^2}{N \sum_{j=1}^{K} \sum_{k=1}^{K} |\rho_{jk}|^2} \tag{299}$$

$$\leq 2 \tag{300}$$

where the inequality is precisely the Welch bound [26]. Thus, we see directly that a set of signature waveforms that satisfies the Welch bound with equality maximizes $\mathcal{S}_0$ (equal to the single-user AWGN slope). In fact, meeting the Welch bound with equality is a necessary and sufficient condition to maximize spectral efficiency for all $\frac{E_b}{N_0}$ as shown in [43].

G. Additive Non-Gaussian Noise

We conclude by returning to the non-Gaussian case studied in Section IV. We saw in Theorem 3 that for the Laplacian noise channel, $\frac{E_b}{N_0}_{\min}$ improved by 3 dB relative to the (worst case) Gaussian noise. However, the wideband slope is quite a bit worse than for the Gaussian channel as the following result shows.

Theorem 22: For the Laplacian noise channel (71) and (72), even if the receiver knows the channel

$$\mathcal{S}_0 = 0. \tag{301}$$

Proof: We need to show that the second derivative of the capacity at $\mathsf{SNR} = 0$ is equal to $-\infty$. For simplicity and without loss of generality, we take $m = 1$ and $N_0 = 1$. Define the following function on the positive-real line:

$$f(z) = -1 + 2\sqrt{z} - 2z + \exp\left(-2\sqrt{z}\right). \tag{302}$$

For any input distribution whose second moment is

$$E[|\mathsf{x}|^2] = \mathsf{SNR} \tag{303}$$

we can write

$$I(X;Y) - \dot{C}(0)\mathsf{SNR}$$
$$\leq D(P_{Y|X=\mathsf{x}} || P_{Y|X=0} | P_{\mathsf{x}}) - \dot{C}(0)\mathsf{SNR} \tag{304}$$
$$= E[f(\Re^2 \mathsf{x})] + E[f(\Im^2 \mathsf{x})] \tag{305}$$
$$\leq 2f\left(\frac{\mathsf{SNR}}{2}\right) \tag{306}$$
$$= -\frac{4}{3\sqrt{2}}\mathsf{SNR}^{3/2} + \frac{1}{3}\mathsf{SNR}^2 + o(\mathsf{SNR}^2) \tag{307}$$

where (304) is a result of dropping the second divergence on the right-hand side of (159); (305) follows by using (75) and the explicit expression for the divergence found in (74); (306) holds because $f(z)$ is a concave function, so it actually holds with equality for deterministic input ($\mathsf{x} = \sqrt{\mathsf{SNR}/2}(1+j)$). To complete the proof that $\ddot{C}(0) = -\infty$, we just need to divide (307) by SNR^2 and let $\mathsf{SNR} \to 0$. □

VI. Conclusion

Spectral efficiency treats time and frequency on an equal footing. Thus, the "wideband regime" studied in this paper encompasses more than the name implies. Low spectral efficiency values are obtained not only when a given data rate (b/s) is transmitted through a very large bandwidth, but when a given bandwidth is used to transmit a very small data rate. The setting encompasses even the case where the bandwidth is not large, the data rate is not low, but the number of receive antennas is large. Thus, the "wideband regime" is to be understood as encompassing any scenario where the number of information bits transmitted per receive dimension is small.

The infinite-bandwidth (or, more generally and precisely, zero bits per dimension) analysis leads to the conclusion that as long as the additive background noise is Gaussian, the received $\frac{E_b}{N_0}$ must be equal to -1.59 dB regardless of receiver/transmitter side information. Channel knowledge at the transmitter may be useful in the infinite bandwidth regime in order to signal along the most favorable dimensions that lead to $\frac{E_b^r}{N_0} = -1.59$ dB with the least power expenditure. In addition to low-duty-cycle on–off keying signaling, which has been the traditional focus of information-theoretic analyses of the infinite-bandwidth channels, we have seen that as long as the input distribution wastes no (or negligible) power in its mean, the $\frac{E_b}{N_0}_{\min}$ is achieved if the receiver knows the channel. We have identified flash signaling, a class of unbounded peak-to-average inputs that is necessary and sufficient to achieve $\frac{E_b}{N_0}_{\min}$ if the receiver does not have perfect channel knowledge.

Transmission of nonzero bits per dimension changes the picture quite radically. Both optimal signaling and the efficiency with which information can be transmitted depend crucially on whether the receiver knows the channel. If the receiver has perfect channel knowledge, we have shown that QPSK is optimal in the wideband regime and that, when compared to QPSK, on–off keying requires more than six times as much bandwidth. The bandwidth required to send a given data rate is proportional to the peakiness of the channel fading quantified by the kurtosis (fourth moment relative to second moment squared) of its amplitude.

In the absence of perfect channel information at the receiver, approaching $\frac{E_b}{N_0}_{\min}$ is prohibitively expensive in terms of both spectral efficiency and peak-to-average ratio.

We have shown that the asymptotic optimality criterion used since [2] (namely, ratio of mutual information to capacity approaching one as signal-to-noise ratio vanishes) is too weak to gauge bandwidth requirements. The weakness of this "traditional" optimality criterion can be gleaned 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. To replace this criterion, we have shown that a signaling format is wideband optimal if it achieves both the first and second derivatives of the capacity function at zero SNR. Note that this new criterion is in no way dependent on the definition of wideband slope with respect to decibel rather than with respect to linear scale. While the logarithmic scale is more analytically convenient and practically insightful, the alternative definition with linear $\frac{E_b}{N_0}$ leads to the same optimality criterion.

The impact of channel impairments and design choices (such as input signaling and coherent versus noncoherent communication) is not equally apparent depending on whether we analyze data rate for given bandwidth and power, or bandwidth for given power and data rate. Since in the low-power regime the bandwidth sensitivity is usually far greater, it is unwise to follow the traditional paradigm of the voiceband telephone channel which maximizes data rate for given power and bandwidth. For example, consider a Rayleigh channel operating at $\frac{E_b^{\mathsf{r}}}{N_0} = 1.25$ dB. In the noncoherent regime, the spectral efficiency is equal to 0.0011 b/s/Hz (Fig. 4). While coherence buys a 92% improvement in rate for fixed bandwidth and power, it reduces bandwidth by a factor of 1000 for fixed power and rate. If we let the initial $\frac{E_b^{\mathsf{r}}}{N_0} \downarrow \frac{E_b^{\mathsf{r}}}{N_0}_{\min}$, then the improvement in rate brought about by coherence vanishes, whereas the bandwidth reduction factor goes to infinity. Thus, in wireless channels where bandwidth is an expensive commodity, it is inadvisable to dictate a choice of bandwidth without careful analysis of information-theoretic limits.

The wideband slope has thus emerged as a new analysis tool that leads to valuable insights and serves to reveal several long-standing misconceptions on the practical significance of low-SNR information-theoretic results.

REFERENCES

[1] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10–21, Jan. 1949.

[2] ——, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July–Oct. 1948.

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

[4] G. L. Turin, "The asymptotic behavior of ideal m-ary systems," *Proc. IRE*, vol. 47, pp. 93–94, Jan. 1959.

[5] I. Jacobs, "The asymptotic behavior of incoherent m-ary communication systems," *Proc. IEEE*, vol. 51, pp. 251–252, Jan. 1963.

[6] J. R. Pierce, "Ultimate performance of m-ary transmissions on fading channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 2–5, Jan. 1966.

[7] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley-Interscience, 1969.

[8] A. J. Viterbi, "Performance of an m-ary orthogonal communication system using stationary stochastic signals," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 414–421, July 1967.

[9] R. G. Gallager, "Power limited channels: Coding, multiaccess, and spread spectrum," in *Proc. 1988 Conf. Information Sciences and Systems*, Mar. 1988. Also, Tech. Rep. LIDS-P-1714.

[10] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[11] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[12] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[13] G. J. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Lab. Tech. J.*, pp. 41–59, Autumn 1996.

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

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

[16] L. Zheng and D.N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

[17] F. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.

[18] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side-information' be?," in *Proc. IEEE Information theory and Communications Workshop*, 1999, pp. 36–38.

[19] M. Pinsker, V. V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inform. Theory*, pp. 1877–1888, Nov. 1995.

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

[21] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.

[22] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[23] E. Visotsky and U. Madhow, "Space–time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.

[24] S. A. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. 2001 IEEE Int. Symp. Information Theory*, June 2001, p. 321.

[25] J. S. Richters, "Communication over fading dispersive channels," Tech. Rep., MIT Res. Lab. Electronics, Nov. 1967.

[26] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.

[27] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.

[28] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.

[29] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication*. Reading, MA: Addison-Wesley, 1977.

[30] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis*. New York: Wiley-Interscience, 2000.

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

[32] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Select. Areas Communications*, vol. 18, pp. 2172–2178, Nov. 2000.

[33] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. 24th Allerton Conf.*, Oct. 1986, pp. 1025–1034.

[34] A. Lozano, A. Tulino, and S. Verdú, "Multiantenna capacity in the low power regime," unpublished manuscript, 2002.

[35] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

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

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

[38] M. Medard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inform. Theory*, vol. 48, pp. 840–853, Apr. 2002.

[39] V. J. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Trans. Inform. Theory*, pp. 809–827, Apr. 2002.

[40] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," unpublished manuscript, 2002.

[41] S. Vembu and A. Viterbi, "Two different philosophies in CDMA—A comparison," in *Proc. 46th IEEE Vehicular Technology Conf.*, Apr.–May 1996, pp. 869–873.

[42] S. Verdú, G. Caire, and D. Tuninetti, "Is TDMA wideband optimal?," in *Proc. 2002 IEEE Int. Symp. Information Theory*, to be published.

[43] M. Rupf and J. L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1261–1266, July 1994.