Error Exponents for Channel Coding and Signal Constellation Design

Jianyi Huang
Marvell Technology, 5400 Bayfront plz
Santa Clara, CA
Email: jianyih@marvell.com

Sean Meyn
Dept. of Electrical and Computer Engg.
and the Coordinated Sciences Lab.
Univ. of Illinois at Urbana-Champaign
Email: meyn@uiuc.edu

Muriel Médard
Lab. for Info. & Decision Systems
Massachusetts Institute of Technology
Email: medard@mit.edu

***Abstract*— This paper summarizes results from the journal article and thesis [10], [9], and the survey article [12].**

Capacity and the random coding exponent is addressed for a general non-coherent stochastic channel model. Under very general assumptions, the distribution optimizing the error exponent has a finite number of mass points, or in the case of a complex channel, the amplitude has finite support. In each numerical example considered, the resulting code significantly out-performs traditional signal constellation schemes such as QAM and PSK.

***Acknowledgements*: S. M. is supported in part by NSF awards ECS 02 17836, and ITR 00-85929. M. M. is supported in part by NSF grant Career 6891730.**

I. Introduction

The problem of constellation design has recently received renewed attention in information theory and communication theory. While many techniques in information theory such as coding have readily found their way into communication applications, the signal constellations that information theory envisages and those generally considered by practitioners differ significantly. In particular, while the optimum constellation for an additive Gaussian noise (AWGN) channel is a continuous constellation that allows for a Gaussian distribution on the input, commonly used constellations over AWGN channels, such as quadrature amplitude modulation (QAM), are not only discrete, but also generally regularly spaced. This gap between theory and practice can be explained in part by the difficulty of deploying, in practical systems, continuous constellations.

However, there is also a body of work that strongly suggests the continuous paradigm favored by theoreticians is inappropriate for realistic channel models in the majority of today's applications, such as wireless communication systems. Under any of the following conditions the optimal capacity achieving distribution has a finite number of mass points, or in the case of a complex channel, the amplitude has finite support:

(i) The AWGN channel under a peak power constraint [17], [16], [15], [3].

(ii) Channels with fading, such as Rayleigh [1] and Rician fading [8], [7]. Substantial generalizations are given in [11], [9].

(iii) Lack of channel coherence [13]. For the noncoherent Rayleigh fading channel, a Gaussian input is shown to generate bounded mutual information as SNR goes to infinity [4], [14].

(iv) Under general conditions a binary distribution is optimal, or approximately optimal for sufficiently low SNR ([6], and [18, Theorem 3].)

The finiteness of the support of the optimal input distribution bodes well for implementation of optimal constellations in communication systems, but the matter of the specific choice of points remains. The problem of selecting such a constellation is one of nonlinear optimization when channel capacity is considered [11]. While capacity provides a fundamental characterization of channel performance, the issue of how to achieve rates close to capacity with low probability of error can also be characterized in a rigorous fashion through the use of error exponents. In this paper, we pose the constellation design problem in the context of error exponent optimization.

We consider a stationary, memoryless channel with input alphabet X, output alphabet Y, and transition density defined by

$$\mathsf{P}(Y \in dy \mid X = x) = p(y|x)\, dy\,, \qquad x \in \mathsf{X},\ y \in \mathsf{Y}. \quad (1)$$

It is assumed that Y is equal to either $\mathbb{R}$ or $\mathbb{C}$, and we assume that X is a closed subset of $\mathbb{R}$.

Many complex channel models in which X is equal to $\mathbb{C}$ are considered by viewing μ as the amplitude of X. In this case the channel input is denoted U, the output is V, and the transition density on $\mathbb{C} \times \mathbb{C}$ it is assumed *symmetric*:

$$p_\bullet(v|u) = p_\bullet(e^{j\alpha}v|e^{j\alpha}u), \qquad u, v \in \mathbb{C},\ \alpha \in \mathbb{R}. \quad (2)$$

This is a natural assumption in many applications, such as Rayleigh and Rician channels, since phase information is lost at high bandwidths.

Throughout the paper we restrict to *noncoherent* channels in which neither the transmitter nor the receiver knows the channel state.

The *channel reliability function* is defined as

$$E(R) = \lim_{N\to\infty} \Big[-\frac{1}{N} \log p_e(N, R)\Big], \qquad R > 0,$$

where $p_e(N, R)$ is the minimal probability of error, over all *block codes* of length N and rate R. The main goal in this paper is to maximize the error exponent, subject to two linear constraints:

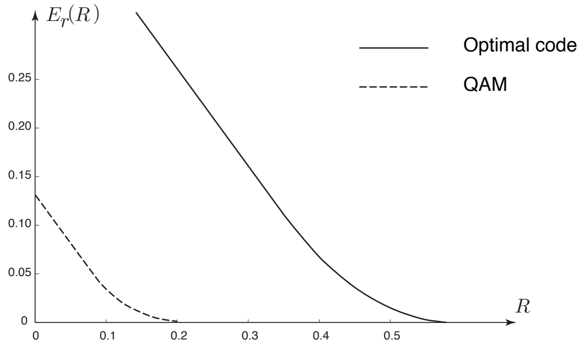

Fig. 1. Error exponent $E_r(R)$ for the two codes shown in Figure 5. The 3-point constellation performs better than 16-point QAM for all rates $R \leq C$.

(i) The *average power constraint* that

$$\langle \mu, \phi \rangle \leq \sigma_P^2$$

where $\langle \mu, \phi \rangle := \int \phi(x)\ \mu(dx)$, and $\phi(x) := x^2$ for $x \in \mathbb{R}$.

(ii) The *peak power constraint* that μ is supported on $\mathsf{X} \cap [-M, M]$ for a given $M \leq \infty$.

Hence the input distribution is constrained to the convex set $\mathcal{M}(\sigma_P^2, M, \mathsf{X})$:=

$$\Big\{\mu \in \mathcal{M} : \langle \mu, \phi \rangle \leq \sigma_P^2,\ \mu\{[-M, M]\} = 1\Big\}, \tag{3}$$

where $\mathcal{M}$ denotes the set of probability distributions on X.

The *random coding exponent* $E_r(R)$ may be expressed,

$$\begin{aligned} E_r(R) &:= \sup_{0 \leq \rho \leq 1} \big[-\rho R + \sup_{\mu}(-\log(G^\rho(\mu)))\big], \\ G^\rho(\mu) &:= \int \Big[\int \mu(dx) p(y|x)^{1/(1+\rho)}\Big]^{1+\rho} dy, \end{aligned} \tag{4}$$

from which we obtain the *random-coding bound*,

$$p_e(N, R) \leq \exp(-N E_r(R)), \qquad N \geq 1, R \geq 0,$$

which holds under the assumptions imposed in this paper.

Moreover, the equality $E(R) = E_r(R)$ holds for rates greater than the *critical rate* R_{crit} [2], [5]. Consequently, if one can design a distribution with a large error exponent, then the associated random code can be constructed with a correspondingly small block-length. This has tremendous benefit in implementation. Figure 1 shows how a better designed code can greatly out-perform QAM for rates R below capacity.

Optimization of the random coding exponent is addressed as follows. Rather than parameterize the optimization problem by the given rate $R > 0$, we consider for each Lagrange multiplier $\rho \in (0, 1]$ the convex program,

$$\begin{aligned} &\textbf{inf} && G^\rho(\mu) \\ &\textbf{subject to} && \mu \in \mathcal{M}(\sigma_P^2, M, \mathsf{X}). \end{aligned} \tag{5}$$

The value is denoted $G^{\rho *}$, and we have from (4)

$$E_r(R) = \sup_{0 \leq \rho \leq 1} \big[-\rho R - \log(G^{\rho *})\big].$$

The alignment conditions obtained in Section II-B are based on the *error exponent sensitivity function*, defined by

$$g_\mu^\rho(x) := \int \Big[\int \mu(dz) p(y|z)^{1/(1+\rho)}\Big]^\rho p(y|x)^{1/(1+\rho)}\, dy. \tag{6}$$

The objective function in (5) can be written $G^\rho(\mu) = \langle \mu, g_\mu^\rho \rangle$. Similarly, mutual information can be expressed $I(\mu) = \langle \mu, g_\mu \rangle$ where the *channel sensitivity function* is defined by

$$g_\mu(x) := \int \log\big[p(y|x)/p_\mu(y)\big] p(y|x)\, dy\,, \tag{7}$$

where p_μ denotes the marginal of the output,

$$p_\mu(dy) = \int \mu(dx) p(y|x), \qquad y \in \mathsf{Y}. \tag{8}$$

We list here the remaining assumptions imposed on the real channel in this paper.

(A1) The input alphabet X is a closed subset of $\mathbb{R}$, $\mathsf{Y} = \mathbb{C}$ or $\mathbb{R}$, and $\min(\sigma_P^2, M) < \infty$.

(A2) Y is large when X is large: For each $n \geq 1$,

$$\lim_{|x| \to \infty} \mathsf{P}(|Y| < n \mid X = x) = 0$$

(A3) The function $\log(p(\,\cdot\,|\,\cdot\,))$ is continuous on $\mathsf{X} \times \mathsf{Y}$ and, for any $y \in \mathsf{Y}$, $\log(p(y|\,\cdot\,))$ is analytic within the interior of X. Moreover, g_μ and g_μ^ρ are analytic functions on $\mathbb{R}$ for any $\rho \in (0, 1]$ and $\mu \in \mathcal{M}(\sigma_P^2, M, \mathsf{X})$.

We occasionally also assume,

(A4) For each ρ, if $\mu^0 \neq \mu^1$ then, for *some* y, $\int [\mu^0(dz) - \mu^1(dz)] p(y|z)^{1/(1+\rho)} \neq 0.$

Conditions (A1)-(A3) are also the standing assumptions in [11]. It is shown there that these assumptions are satisfied in all of the standard models, including the AWGN, phase-noise, Rayleigh, and Rician channels.

It can be shown that the functional G^ρ is strictly convex under (A4).

The capacity-achieving input distribution is discrete under the conditions imposed here when M is finite [11]. We find that the distribution optimizing the error exponent E_r for a given positive rate $R < C$ is *always* discrete in the real channel, with or without a peak power constraint. In the symmetric complex channel, the distribution is symmetric with finite support.

The following result provides a summary of results for a symmetric complex channel. For any $\mu \in \mathcal{M}$, we define $\mu_\bullet$ as the symmetric distribution on $\mathbb{C}$ whose magnitude has distribution μ. That is, we have the polar-coordinates representation,

$$\mu_\bullet(dx \times d\alpha) = \frac{1}{2\pi x} \mu(dx) d\alpha\,, \qquad x > 0,\ 0 \leq \alpha \leq 2\pi,$$

and we set $\mu(\{0\}) = \mu_\bullet(\{0\})$.

Theorem 1.1: Consider a complex channel model satisfying the symmetry condition (2) and Assumptions (A1)–(A3). Then,

(i) If $M < \infty$ then there exists an optimizer $\mu^*_\bullet$ of the convex program defininining capacity,

$$\begin{array}{ll} \textbf{sup} & I(\mu) \\ \textbf{subject to} & \mu \in \mathcal{M}(\sigma_P^2, M, \mathsf{X}). \end{array} \tag{9}$$

The distribution $\mu^*_\bullet$ is symmetric with finite magnitude.

(ii) Consider the convex program (5) under the relaxed condition that $\min(\sigma_P^2, M) < \infty$. For each ρ there exists an optimizer $\mu^\rho_\bullet$ that is symmetric, and its magnitude μ^ρ has finite support. Moreoever, for each $R \in (0, C)$ there exists ρ^* achieving the maximum in (4) so that,

$$E_r(R) = -\rho^* R - \log(G^{\rho*}) = -\rho^* R - \log(G^{\rho^*}(\mu^{\rho^*}))\,.$$

□

II. Convex optimization and channel coding

A. Error exponents and robust hypothesis testing

Here we show that the random coding exponent bound can be derived by casting the signal detection process as a *robust* hypothesis testing problem.

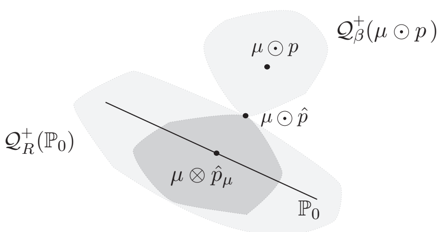

Fig. 2. The random coding exponent $E_r(R)$ is equal to the solution of a robust N-P hypothesis testing problem with 'false alarm rate' R, and 'missed detection rate' β as shown in the figure.

For a given $\mu \in \mathcal{M}$, $\eta > 0$, we denote the *divergence neighborhood* of μ by,

$$\mathcal{Q}^+_\eta(\mu) = \{\nu : D(\nu \parallel \mu) \le \eta\}.$$

We denote by $\mathbb{P}_0$ the space of product measures on $\mathsf{X} \times \mathsf{Y}$,

$$\mathbb{P}_0 = \{\mu \otimes \nu : \nu \text{ is a probability measure on } \mathsf{Y}\},$$

and define the corresponding divergence set for a given $R > 0$,

$$\begin{aligned} \mathcal{Q}^+_R(\mathbb{P}_0) &:= \bigcup_\nu \mathcal{Q}^+_R(\mu \otimes \nu). \\ &= \big\{\gamma : \min_\nu D(\gamma \parallel \mu \otimes \nu) \le R\big\}. \end{aligned}$$

The robust hypothesis testing problem is binary, with H_1 as defined in the channel capacity problem, but with H_0 defined using $\mathbb{P}_0$:

H_0: $\{(X^i_j, Y_j) : j = 1, \dots, N\}$ has marginal $\pi^0 \in \mathbb{P}_0$.

H_1: $\{(X^i_j, Y_j) : j = 1, \dots, N\}$ has marginal $\pi^1 := \mu \odot p$.

Proposition 2.1:

$$E_r(R) = \sup_\mu \Big(\inf_\beta \big\{\beta : \mathcal{Q}^+_\beta(\mu \odot p) \cap \mathcal{Q}^+_R(\mathbb{P}_0) \neq \emptyset\big\}\Big). \tag{10}$$

Suppose that there exists a triple (μ^*, ν^*, γ^*) that solve (10) in the sense that

$$D(\gamma^* \parallel \mu^* \odot p) = E_r(R), \quad D(\gamma^* \parallel \mu^* \otimes \nu^*) = R.$$

Then, there exists a channel transition density $\hat{p}$ such that

$$\gamma^* = \mu^* \odot \hat{p}, \quad \nu^* = \hat{p}_{\mu^*},$$

and the rate can be expressed as mutual information,

$$R = I(\mu^*; \hat{p}) := D(\mu^* \odot \hat{p} \parallel \mu^* \otimes \hat{p}_\mu).$$

□

For fixed μ, the error exponent obtained in the proof [10] is given by,

$$\beta^*(\mu) := \inf\big\{\beta : \mathcal{Q}^+_\beta(\mu \odot p) \cap \mathcal{Q}^+_R(\mathbb{P}_0) \neq \emptyset\big\} \tag{11}$$

The solution to (11) is illustrated in Figure 2. The channel transition density $\hat{p}$ shown in the figure solves

$$\begin{aligned} \beta^*(\mu) &= \inf\big\{\beta : \mathcal{Q}^+_\beta(\mu \odot p) \cap \mathcal{Q}^+_R(\mu \otimes \hat{p}_\mu) \neq \emptyset\big\} \\ &= D(\mu \odot \hat{p} \parallel \mu \odot p). \end{aligned}$$

The error exponent is equal to the maximal relative entropy $\beta^*(\mu)$ over all μ, and the rate can be expressed as mutual information $R = I(\mu^*; \hat{p}) := D(\mu^* \odot \hat{p} \parallel \mu^* \otimes \hat{p}_\mu)$ where μ^* is the optimizing distribution.

B. Convexity and alignment

Boundedness of the sensitivity function is central to the analysis that follows.

Proposition 2.2: The following hold under (A1)-(A3):

(i) $0 < g^\rho_\mu(x) \le 1$ for each x.

(ii) $g^\rho_\mu \to 0$ as $x \to \infty$.

(iii) Suppose that there is a peak power constraint, so that $M < \infty$. For each finite $N > 0$, the mapping $\mu \to g_\mu$ is continuous from $\mathcal{M}(\sigma_P^2, M, \mathsf{X})$ to $L_\infty[-N, N]$. That is if $\mu_n \to \mu$ weakly, with $\mu_n \in \mathcal{M}(\sigma_P^2, M, \mathsf{X})$ for all n, then

$$\lim_{n\to\infty} \sup_{|x|\le N} |g^\rho_{\mu_n}(x) - g^\rho_\mu(x)| = 0. \tag{12}$$

□

The following set of results establishes convexity and differentiability of G^ρ with respect to μ. For $\mu, \mu^\circ \in \mathcal{M}$ and $\theta \in [0, 1]$ we denote $\mu_\theta := (1 - \theta)\mu^\circ + \theta\mu$.

Proposition 2.3: The following hold under (A1)–(A3): For any given $\mu, \mu^\circ \in \mathcal{M}(\sigma_P^2, M, \mathsf{X})$ and $\rho > 0$,

(i) $G^\rho(\mu) = \langle \mu, g^\rho_\mu \rangle$.

(ii) The functional G^ρ is convex, and can be expressed as the maximum of linear functionals,

$$G^\rho(\mu^\circ) = \max_{\mu\in\mathcal{M}} \{(1 + \rho)\langle \mu^\circ, g^\rho_\mu \rangle - \rho G^\rho(\mu)\}$$

(iii) Fix $\rho \ge 0$, $\mu^\circ \in \mathcal{M}$. The first order sensitivity is given by

$$\frac{d}{d\theta} G^\rho(\mu_\theta)\Big|_{\theta=0} = (1 + \rho)\langle \mu - \mu^\circ, g^\rho_{\mu^\circ} \rangle.$$

(iv) If (A4) holds then G^ρ is strictly convex for each $\rho > 0$.

□

For fixed ρ, the optimization problem (5) is a convex program since G^ρ is convex. Continuity then leads to existence of an optimizer. The following result summarizes the structure of the optimal input distribution. It is similar to Theorem 2.8 of [11], which required the peak power constraint $M < \infty$. We stress that this condition is not required here.

Theorem 2.4: Suppose that (A1)–(A3) hold. Then,

(i) For each $\rho \geq 0$, there exists $\mu^\rho \in \mathcal{M}(\sigma_P^2, M, \mathsf{X})$ that achieves $G^{\rho *}$.

(ii) A given distribution $\mu^\circ \in \mathcal{M}(\sigma_P^2, M, \mathsf{X})$ is optimal if and only if there exists a real number λ_1^* and a positive real number λ_2^* such that for a.e. $x \quad [\mu^\circ]$

$$g_{\mu^\circ}^\rho(x) \geq \lambda_1^* - \lambda_2^* x^2, \quad \text{and} \quad g_{\mu^\circ}^\rho(x) = \lambda_1^* - \lambda_2^* x^2.$$

If these conditions hold, then

$$G^{\rho *} := \min_\mu G^\rho(\mu) = G^\rho(\mu^\circ) = \frac{\lambda_1^* - \lambda_2^* \sigma_P^2}{1+\rho}.$$

(iii) In the absence of an average power constraint $\lambda_2^* = 0$.

(iv) If (A4) holds, then the optimizer μ^ρ is unique.

□

III. Signal constellation design

For a symmetric, complex channel we have seen in Theorem 1.1 that the optimal input distribution is circularly symmetric on $\mathbb{C}$, and discrete in magnitude. We consider in this section discrete approximations of this optimal distribution on $\mathbb{C}$.

A *signal constellation* is a finite set of points in $\mathbb{C}$ that is used to define possible codewords. Two well known examples are *quadrature-amplitude modulation* (QAM), and *phase-shift keyed* (PSK) coding. In these coding schemes the codewords are chosen using a random code constructed with a uniform distribution across the given signal constellation. *Multi-dimensional constellations* and *non-equiprobable signaling* have been proposed to reduce mean-square energy. These methods are largely motivated by properties of the AWGN channel.

We propose the following approach to signal constellation design and coding in which the signal alphabet and associated probabilities are chosen to provide a random code that approximates the random code obtained through the nonlinear program (9). We conclude this paper with examples to illustrate the performance of this approach, as compared to standard approaches based on QAM or PSK.

a) ***Complex AWGN channel:*** This is the complex channel model given by $Y = X + N$ with N complex Gaussian. Examples considered in [2] suggest that QAM typically outperforms PSK when the constellation sizes are fixed, and the signal to noise ratio is large. For small SNR, it is known that QAM and PSK are almost optimal (see [2, Figure 7.11], [19], and related results in [16]).

We illustrate application of the approach proposed here using the symmetric AWGN complex channel with $\sigma_P^2 = 9$ and $\sigma_N^2 = 1$, i.e. SNR $= 9$ $(9.54dB)$. Figure 3 shows results using two signal constellation methods: 4-point QAM and a 5-point distribution which contains the 4-point QAM plus a point at origin. The 5 point distribution is an approximation to the optimal input distribution, which is binary in magnitude, and uniformly distributed in phase. The 5-point constellation performs better than 4-point QAM by about 13%, with lower power consumption. □

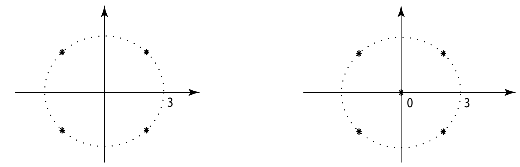

Fig. 3. Signal constellations for thecomplex AWGN channel with $\sigma_N^2 = 1$. At left is the 4-point QAM signal constellation, and at right an additional point is added at the origin. The distribution μ is uniform in each case. The constellation at right achieves 1.52 nats/symbol mutual information, while QAM achieves 1.38 nats/symbol.

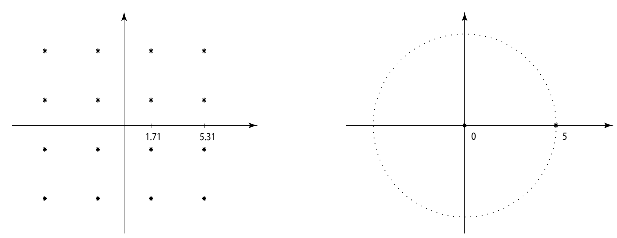

Fig. 4. Signal constellations for the Rayleigh channel $Y = AX + N$ with $\sigma_A^2 = 1$, $\sigma_N^2 = 1$, and average power constraint $\sigma_P^2 = 11.7$. The 16-point QAM constellation shown at left (with uniform distribution) achieves 0.1951 nats/symbol. The optimal constellation shown at right has two points of support, with one point at origin (with probability 0.5346.) The mutual information achieved is 0.4879 nats/symbol, which is 2.5 times more than that achieved by QAM.

b) ***Rayleigh channel with low SNR:*** This is the complex channel model $Y = AX + N$, with A, N, each Gaussian, and mutually independent. We assume that each is circularly symmetric, with $\sigma_A^2 = 1$, $\sigma_N^2 = 1$. Note that the channel output depends only on the magnitude of channel input in this model.

We compare codes obtained from the two constellations illustrated in Figure 4. The first constellation is a 16-point QAM. Since the code used in QAM is a random code with uniform distribution, the average power is given by $\sigma_P^2 = 11.7$. The second constellation has only two elements: one point at origin and another point at position $5 \in \mathbb{C}$. The weights are chosen so that the average power is again $\sigma_P^2 = 11.7$, which results in $\mu\{0\} = 1 - \mu\{5\} = 0.5346$. This is the optimal input distribution when the peak-power constraint $M = 5$ is imposed.

Computations show that the simpler coding scheme achieves mutual information 0.4879 nats/symbol, which is about 2.5 times more than the mutual information achieved by the 16-point QAM code. □

*c) **Rayleigh channel with high SNR:*** In this final example the same parameters used in the previous experiment are maintained, except now the average power is increased to $\sigma_P^2 = 26.4$. The optimal input distribution is given as follows when the channel is subject to the peak power constraint $|X| \leq 8$: The phase may be taken uniformly distributed without any loss of generality, and the magnitude has three points of support at $\{0.0, 2.7, 8.0\}$ with respective probabilities $\{0.465, 0.138, 0.397\}$. Consequently, we propose a constellation whose magnitude is restricted to these three radii. This is compared to 16-point QAM. The two constellation designs are illustrated in Figure 5. □

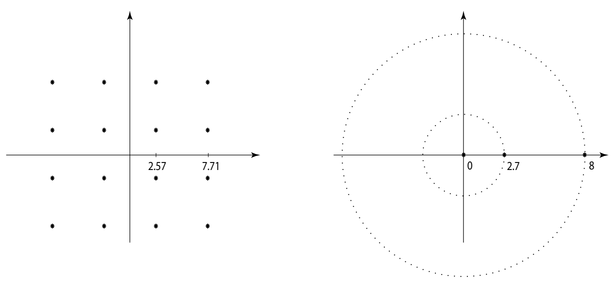

Fig. 5. The Rayleigh channel with average power increased to $\sigma_P^2 = 26.4$. The optimal constellation shown at right has three-points of support $\{0, 2.57, 8\}$, and respective probabilities $\{0.465, 0.138, 0.397\}$. The resulting mutual information is 0.5956nats/symbol, which is about 3 times larger than the mutual information achieved by the 16-point QAM.

IV. Conclusions

Many problems in information theory may be cast as a convex program over a set of probability distributions. Examples included in recent research include hypothesis testing, channel capacity, computation of the random coding exponent, and computation of the distortion function in source coding. Algorithms to compute optimal distributions based on the cutting plane proposed in [11], [9], [10] are remarkably efficient.

There are many unanswered questions:

(i) In this paper we restrict attention to the random coding exponent $E_r(R)$. However, for rates below the critical rate $R < R_{crit}$ there may be significantly better bounds such as the sphere-packing bound, the expurgated bound, or the straight-line bound (see Section 5.8 of [5].) These bounds have a representation similar to the random coding exponent [10]. Hence it is likely that analogous theory can be developed in these cases.

(ii) The theory described here sets the stage for further research on channel sensitivity. For example, how sensitive is the error exponent to SNR, coherence, channel memory, or other parameters?

(iii) It is possible to extend most of these results to multiple access channels. However, we have not yet extended the cutting plane algorithm to MIMO channels, and we do not know if the resulting algorithms will be computationally feasible.

(iv) In some applications optimal distributions may not be feasible due to the resulting 'peakiness' of the code book. There are then tradeoffs, and resource investment issues to be addressed. For example, one can consider if increasing the number of transmitter antennas will reduce the peakiness of the optimal random codebook.

(v) It is known that QAM and PSK perform well for the AWGN channel at specific SNR. In the case of fading channels, it is of interest to see if efficient constellations can be adapted to a changing environment.

References

[1] I. C. Abou-Faycal, M. D. Trott, and S. Shamai. The capcity of discrete-time memoryless Rayleigh-fading channels. *TIT*, 47(4):1290–1301, May 2001.

[2] R.E Blahut. *Principles and Practice of Information Theory*. McGraw-Hill, New York, 1995.

[3] T.H. Chan, S. Hranilovic, and F.R. Kschischang. Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs. *IEEE Trans. Inform. Theory*, 51 51, Issue: 6 51(6):2073– 2088, 2004.

[4] Rong-Rong Chen, B. Hajek, R. Koetter, and U. Madhow. On fixed input distributions for noncoherent communication over high SNR Rayleigh fading channels. *IEEE Trans. Inform. Theory*, 50(12):3390–3396, 2004.

[5] R.G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[6] R.G. Gallager. Power limited channels: Coding, multiaccess, and spread spectrum. In R.E. Blahut and R. Koetter, editors, *Codes, Graphs, and Systems*, pages 229–257. Kluwer Academic Publishers, Boston, 2002.

[7] M.C. Gursoy, H.V. Poor, and S. Verdu. The noncoherent Rician fading channel - part I: Structure of capacity achieving input. *IEEE Trans. Wireless Commun.*, 4(5):2193–2206, 2005.

[8] M.C. Gursoy, H.V. Poor, and S. Verdu. The noncoherent Rician fading channel - part II: Spectral efficiency in the low power regime. *IEEE Trans. Wireless Commun.*, 4(5):2207–2221, 2005.

[9] J. Huang. *Characterization and computation of optimal distribution for channel coding*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, 2004.

[10] J. Huang, S. Meyn, and M. Medard. Error exponents for channel coding and signal constellation design. *Journal of Selected Areas in Comm. Special Issue on Nonlinear Optimization of Communication Systems (submitted)*, 2005.

[11] J. Huang and S. P. Meyn. Characterization and computation of optimal distribution for channel coding. *IEEE Trans. Inform. Theory*, 51(7):1–16, 2005.

[12] J. Huang, C. Pandit, S.P. Meyn, M. Medard, and V. Veeravalli. Entropy, inference, and channel coding. In Prathima Agrawal, Matthew Andrews, Philip J. Fleming, George Yin, and Lisa Zhang, editors, *Proceedings of the IMA Workshop on Wireless Communications*, IMA Volumes in Mathematics and its Applications. Springer, New York, 2006.

[13] M. Katz and S. Shamai. On the capacity-achieving distribution of the discrete-time non-coherent additive white gaussian noise channel. In *Proc. IEEE Int'l. Symp. Inform. Theory, Lausanne, Switzerland, June 30 - July 5.*, page 165, 2002.

[14] A. Lapidoth and S.M. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels. *IEEE Trans. Inform. Theory*, 49(10):2426–2467, 2003.

[15] R. Palanki. On the capacity-achieving distributions of some fading channels. Presented at 40th Allerton Conference on Communication, Control, and Computing, 2002.

[16] S. Shamai and I. Bar-David. The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Trans. Inform. Theory*, 41(4):1060–1071, 1995.

[17] J. G. Smith. The information capacity of amplitude and variance-constrained scalar gaussian channels. *Inform. Contr.*, 18:203–219, 1971.

[18] S. Verdu. On channel capacity per unit cost. *IEEE Trans. Inform. Theory*, 36(5):1019–1030, 1990.

[19] S. Verdu. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48(6):1319–1343, June 2002.