On the Information Capacity of Gaussian Channels Under Small Peak Power Constraints

Maxim Raginsky
Department of Electrical and Computer Engineering
Duke University, Durham NC, 27708, USA
E-mail: m.raginsky@duke.edu

***Abstract*— This paper deals with information capacities of Gaussian channels under small (but nonvanishing) peak power constraints. We prove that, when the peak amplitude is below 1.05, the capacity of the scalar Gaussian channel is achieved by symmetric equiprobable signaling and is equal to at least 80% of the corresponding average-power capacity. The proof uses the identity of Guo, Shamai and Verdú that relates mutual information and minimum mean square error in Gaussian channels, together with several results on the minimax estimation of a bounded parameter in white Gaussian noise. We also give upper and lower bounds on peak-power capacities of vector Gaussian channels whose inputs are constrained to lie in suitably small ellipsoids and show that we can achieve at least 80% of the average-power capacity by having the transmitters use symmetric equiprobable signaling at amplitudes determined from the usual water-filling policy. The 80% figure comes from an upper bound on the ratio of the nonlinear and the linear minimax risks of estimating a bounded parameter in white Gaussian noise.**

I. Introduction

Consider the scalar Gaussian channel

$$Y = X + Z, \tag{1}$$

where X is a real-valued random input and $Z \sim \mathrm{Normal}(0,1)$ is independent of X. If the input X is restricted to the interval $[-\sqrt{\rho}, \sqrt{\rho}]$ for some $\rho > 0$, we say that the channel is constrained to peak input power ρ. We can define the capacity of (1) under this constraint by

$$C_P(\rho) \triangleq \sup_{X \in [-\sqrt{\rho},\sqrt{\rho}] \text{ w.p.1}} I(X; X+Z), \tag{2}$$

where the supremum is over all random inputs X that are bounded between $-\sqrt{\rho}$ and $\sqrt{\rho}$ with probability one (w.p.1). Channels with input peak power constraints are important from the practical point of view as they describe situations where the transmitter has a limited energy budget. A general solution to (2) is not available. However, Smith [1] showed that, for each ρ, the capacity-achieving distribution is unique and discrete with finite support in $[-\sqrt{\rho}, \sqrt{\rho}]$. He also showed that, for $\sqrt{\rho} \le 0.1$, the capacity is achieved by symmetric equiprobable signaling, i.e., by the distribution that places equal mass on the points $\pm\sqrt{\rho}$.

In this paper, we study the capacity $C_P(\rho)$ for small (but nonvanishing) values of ρ. We use a remarkable recent result of Guo, Shamai and Verdú [2] that relates mutual information in Gaussian channels to minimum mean-squared error (MMSE) estimation, and then connect the problem of finding $C_P(\rho)$ to the problem known in statistics under the heading of *minimax estimation of a bounded normal mean* [3]–[6]. In fact, the Gaussian channel (1) plays as prominent a role in theoretical statistics as it does in information theory. For example, a whole class of nonparametric statistical estimation problems can be reduced to the problem of estimating a sequence of parameters corrupted by additive Gaussian white noise (for details and further references see, e.g., an excellent set of notes by I. Johnstone [7]).

In this paper, we make the following contributions:

- We prove that symmetric equiprobable signaling achieves capacity for all ρ satisfying $\sqrt{\rho} \le 1.05$. This improves on the result of Smith, who gave the bound $\sqrt{\rho} \le 0.1$.
- We prove that, for $\sqrt{\rho} \le 1.05$, the peak-power constrained capacity $C_P(\rho)$ is equal to at least 80% of the corresponding *average-power* constrained capacity

$$C_A(\rho) \triangleq \sup_{X : \mathbb{E}X^2 \le \rho} I(X; X+Z) \equiv \frac{1}{2}\log(1+\rho).$$

- We demonstrate that, for $\sqrt{\rho} \le 1.05$, the relation between the peak- and the average-power constrained capacities $C_P(\rho)$ and $C_A(\rho)$ is analogous to that between nonlinear and linear minimax estimation of a parameter in the interval $[-\sqrt{\rho}, \sqrt{\rho}]$ from an observation corrupted by additive Gaussian noise.

- We derive bounds on the capacity of vector Gaussian channels whose inputs are constrained to lie in suitably small ellipsoids, and show how to do at least as well as the lower bound using a simple suboptimal signaling scheme.

The paper is organized as follows. In the remainder of this section, we define the notation used throughout the paper. Section II contains the necessary background from statistics on the problem of minimax estimation of bounded normal means. Then, in Section III we state and prove our main results on the capacity of Gaussian channels under small peak power constraints.

A. Notation

We shall use the following notation. $\mathcal{M}_n$ will denote the set of all Borel probability measures on $\mathbb{R}^n$; when $n = 1$, we shall simply write $\mathcal{M}$. If a vector $\boldsymbol{a} \in \mathbb{R}^n$ has $a_i \ge 0$ for all $1 \le i \le n$, then we shall write $\boldsymbol{a} \succeq 0$. Given some $\boldsymbol{\rho}, \boldsymbol{w} \in \mathbb{R}^n$ satisfying $\boldsymbol{\rho}, \boldsymbol{w} \succeq 0$, we define the sets

$$\Theta_{\boldsymbol{\rho}} \triangleq \{\boldsymbol{x} \in \mathbb{R}^n : |x_i| \le \sqrt{\rho_i}, 1 \le i \le n\}$$

and

$$\Lambda_{\boldsymbol{w}} \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n w_i x_i^2 \le 1\right\}$$

in $\mathbb{R}^n$. When $n = 1$ and $\rho = 1$, we shall write Θ instead of $\Theta_1 \equiv [-1, 1]$. Given any closed subset S of $\mathbb{R}^n$, we denote by $\mathcal{M}_P(S)$ the set of all $\pi \in \mathcal{M}_n$ with support in S:

$$\mathcal{M}_P(S) \triangleq \{\pi \in \mathcal{M}_n : \mathrm{supp}(\pi) \subseteq S\},$$

and by $\mathcal{M}_A(S)$ the set of all $\pi \in \mathcal{M}_n$, such that $\boldsymbol{X} \sim \pi$ lies in S in the mean square sense:

$$\mathcal{M}_A(S) \triangleq \left\{\pi \in \mathcal{M}_n : \left(\int x_1^2 \pi(\mathrm{d}\boldsymbol{x}), \cdots, \int x_n^2 \pi(\mathrm{d}\boldsymbol{x})\right) \in S\right\}.$$

When $n = 1$ and $S = \Theta$, we shall write $\mathcal{M}_P$ and $\mathcal{M}_A$ instead of $\mathcal{M}_P(\Theta)$ and $\mathcal{M}_A(\Theta)$, respectively.

II. Background on minimax estimation

In this paper, we use several results on minimax estimation of bounded normal means [3]–[7]. This section contains a brief summary of these results.

A. The scalar case

Consider the problem of estimating a deterministic parameter $x \in \Theta_\rho \equiv [-\sqrt{\rho}, \sqrt{\rho}]$ from a single noisy observation $Y = x + Z$, where $Z \sim \mathrm{Normal}(0, 1)$. An estimator $\varphi^* : \mathbb{R} \to \mathbb{R}$ is called *minimax* if

$$\sup_{x \in \Theta_\rho} \mathbb{E}\left\{(x - \varphi^*(Y))^2\right\} = R_N^*(\rho),$$

where

$$R_N^*(\rho) \triangleq \inf_{\varphi} \sup_{x \in \Theta_\rho} \mathbb{E}\left\{(x - \varphi(Y))^2\right\} \quad (3)$$

is the *minimax risk* at ρ. The infimum in (3) is over all measurable $\varphi : \mathbb{R} \to \mathbb{R}$. The subscript N indicates the fact that arbitrary, nonlinear estimators are allowed.

It is not hard to show that the worst-case risk of estimating a *deterministic* $x \in \Theta_\rho$ is equal to the worst-case MSE of estimating a *random* X distributed according to some $\pi \in \mathcal{M}_P(\Theta_\rho)$ independently of Z. Any distribution $\pi \in \mathcal{M}_P(\Theta_\rho)$ that achieves the worst-case MSE is called a *least favorable prior distribution* (or simply least favorable prior) for the problem of estimating a bounded normal mean. Let δ_x denote the Dirac measure (point mass) concentrated at $x \in \mathbb{R}$. Since $\delta_x \in \mathcal{M}_P(\Theta_\rho)$ for every $x \in \Theta_\rho$, and since the set $\mathcal{M}_P(\Theta_\rho)$ is convex, every measurable $\varphi : \mathbb{R} \to \mathbb{R}$ satisfies

$$\begin{aligned}\sup_{x \in \Theta_\rho} & \mathbb{E}\left\{(x - \varphi(Y))^2\right\} \\ &= \sup_{\pi \in \mathcal{M}_P(\Theta_\rho)} \int \mathbb{E}\left\{(x - \varphi(Y))^2\right\} \mathrm{d}\pi(x) \\ &\equiv \sup_{\pi \in \mathcal{M}_P(\Theta_\rho)} R(\varphi, \pi),\end{aligned}$$

where $R(\varphi, \pi) \triangleq \mathbb{E}\left\{(X - \varphi(Y))^2\right\}$ is the MSE of the estimator φ when $X \sim \pi$. Hence, we can express the minimax risk (3) as

$$R_N^*(\rho) = \inf_{\varphi} \sup_{\pi \in \mathcal{M}_P(\Theta_\rho)} R(\varphi, \pi).$$

Since both $\mathcal{M}_P(\Theta_\rho)$ and the set of all measurable $\varphi : \mathbb{R} \to \mathbb{R}$ are convex, we can invoke the minimax theorem [8] (see also Appendix A in [7]) to get

$$R_N^*(\rho) = \sup_{\pi \in \mathcal{M}_P(\Theta_\rho)} \inf_{\varphi} R(\varphi, \pi). \quad (4)$$

For a fixed $\pi \in \mathcal{M}_P(\Theta_\rho)$, the infimum over φ on the right-hand side of (4) is the MMSE achievable by any estimator of $X \sim \pi$ based on $Y = X + Z$, or, using statistical terminology, the *Bayes risk* at π. We shall denote it by $B(\pi)$. An estimator $\varphi : \mathbb{R} \to \mathbb{R}$ that satisfies $R(\varphi, \pi) = B(\pi)$ is said to be *Bayes* relative to

π. It is well-known that the MMSE is achieved by the conditional mean $\varphi_\pi(y) = \mathbb{E}\{X|Y=y\}$:

$$B(\pi) = R(\varphi_\pi, \pi) = \mathbb{E}\left\{(X - \mathbb{E}\{X|Y\})^2\right\}.$$

That is, φ_π is Bayes relative to π.

An important consequence of the minimax theorem is the existence of a *saddlepoint*, i.e., a pair $(\varphi^*_\rho, \pi^*_\rho)$, such that, for any other $\varphi : \mathbb{R} \to \mathbb{R}$ and $\pi \in \mathcal{M}_P(\Theta_\rho)$, we have

$$R(\varphi^*_\rho, \pi) \le R(\varphi^*_\rho, \pi^*_\rho) \le R(\varphi, \pi^*_\rho). \quad (5)$$

The second inequality in (5) implies that $R(\varphi^*_\rho, \pi^*_\rho) = B(\pi^*_\rho)$. Together with the first inequality in (5), this implies that $B(\pi) \le B(\pi^*_\rho)$ for any other $\pi \in \mathcal{M}_P(\Theta_\rho)$. In other words, π^*_ρ has the largest Bayes risk among all $\pi \in \mathcal{M}_P(\Theta_\rho)$. For this reason, π^*_ρ is referred to as the *least favorable prior*. Thus, $R^*_N(\rho) = B(\pi^*_\rho) = R(\varphi^*_\rho, \pi^*_\rho)$. The estimator φ^*_ρ is called the *minimax estimator* for the problem at hand.

No general expression is known for the minimax risk $R^*_N(\rho)$. However, there are some partial results and bounds. For example, from symmetry and analyticity arguments (see, e.g., Ghosh [9]) it follows that the least favorable prior π^*_ρ is unique and discrete with a finite support. Casella and Strawderman [3] proved that, for $\sqrt{\rho} \in (0, 1.05]$, π^*_ρ is given by the symmetric two-point distribution $(1/2)(\delta_{-\sqrt{\rho}} + \delta_{\sqrt{\rho}})$. Moreover, for any $\rho > 0$, we can get upper and lower bounds on $R^*_N(\rho)$ in terms of the *linear* minimax risk

$$\begin{aligned} R^*_L(\rho) &\triangleq \inf_{c\in\mathbb{R}} \sup_{x\in\Theta_\rho} \mathbb{E}\left\{(x - cY)^2\right\} \\ &\equiv \inf_{c\in\mathbb{R}} \sup_{\pi\in\mathcal{M}_P(\Theta_\rho)} \mathbb{E}\left\{(X - cY)^2\right\}. \end{aligned}$$

Using calculus, one can prove that the minimax linear risk and the corresponding minimax linear estimator are given by

$$R^*_L(\rho) = \frac{\rho}{\rho+1} \qquad \text{and} \qquad \varphi^L_\rho(y) = \frac{\rho}{\rho+1} y. \quad (6)$$

It is clear that $R^*_N(\rho) \le R^*_L(\rho)$. A surprising fact is that $R^*_L(\rho)$ is not much larger than $R^*_N(\rho)$. Let $\mu(\rho)$ denote the ratio $R^*_L(\rho)/R^*_N(\rho)$, and let

$$\mu^* \triangleq \sup_{\rho>0} \mu(\rho) \equiv \sup_{\rho>0} \frac{R^*_L(\rho)}{R^*_N(\rho)} \quad (7)$$

be the worst-case ratio of the linear and the nonlinear minimax risks. Donoho, Liu and MacGibbon [6] obtained the bound $\mu^* \le 1.25$ via numerical computation; they dubbed μ^* the *Ibragimov–Has'minskii constant* because Ibragimov and Has'minskii [5] first introduced the function $\mu(\rho)$, described several of its properties, and proved that $\mu^* < +\infty$. Gourdin, Jaumard and MacGibbon [10] later used more sophisticated numerical methods to get more precise bounds $1.246509 \le \mu^* \le 1.246609$.

The minimax linear risk $R^*_L(\rho)$ can be alternatively introduced through the so-called *minimax Bayes* framework, which originated with the work of Pinsker [11] on minimax filtering of smooth functions in white Gaussian noise. Its main idea is to construct an appropriate relaxation of the constraint on the distribution of X, which can lead to useful (and, possibly, asymptotically tight) bounds on the original minimax risk. For the problem at hand, we can apply this reasoning as follows. Instead of assuming that $X \in \Theta_\rho$ w.p.1, let us consider the case when X belongs to Θ_ρ in mean square, i.e., $X \sim \pi$ for some $\pi \in \mathcal{M}_A(\Theta_\rho)$. Then we define the *minimax Bayes risk* at ρ via

$$R^*_B(\rho) \triangleq \inf_{\varphi} \sup_{\pi\in\mathcal{M}_A(\Theta_\rho)} R(\varphi, \pi).$$

Since $\mathcal{M}_P(\Theta_\rho) \subset \mathcal{M}_A(\Theta_\rho)$, we have $R^*_N(\rho) \le R^*_B(\rho)$. Moreover, $\mathcal{M}_A(\Theta_\rho)$ is convex, so we can apply the minimax theorem again to get

$$\begin{aligned} R^*_B(\rho) &= \sup_{\pi\in\mathcal{M}_A(\Theta_\rho)} \inf_{\varphi} R(\varphi, \pi) \\ &= \sup_{\pi\in\mathcal{M}_A(\Theta_\rho)} B(\pi). \end{aligned}$$

Now, let Φ_ρ denote the normal distribution with mean zero and variance ρ. Then $\Phi_\rho \in \mathcal{M}_A(\Theta_\rho)$, and it is an easy consequence of the Cramér–Rao bound [12] that

$$\sup_{\pi\in\mathcal{M}_A(\Theta_\rho)} B(\pi) = B(\Phi_\rho) = \frac{\rho}{\rho+1},$$

which is precisely the linear minimax risk $R^*_L(\rho)$. Thus, the linear minimax risk $R^*_L(\rho)$ is also the minimax Bayes risk $R^*_B(\rho)$, and that the linear minimax estimator in (6) is Bayes relative to the Gaussian prior Φ_ρ.

B. *The vector case*

Now we pass to the vector (or multivariate) case. Consider the problem of the minimax estimation of an n-dimensional vector $\boldsymbol{x} \in \Theta_{\boldsymbol{\rho}}$, where $\boldsymbol{\rho} \succ 0$,[1] from a vector of noisy observations $\boldsymbol{Y} = \boldsymbol{x} + \boldsymbol{Z}$, where $\boldsymbol{Z}$ is a vector of n independent standard Gaussians. Just as before, we can express the corresponding minimax risk in terms of the worst-case MMSE (or Bayes risk) achievable in estimating a random vector $\boldsymbol{X}$ that lies in

[1] We assume that the components of $\boldsymbol{\rho}$ are strictly positive because otherwise we can reduce the dimension of the problem by $|\{1 \le i \le n : \rho_i = 0\}|$.

$\Theta_{\boldsymbol{\rho}}$ w.p. 1 from $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$, where $\boldsymbol{Z}$ is independent of $\boldsymbol{X}$. The corresponding minimax risk is

$$\begin{aligned} R_N^*(\boldsymbol{\rho}) &\triangleq \inf_{\boldsymbol{\varphi}} \sup_{\boldsymbol{x} \in \Theta_{\boldsymbol{\rho}}} \mathbb{E}\, \|\boldsymbol{x} - \boldsymbol{\varphi}(\boldsymbol{Y})\|^2 \\ &\equiv \inf_{\boldsymbol{\varphi}} \sup_{\pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}})} \mathbb{E}\, \|\boldsymbol{X} - \boldsymbol{\varphi}(\boldsymbol{Y})\|^2, \end{aligned}$$

where the infimum is over all measurable $\boldsymbol{\varphi} : \mathbb{R}^n \to \mathbb{R}^n$. Once again, an application of the minimax theorem leads to

$$\begin{aligned} R_N^*(\boldsymbol{\rho}) &= \inf_{\boldsymbol{\varphi}} \sup_{\pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}})} R(\boldsymbol{\varphi}, \pi) \\ &= \sup_{\pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}})} \inf_{\boldsymbol{\varphi}} R(\boldsymbol{\varphi}, \pi), \end{aligned}$$

where $R(\boldsymbol{\varphi}, \pi)$ is the MSE of $\boldsymbol{\varphi}$ relative to π. To characterize the least favorable prior, we note that (i) $\Theta_{\boldsymbol{\rho}}$ is the Cartesian product $\Theta(\rho_1) \times \cdots \times \Theta(\rho_n)$, and (ii) the risk is additive, i.e.,

$$R(\boldsymbol{\varphi}, \pi) = \sum_{i=1}^n \mathbb{E}\left\{ (X_i - \varphi_i(\boldsymbol{Y}))^2 \right\},$$

where $\varphi_i(\boldsymbol{Y})$ is the ith component of $\boldsymbol{\varphi}(\boldsymbol{Y})$. Using this, it is not hard to show (see, e.g., Chapter 6 in [7]) that, if we take any $\pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}})$ and let $\bar{\pi}$ be the product of its marginal distributions, then the Bayes risk of $\bar{\pi}$ will be at least as great as that of π: $B(\bar{\pi}) \ge B(\pi)$. This in turn implies that:

1) The least favorable prior is the product distribution $\pi^*_{\boldsymbol{\rho}} = \pi^*_{\rho_1} \times \cdots \times \pi^*_{\rho_n}$.
2) The corresponding minimax estimator $\boldsymbol{\varphi}^*_{\boldsymbol{\rho}}$ is separable, i.e., $\boldsymbol{\varphi}^*_{\boldsymbol{\rho}}(\boldsymbol{y}) = (\varphi^*_{\rho_1}(y_1), \cdots, \varphi^*_{\rho_n}(y_n))^T$.
3) The minimax risk $R_N^*(\boldsymbol{\rho})$ is the sum of corresponding scalar minimax risks: $R_N^*(\boldsymbol{\rho}) = \sum_{i=1}^n R_N^*(\rho_i)$.

As for the linear minimax risk

$$\begin{aligned} R_L^*(\boldsymbol{\rho}) &\triangleq \inf_{\boldsymbol{C} \in \mathbb{R}^{n \times n}} \sup_{\boldsymbol{x} \in \Theta_{\boldsymbol{\rho}}} \mathbb{E}\, \|\boldsymbol{x} - \boldsymbol{C}\boldsymbol{Y}\|^2 \\ &\equiv \inf_{\boldsymbol{C} \in \mathbb{R}^{n \times n}} \sup_{\pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}})} \mathbb{E}\, \|\boldsymbol{X} - \boldsymbol{C}\boldsymbol{Y}\|^2, \end{aligned}$$

where the infimum is over all $n \times n$ matrices $\boldsymbol{C}$ with real entries, it can be shown that it suffices to consider only diagonal matrices with diagonal entries bounded between 0 and 1, and that $R_L^*(\boldsymbol{\rho}) = \sum_{i=1}^n \rho_i/(\rho_i + 1)$ (see, e.g., Chapter 7 in [7]). Moreover, the linear minimax risk coincides with the minimax Bayes risk

$$\begin{aligned} R_{\mathrm{B}}^*(\boldsymbol{\rho}) &\triangleq \inf_{\boldsymbol{\varphi}} \sup_{\pi \in \mathcal{M}_A(\Theta_{\boldsymbol{\rho}})} R(\boldsymbol{\varphi}, \pi) \\ &= \sup_{\pi \in \mathcal{M}_A(\Theta_{\boldsymbol{\rho}})} B(\pi), \end{aligned}$$

and the minimax linear estimator is Bayes relative to the product Gaussian distribution $\Phi_{\boldsymbol{\rho}} = \Phi_{\rho_1} \times \cdots \times \Phi_{\rho_n}$. Because for each i we have $R_L^*(\rho_i) \le \mu^* R_N^*(\rho_i)$, the Ibragimov–Has'minskii bound applies to this multivariate problem: $R_L^*(\boldsymbol{\rho}) \le \mu^* R_N^*(\boldsymbol{\rho})$ [6].

III. Gaussian channel capacity under small peak power constraints

A. The scalar case

In this section, we prove that the information capacity of the scalar Gaussian channel (1) under the peak power constraint $X \in \Theta_\rho$ w.p. 1 is achieved by symmetric equiprobable signaling when $\sqrt{\rho} \le 1.05$. The capacity of interest $C_P(\rho)$ is given by the value of the convex program

$$\max_{\pi} I(\pi) \quad \text{subject to } \pi \in \mathcal{M}_P(\Theta_\rho), \tag{8}$$

where $I(\pi)$ denotes the mutual information between X and $X + Z$ when $X \sim \pi$. We have the following result:

Theorem 3.1. Consider the Gaussian channel (1). Then for any ρ satisfying the condition $0 < \sqrt{\rho} \le 1.05$, the capacity $C_P(\rho)$ is given by the formula

$$C_P(\rho) = \rho - \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \log\cosh(\rho - \sqrt{\rho}y)\mathrm{d}y. \tag{9}$$

The unique capacity-achieving distribution is given by

$$\pi^\circ_\rho = \frac{1}{2}(\delta_{-\sqrt{\rho}} + \delta_{\sqrt{\rho}}),$$

where δ_x is the Dirac measure (point mass) concentrated at x.

Proof: Given a random variable X and some $\rho > 0$, we can define another random variable $U = X/\sqrt{\rho}$ and treat it as the input to the channel (1): $Y = \sqrt{\rho}U + Z$. Conversely, given a random variable U with distribution $\pi \in \mathcal{M}_P$ and some $\rho > 0$, let $\pi_\rho \in \mathcal{M}_P(\Theta_\rho)$ denote the distribution of $X = \sqrt{\rho}U$. Since $\rho > 0$, we can write

$$I(X; Y) = I(\sqrt{\rho}U; \sqrt{\rho}U + Z) = I(U; \sqrt{\rho}U + Z),$$

so that

$$\begin{aligned} C_P(\rho) &= \sup_{U \in [-1,1]\ \text{w.p.}1} I(U; \sqrt{\rho}U + Z) \\ &= \sup_{\pi \in \mathcal{M}_P} I(\pi, \rho), \end{aligned}$$

where $I(\pi, \rho)$ denotes the mutual information between U and $\sqrt{\rho}U + Z$ when $U \sim \pi$. Moreover, if some $\pi^\circ \in \mathcal{M}_P$ maximizes $I(\pi, \rho)$ over all $\pi \in \mathcal{M}_P$, then its rescaled version $\pi^\circ_\rho \in \mathcal{M}_P(\Theta_\rho)$ maximizes $I(\pi)$ over all $\pi \in \mathcal{M}_P(\Theta_\rho)$. Hence, to find the capacity $C_P(\rho)$ we must solve the convex program

$$\max_\pi I(\pi,\rho) \quad \text{subject to } \pi \in \mathcal{M}_P.$$

Now, we recall the fundamental identity of Guo, Shamai and Verdú [2]. Let $B(\pi,\rho)$ denote the MMSE (or Bayes risk) achievable in estimating $U \sim \pi$ based on $Y = \sqrt{\rho}U + Z$:

$$\begin{aligned} B(\pi,\rho) &\triangleq \inf_\varphi \mathbb{E}\left\{(U-\varphi(Y))^2\right\} \\ &\equiv \mathbb{E}\left\{(U-\mathbb{E}\{U|\sqrt{\rho}U+Z\})^2\right\}. \end{aligned}$$

Then we have

$$\frac{\mathrm{d}}{\mathrm{d}\rho} I(\pi,\rho) = \frac{1}{2}B(\pi,\rho) \tag{10}$$

for any $\pi \in \mathcal{M}$ with a finite second moment.[2] Integrating both sides of Eq. (10) with respect to ρ and using the fact that $I(\pi,0)=0$ for any π, we obtain

$$I(\pi,\rho) = \frac{1}{2}\int_0^\rho B(\pi,\lambda)\mathrm{d}\lambda. \tag{11}$$

Given $\lambda > 0$, we let $W = \sqrt{\lambda}U$ and write

$$\begin{aligned} B(\pi,\lambda) &= \mathbb{E}\left\{(U-\mathbb{E}\{U|\sqrt{\lambda}U+Z\})^2\right\} \\ &= \frac{1}{\lambda}\mathbb{E}\left\{(W-\mathbb{E}\{W|W+Z\})^2\right\} \\ &= \frac{1}{\lambda}B(\pi_\lambda), \end{aligned} \tag{12}$$

where $B(\pi_\lambda)$ denotes the Bayes risk of estimating $W \sim \pi_\lambda$ based on $Y = W + Z$. Thus, we can write

$$I(\pi,\rho) = \frac{1}{2}\int_0^\rho \frac{B(\pi_\lambda)}{\lambda}\mathrm{d}\lambda.$$

Since $B(\pi_\lambda) \le R_N^*(\lambda)$, we obtain the bound

$$I(\pi,\rho) \le \frac{1}{2}\int_0^\rho \frac{R_N^*(\lambda)}{\lambda}\mathrm{d}\lambda. \tag{13}$$

Now we apply the result of Casella and Strawderman [3], which says that, for $\sqrt{\lambda} \le 1.05$, $R_N^*(\lambda) = B(\pi_\lambda^\circ)$, where $\pi_\lambda^\circ = (1/2)(\delta_{-\sqrt{\lambda}} + \delta_{\sqrt{\lambda}})$. According to our notation, π_λ° is obtained by rescaling the distribution $\pi^\circ = (1/2)(\delta_{-1}+\delta_1)$ to the interval $[-\sqrt{\lambda},\sqrt{\lambda}]$. Therefore, if $\sqrt{\rho} \le 1.05$, we can write

$$\begin{aligned} \frac{1}{2}\int_0^\rho \frac{R_N^*(\lambda)}{\lambda}\mathrm{d}\lambda &= \frac{1}{2}\int_0^\rho \frac{B(\pi_\lambda^\circ)}{\lambda}\mathrm{d}\lambda \\ &= \frac{1}{2}\int_0^\rho B(\pi^\circ,\lambda)\mathrm{d}\lambda \\ &= I(\pi^\circ,\rho), \end{aligned}$$

where the second equality follows from (12) and the third one from (11). This, together with (13), implies that $I(\pi,\rho) \le I(\pi^\circ,\rho)$ whenever $\sqrt{\rho} \le 1.05$. In other words,

$$I(\pi^\circ,\rho) = \max_{\pi\in\mathcal{M}_P} I(\pi,\rho)$$

and, consequently,

$$I(\pi_\rho^\circ) = \max_{\pi\in\mathcal{M}_P(\Theta_\rho)} I(\pi) \equiv C_P(\rho).$$

Hence, π_ρ° drives the channel (1) at capacity under the peak power constraint $X \in [-\sqrt{\rho},\sqrt{\rho}]$ w.p.1. We know from [1] that the capacity-achieving distribution [i.e., the maximizer of (8)] is unique. The expression (9) is from [2]. ■

We can also consider the following relaxation of (8):

$$\max_\pi I(\pi) \quad \text{subject to } \pi \in \mathcal{M}_A(\Theta_\rho). \tag{14}$$

The value of this program is, of course, the information capacity $C_A(\rho)$ of (1) under the average power constraint $\mathbb{E}\,X^2 \le \rho$. It is well known that

$$C_A(\rho) = \frac{1}{2}\log(1+\rho),$$

and that the capacity-achieving distribution is the Gaussian distribution Φ_ρ [13]. It is evident from the form of the programs (8) and (14) that the relationship between the peak-power and the average-power capacities C_P and C_A is analogous to that between the minimax and the minimax Bayes risks for the problem of estimating a bounded normal mean. In fact, for small peak power constraints ($\sqrt{\rho} \le 1.05$), this analogy can be made precise. To begin with, it is easy to verify that the following formula holds for *all* values of ρ:

$$C_A(\rho) = \frac{1}{2}\int_0^\rho \frac{R_L^*(\lambda)}{\lambda}\mathrm{d}\lambda. \tag{15}$$

Moreover, the capacity-achieving Gaussian distribution Φ_ρ is also the least favorable prior for the minimax Bayes problem over $\mathcal{M}_A(\Theta_\rho)$. On the other hand, for $\sqrt{\rho} \le 1.05$, the capacity-achieving distribution π_ρ° is also the least favorable prior for the minimax problem over $\mathcal{M}_P(\Theta_\rho)$. These observations lead to the following corollary:

[2]There is a caveat here: π cannot depend on ρ or, equivalently, its cumulative distribution function cannot have ρ as a parameter. This restriction should be understood in the sense that

$$\frac{\partial}{\partial\lambda} I(\pi,\lambda)\Big|_{\lambda=\rho} = \frac{1}{2}B(\pi,\rho).$$

Corollary 3.2. Whenever $\sqrt{\rho} \le 1.05$, we have

$$\frac{1}{\mu^*} C_A(\rho) \le C_P(\rho) \le C_A(\rho),$$

where μ^* is the Ibragimov–Has'minskii constant defined in (7).

Proof: We have already established that

$$C_P(\rho) = \frac{1}{2} \int_0^\rho \frac{R_N^*(\lambda)}{\lambda} \mathrm{d}\lambda, \qquad \sqrt{\rho} \le 1.05.$$

The integrand can be bounded from below by $R_L^*(\lambda)/(\mu^* \lambda)$. This, together with Eq. (15), proves the lower bound on C_P. The upper bound is obvious. ■

From the work of Donoho, Liu and MacGibbon [6], we know that $\mu^* \le 1.25$. Therefore, the small peak power capacity of the Gaussian channel is at least 80% of the corresponding average power capacity. We also note that Eq. (13) leads to the bound

$$C_P(\rho) \le \frac{1}{2} \int_0^\rho \frac{R_N^*(\lambda)}{\lambda} \mathrm{d}\lambda,$$

which is tighter than the upper bound in Corollary 3.2, but does not admit a simple analytical expression.

An interesting question is whether the distribution that achieves $C_P(\rho)$ always coincides with the least favorable prior π_ρ^* that achieves $R_N^*(\rho)$. Intuitively, it makes sense to conjecture that it does: for both problems, the corresponding extremal distributions are discrete and symmetric with finite support, and the spacing between the points of support should be as large as possible in order to ensure that the they can be reliably distinguished from one another in the presence of noise (in the information capacity case) and that the prior is as "uninformative" as possible about the true value of the mean (in the bounded mean estimation case). However, the answer turns out to be negative. As shown by Casella and Strawderman [3], the symmetric two-point prior π_ρ° is no longer least favorable for $\sqrt{\rho} > 1.05$. In fact, the least favorable prior for $\sqrt{\rho} > 1.05$ has at least three points of support, and the weights of these points vary as ρ increases. On the other hand, as shown recently by Sharma and Shamai [14], symmetric equiprobable signaling achieves $C_P(\rho)$ for all $\sqrt{\rho} \le 1.671$; when $\sqrt{\rho} > 1.671$, the capacity-achieving distribution has at least three points of support. Thus, the capacity-achieving distribution coincides with the least favorable prior only for $\sqrt{\rho} \le 1.05$.

These considerations also imply that the method used to prove Theorem 3.1 would not work for $\sqrt{\rho} > 1.05$. Indeed, our approach makes use of the fact that the least favorable prior for $\sqrt{\rho} \le 1.05$ is independent of ρ, which permits the application of the Guo–Shamai–Verdú identity (10). When $\sqrt{\rho} > 1.05$, however, the least favorable prior depends on ρ [3]. (A "threshold effect" of this kind has been noted before in similar contexts [15].) On the other hand, by relating the problem of achieving the capacity of (1) under a peak power constraint to the problem of minimax estimation of a bounded normal mean, we were also able to show that the peak-power capacity for peak amplitudes below 1.05 is equal to at least 80% of the corresponding average-power capacity.

B. The vector case

We now consider vector Gaussian channels of the form

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}, \tag{16}$$

where the input $\boldsymbol{X}$ is an n-dimensional real vector, and $\boldsymbol{Z}$ is a vector of n independent standard Gaussian variables that are also independent of $\boldsymbol{X}$. We are interested in peak power constraints defined by an ellipsoid $\Lambda_{\boldsymbol{w}} \subset \mathbb{R}^n$ for some $\boldsymbol{w} \succ 0$:

$$\sum_{i=1}^n w_i X_i^2 \le 1 \quad \text{w.p.1.}$$

This can be used to model a MIMO communication system with n transmitters and n receivers operating under a fixed transmit energy budget, where the weights $\{w_i\}$ control the amount of power allocated to individual transmitters. The corresponding information capacity, which we denote by $C_P[\boldsymbol{w}]$, is the value of the convex program

$$\max_\pi I(\pi) \qquad \text{subject to } \pi \in \mathcal{M}_P(\Lambda_{\boldsymbol{w}}), \tag{17}$$

where $I(\pi)$ stands for the mutual information $I(\boldsymbol{X}; \boldsymbol{X} + \boldsymbol{Z})$ when $\boldsymbol{X} \sim \pi$. Unlike in the scalar case, we do not have an exact solution of (17), even when the ellipsoid $\Lambda_{\boldsymbol{w}}$ is suitably "small." We will, however, derive a useful lower bound on $C_P[\boldsymbol{w}]$ under some additional conditions on $\boldsymbol{w}$.

As before, let us consider the following relaxation of (17):

$$\max_\pi I(\pi) \qquad \text{subject to } \pi \in \mathcal{M}_A(\Lambda_{\boldsymbol{w}}),$$

whose value we denote by $C_A[\boldsymbol{w}]$. The capacity-achieving distribution in this case is the Gaussian product distribution $\Phi_{\boldsymbol{\rho}^\circ} = \Phi_{\rho_1^\circ} \times \ldots \times \Phi_{\rho_n^\circ}$, where the variances $\{\rho_i^\circ\}$ are given by the well-known *waterfilling* solution

$$\rho_i^\circ = (\alpha/w_i - 1)^+, \qquad 1 \le i \le n \tag{18}$$

where α is determined from the equation $\sum_{i=1}^{n}(\alpha - w_i)^+ = 1$ [13]. Moreover, a given point $\boldsymbol{\rho} \succeq 0$ satisfies

$$\sum_{i=1}^{n} w_i \rho_i \le 1$$

if and only if the hyperrectangle $\Theta_{\boldsymbol{\rho}}$ lies within $\Lambda_{\boldsymbol{w}}$. Consequently,

$$\begin{aligned} C_A[\boldsymbol{w}] &= \frac{1}{2} \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} \sum_{i=1}^{n} \log(1+\rho_i) \\ &= \frac{1}{2} \sum_{i=1}^{n} \log(1+\rho_i^\circ). \end{aligned}$$

First, we consider a simpler case of *hyperrectangular* constraints. Namely, given any $\boldsymbol{\rho} \in \mathbb{R}^n$ satisfying the condition $\boldsymbol{\rho} \succeq 0$, let $C_P(\boldsymbol{\rho})$ denote the value of the program

$$\max_{\pi} I(\pi) \qquad \text{subject to } \pi \in \mathcal{M}_P(\Theta_{\boldsymbol{\rho}}).$$

Consider also the corresponding relaxation

$$\max_{\pi} I(\pi) \qquad \text{subject to } \pi \in \mathcal{M}_A(\Theta_{\boldsymbol{\rho}}),$$

whose value we denote by $C_A(\boldsymbol{\rho})$. Then we have the following:

Proposition 3.3. The capacity $C_P(\boldsymbol{\rho})$ is equal to

$$C_P(\boldsymbol{\rho}) = \sum_{i=1}^{n} C_P(\rho_i), \tag{19}$$

where $C_P(\rho)$ is the value of (8). Suppose now that $\boldsymbol{\rho}$ satisfies

$$\|\boldsymbol{\rho}\|_\infty^{1/2} \equiv \left(\max_{1\le i\le n} \rho_i\right)^{1/2} \le 1.05.$$

Then each term on the right-hand side of (19) can be computed using (9), and the capacity-achieving distribution is unique and equal to the product measure $\pi^\circ_{\rho_1} \times \cdots \times \pi^\circ_{\rho_n}$. Moreover,

$$C_A(\boldsymbol{\rho}) = \sum_{i=1}^{n} C_A(\rho_i) = \frac{1}{2}\sum_{i=1}^{n} \log(1+\rho_i),$$

and we have the bounds $(1/\mu^*)C_A(\boldsymbol{\rho}) \le C_P(\boldsymbol{\rho}) \le C_A(\boldsymbol{\rho})$, where μ^* is the Ibragimov–Has'minskii constant.

Proof: The proposition follows immediately from Theorem 3.1 and the observation that, since the Z_i are i.i.d. and also independent of $\boldsymbol{X}$, the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ can be bounded as

$$I(\boldsymbol{X};\boldsymbol{Y}) \le \sum_{i=1}^{n} I(X_i;Y_i) = \sum_{i=1}^{n} I(X_i; X_i + Z_i),$$

where equality holds if and only if the X_i are mutually independent. ■

Just as in the scalar case, the capacity-achieving distribution (for $\|\boldsymbol{\rho}\|_\infty^{1/2} \le 1.05$) coincides with the least favorable prior for the corresponding minimax problem, and the peak-power capacity can be expressed in terms of the minimax risks as

$$C_P(\boldsymbol{\rho}) = \frac{1}{2}\sum_{i=1}^{n} \int_0^{\rho_i} \frac{R_N^*(\lambda)}{\lambda} \mathrm{d}\lambda. \tag{20}$$

Similarly, the average-power capacity can be expressed in terms of the linear minimax risks as

$$C_A(\boldsymbol{\rho}) = \frac{1}{2}\sum_{i=1}^{n} \int_0^{\rho_i} \frac{R_L^*(\lambda)}{\lambda} \mathrm{d}\lambda, \tag{21}$$

and the corresponding capacity-achieving distribution $\Phi_{\rho_1} \times \ldots \times \Phi_{\rho_n}$ coincides with the least favorable prior for the appropriate minimax Bayes problem. In fact, the formula (21) holds for all $\boldsymbol{\rho} \succeq 0$.

We now turn to $C_P[\boldsymbol{w}]$ and $C_A[\boldsymbol{w}]$. The average-power capacity $C_A[\boldsymbol{w}]$ can be expressed as

$$\begin{aligned} C_A[\boldsymbol{w}] &= \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} C_A(\boldsymbol{\rho}) \\ &= \frac{1}{2} \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} \sum_{i=1}^{n} \log(1+\rho_i) \\ &= \frac{1}{2} \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} \sum_{i=1}^{n} \int_0^{\rho_i} \frac{R_L^*(\lambda)}{\lambda} \mathrm{d}\lambda \\ &= \frac{1}{2}\sum_{i=1}^{n} \int_0^{\rho_i^\circ} \frac{R_L^*(\lambda)}{\lambda} \mathrm{d}\lambda, \end{aligned}$$

where $\rho_1^\circ, \ldots, \rho_n^\circ$ are given by the waterfilling policy (18). As for $C_P[\boldsymbol{w}]$, we have the following:

Theorem 3.4. Suppose that $\boldsymbol{w} \succ 0$ satisfies

$$\left(\max_{1\le i\le n} \frac{1}{w_i}\right)^{1/2} \le 1.05. \tag{22}$$

Then we have the bounds

$$\frac{1}{\mu^*} C_A[\boldsymbol{w}] \le C_P[\boldsymbol{w}] \le C_A[\boldsymbol{w}],$$

where μ^* is the Ibragimov–Has'minskii constant. Moreover, the product distribution $\tilde{\pi}_{\boldsymbol{\rho}^\circ} = \pi^\circ_{\rho_1^\circ} \times \ldots \times \pi^\circ_{\rho_n^\circ}$ achieves at least 80% of the average-power constrained capacity:

$$I(\tilde{\pi}_{\boldsymbol{\rho}^\circ}) \ge \frac{1}{\mu^*} C_A[\boldsymbol{w}] \ge 0.8 C_A[\boldsymbol{w}].$$

Proof: We first observe that, if (22) holds, then $\Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}$ for some $\boldsymbol{\rho} \succeq 0$ if and only if $\|\boldsymbol{\rho}\|_\infty^{1/2} \le 1.05$. Therefore, we can write

$$\begin{aligned} C_P[\boldsymbol{w}] &\ge \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} C_P(\boldsymbol{\rho}) \\ &= \sup \left\{ C_P(\boldsymbol{\rho}) : \|\boldsymbol{\rho}\|_\infty^{1/2} \le 1.05 \right\}. \end{aligned}$$

Applying Proposition 3.3 to each term in the second supremum, we can write

$$\begin{aligned} C_P[\boldsymbol{w}] &\ge \frac{1}{\mu^*} \sup \left\{ C_A(\boldsymbol{\rho}) : \|\boldsymbol{\rho}\|_\infty^{1/2} \le 1.05 \right\} \\ &= \frac{1}{\mu^*} \sup_{\boldsymbol{\rho}\succeq 0, \Theta_{\boldsymbol{\rho}} \subset \Lambda_{\boldsymbol{w}}} C_A(\boldsymbol{\rho}) \\ &\equiv \frac{1}{\mu^*} C_A[\boldsymbol{w}]. \end{aligned}$$

This proves the lower bound on $C_P[\boldsymbol{w}]$, the upper bound is obvious. Finally, using Proposition 3.3 and Eqs. (20) and (21), we can write

$$\begin{aligned} I(\tilde{\pi}_{\boldsymbol{\rho}^\circ}) &= C_P(\boldsymbol{\rho}^\circ) \\ &= \frac{1}{2} \sum_{i=1}^{n} \int_0^{\rho_i^\circ} \frac{R_N^*(\lambda)}{\lambda} \mathrm{d}\lambda \\ &\ge \frac{1}{2\mu^*} \sum_{i=1}^{n} \int_0^{\rho_i^\circ} \frac{R_L^*(\lambda)}{\lambda} \mathrm{d}\lambda \\ &= \frac{1}{\mu^*} C_A(\boldsymbol{\rho}^\circ) \\ &= \frac{1}{\mu^*} C_A[\boldsymbol{w}]. \end{aligned}$$

The theorem is proved. ■

Although we were not able to derive an exact expression for the peak-power capacity $C_P[\boldsymbol{w}]$, we have shown that, provided the ellipsoid $\Lambda_{\boldsymbol{w}}$ is sufficiently small, $C_P[\boldsymbol{w}]$ is equal to at least 80% of the average-power capacity $C_A[\boldsymbol{w}]$. Moreover, we can do at least as well as this lower bound if all the transmitters use symmetric equiprobable signaling at amplitudes determined from the waterfilling power allocation policy for achieving $C_A[\boldsymbol{w}]$. This scheme, however, does not achieve $C_P[\boldsymbol{w}]$ because the capacity-achieving distribution is known to be a finite mixture of uniform distributions on concentric ellipsoidal shells [16]. The least favorable prior for the corresponding minimax problem has the same structure [4]. An interesting problem for future work is to determine whether these two extremal probability distributions coincide when (22) holds.

ACKNOWLEDGMENT

The author would like to thank Andrew R. Barron for the reference to I. Johnstone's lecture notes on functional estimation and Gaussian sequence models, and Shlomo Shamai and Sergio Verdú for their comments and suggestions.

REFERENCES

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

[2] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual informaiton and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[3] G. Casella and W. E. Strawderman, "Estimating a bounded normal mean," *Ann. Statist.*, vol. 9, no. 4, pp. 870–878, 1981.

[4] P. J. Bickel, "Minimax estimation of the mean of a normal distribution when the parameter space is restricted," *Ann. Statist.*, vol. 9, no. 6, pp. 1301–1309, 1981.

[5] I. A. Ibragimov and R. Z. Has'minskii, "On nonparametric estimation of the value of a linear functional in Gaussian white noise," *Theory Probab. Appl.*, vol. 29, no. 1, pp. 18–32, 1984.

[6] D. L. Donoho, R. C. Liu, and B. MacGibbon, "Minimax risk over hyperrectangles, and implications," *Ann. Statist.*, vol. 18, no. 3, pp. 1416–1437, 1990.

[7] I. M. Johnstone, "Function estimation and Gaussian sequence models," 2002, available [Online] at http://www-stat.stanford.edu/~imj/baseb.pdf.

[8] M. Sion, "On general minimax theorems," *Pacific J. Math.*, vol. 8, pp. 171–176, 1958.

[9] M. N. Ghosh, "Uniform approximation of minimax point estimates," *Ann. Statist.*, vol. 35, pp. 1031–1047, 1964.

[10] E. Gourdin, B. Jaumard, and B. MacGibbon, "Global optimization decomposition methods for bounded parameter minimax risk evaluation," *SIAM J. Sci. Comput.*, vol. 15, no. 1, pp. 16–35, January 1994.

[11] M. S. Pinsker, "Optimal filtering of square integrable signals in Gaussian white noise," *Problems Inform. Transmission*, vol. 16, pp. 120–123, 1980.

[12] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Berlin: Springer, 1994.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[14] N. Sharma and S. Shamai (Shitz), "Characterization of the discrete-capacity achieving distribution when mass points increases," in *Proc. Int. Symp. on Information Theory and Its Applications (ISITA)*, Auckland, New Zealand, December 2008, to appear.

[15] L. D. Brown and R. C. Liu, "Bounds on the Bayes and minimax risk for signal parameter estimation," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1386–1394, July 1993.

[16] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inform. Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.