

# Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error

Dongning Guo, *Member, IEEE*, Yihong Wu, *Student Member, IEEE*, Shlomo Shamai (Shitz), *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

**Abstract**—Consider the minimum mean-square error (MMSE) of estimating an arbitrary random variable from its observation contaminated by Gaussian noise. The MMSE can be regarded as a function of the signal-to-noise ratio (SNR) as well as a functional of the input distribution (of the random variable to be estimated). It is shown that the MMSE is concave in the input distribution at any given SNR. For a given input distribution, the MMSE is found to be infinitely differentiable at all positive SNR, and in fact a real analytic function in SNR under mild conditions. The key to these regularity results is that the posterior distribution conditioned on the observation through Gaussian channels always decays at least as quickly as some Gaussian density. Furthermore, simple expressions for the first three derivatives of the MMSE with respect to the SNR are obtained. It is also shown that, as functions of the SNR, the curves for the MMSE of a Gaussian input and that of a non-Gaussian input cross at most once over all SNRs. These properties lead to simple proofs of the facts that Gaussian inputs achieve both the secrecy capacity of scalar Gaussian wiretap channels and the capacity of scalar Gaussian broadcast channels, as well as a simple proof of the entropy power inequality in the special case where one of the variables is Gaussian.

**Index Terms**—Entropy, estimation, Gaussian broadcast channel, Gaussian noise, Gaussian wiretap channel, minimum mean square error (MMSE), mutual information.

## I. INTRODUCTION

THE concept of mean-square error has assumed a central role in the theory and practice of estimation since the time of Gauss and Legendre. In particular, minimization of mean-square error underlies numerous methods in statistical sciences. The focus of this paper is the minimum mean-square

error (MMSE) of estimating an arbitrary random variable contaminated by additive Gaussian noise.

Let  $(X, Y)$  be random variables with arbitrary joint distribution. Throughout the paper,  $E\{\cdot\}$  denotes the expectation with respect to the joint distribution of all random variables in the braces, and  $E\{X|Y\}$  denotes the conditional mean estimate of  $X$  given  $Y$ . The corresponding conditional variance is a function of  $Y$  which is denoted by

$$\text{var}\{X|Y\} = E\{(X - E\{X|Y\})^2|Y\}. \quad (1)$$

It is well known that the conditional mean estimate is optimal in the mean-square sense. In fact, the MMSE of estimating  $X$  given  $Y$  is nothing but the average conditional variance:

$$\text{mmse}(X|Y) = E\{\text{var}\{X|Y\}\}. \quad (2)$$

In this paper, we are mainly interested in random variables related through models of the following form:

$$Y = \sqrt{\text{snr}}X + N \quad (3)$$

where  $N \sim \mathcal{N}(0, 1)$  is standard Gaussian throughout this paper unless otherwise stated. The MMSE of estimating the *input*  $X$  of the model given the noisy *output*  $Y$  is alternatively denoted by

$$\text{mmse}(X, \text{snr}) = \text{mmse}(X|\sqrt{\text{snr}}X + N) \quad (4)$$

$$= E\{(X - E\{X|\sqrt{\text{snr}}X + N\})^2\}. \quad (5)$$

The MMSE (4) can be regarded as a function of the signal-to-noise ratio (SNR) for every given distribution  $P_X$ , and as a functional of the input distribution  $P_X$  for every given SNR.<sup>1</sup> In particular, for a Gaussian input with mean  $m$  and variance  $\sigma_X^2$ , denoted by  $X \sim \mathcal{N}(m, \sigma_X^2)$

$$\text{mmse}(X, \text{snr}) = \frac{\sigma_X^2}{1 + \sigma_X^2 \text{snr}}. \quad (6)$$

If  $X$  is equally likely to take  $\pm 1$ , then

$$\text{mmse}(X, \text{snr}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \tanh(\text{snr} - \sqrt{\text{snr}}y) dy. \quad (7)$$

The function  $\text{mmse}(X, \text{snr})$  is illustrated in Fig. 1 for four special inputs: the standard Gaussian variable, a Gaussian variable

<sup>1</sup>Note that  $\text{snr}$  in (3) coincides with the usual notion of signal-to-noise power ratio only if  $E\{X^2\} = 1$ . For simplicity, we refer to  $\text{snr}$  as SNR regardless of the input power.

Manuscript received June 19, 2008; revised April 19, 2010; accepted August 21, 2010. Date of current version March 16, 2011. This work was supported in part by the National Science Foundation (NSF) under Grants CCF-0644344 and CCF-0635154 and in part by the Binational U.S.-Israel Scientific Foundation. This work was presented in part at the IEEE International Symposium on Information Theory, Toronto, ON, Canada, July 2008.

D. Guo is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA (e-mail: dGuo@northwestern.edu).

Y. Wu and S. Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: yihongwu@princeton.edu; verdu@princeton.edu).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by H. Bölcskei, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2011.2111010

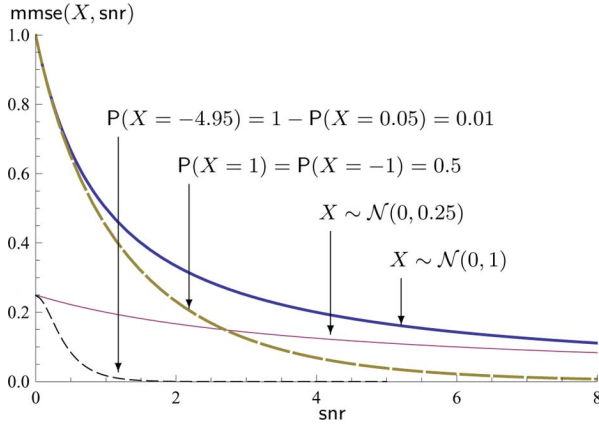


Fig. 1. MMSE of Gaussian and binary inputs as a function of the SNR.

with variance  $1/4$ , as well as symmetric and asymmetric binary random variables, all of zero mean.

Optimal estimation underlies many fundamental information theoretic results, which describe the boundary between what is achievable and what is not. In addition, simple quantitative connections between the MMSE and information measures are revealed in [1]. One such result is that, for arbitrary but fixed  $P_X$

$$\text{mmse}(X, \text{snr}) = 2 \frac{d}{d\text{snr}} I(X; \sqrt{\text{snr}}X + N) \quad (8)$$

for every  $\text{snr} \geq 0$ . This relationship implies the following integral expression for the mutual information:

$$I(X; \sqrt{\text{snr}}g(X) + N) = \frac{1}{2} \int_0^{\text{snr}} \text{mmse}(g(X), \gamma) d\gamma \quad (9)$$

which holds for any one-to-one real-valued function  $g$ . By sending  $\text{snr} \rightarrow \infty$  in (9), the entropy of every discrete random variable  $X$  can be expressed as (see [1] and [2]) follows:

$$H(X) = \frac{1}{2} \int_0^{\infty} \text{mmse}(g(X), \gamma) d\gamma \quad (10)$$

whereas the differential entropy of any continuous random variable  $X$  can be expressed as

$$h(X) = \frac{1}{2} \int_0^{\infty} \text{mmse}(g(X), \gamma) - \frac{1}{2\pi e + \gamma} d\gamma. \quad (11)$$

The preceding information-estimation relationships have found a number of applications, e.g., in nonlinear filtering [1], [3], in multiuser detection [4], in power allocation over parallel Gaussian channels [5], [6], in the proof of Shannon's entropy power inequality (EPI) and its generalizations [2], [7], [8], and in the treatment of the capacity region of several multiuser channels [9]–[11]. Relationships between relative entropy and mean-square error are also found in [12] and [13]. Moreover, many such results have been generalized to vector-valued inputs and multiple-input multiple-output (MIMO) models [1], [7], [14].

Partially motivated by the important role played by the MMSE in information theory, this paper presents a detailed study of the key mathematical properties of  $\text{mmse}(X, \text{snr})$ . The remainder of the paper is organized as follows.

In Section II, we establish bounds on the MMSE as well as on the conditional and unconditional moments of the conditional mean estimation error. In particular, it is shown that the tail of the posterior distribution of the input given the observation vanishes at least as quickly as that of some Gaussian density. Simple properties of input shift and scaling are also shown.

In Section III,  $\text{mmse}(X, \text{snr})$  is shown to be an infinitely differentiable function of  $\text{snr}$  on  $(0, \infty)$  for every input distribution regardless of the existence of its moments (even the mean and variance of the input can be infinite). Furthermore, under certain conditions, the MMSE is found to be real analytic at all positive SNRs, and hence can be arbitrarily well-approximated by its Taylor series expansion.

In Section IV, the first three derivatives of the MMSE with respect to the SNR are expressed in terms of the average central moments of the input conditioned on the output. The result is then extended to the conditional MMSE. We note that a contemporary work [15] derives the Jacobian (a generalized first derivative) of the MMSE in the context of vector Gaussian channels.

Section V shows that the MMSE is concave in the distribution  $P_X$  at any given SNR. The monotonicity of the MMSE of a partial sum of independent identically distributed (i.i.d.) random variables is also investigated. It is well known that the MMSE of a non-Gaussian input is dominated by the MMSE of a Gaussian input of the same variance. It is further shown in this paper that the MMSE curve of a non-Gaussian input and that of a Gaussian input cross each other at most once over  $\text{snr} \in (0, \infty)$ , regardless of their variances.

In Section VI, properties of the MMSE are used to establish Shannon's EPI in the special case where one of the variables is Gaussian. Sidestepping the EPI, the properties of the MMSE lead to simple and natural proofs of the fact that Gaussian input is optimal for both the Gaussian wiretap channel and the scalar Gaussian broadcast channel.

## II. BASIC PROPERTIES

### A. The MMSE

The input  $X$  and the observation  $Y$  in the model described by  $Y = \sqrt{\text{snr}}X + N$  are tied probabilistically by the conditional Gaussian probability density function

$$p_{Y|X}(y|x; \text{snr}) = \varphi(y - \sqrt{\text{snr}}x) \quad (12)$$

where  $\varphi$  stands for the standard Gaussian density

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \quad (13)$$

Let us define for every  $a \in \mathbb{R}$  and  $i = 0, 1, \dots$

$$h_i(y; a) = \mathbb{E}\{X^i \varphi(y - aX)\} \quad (14)$$

which is always well defined because  $\varphi(y - ax)$  is bounded and vanishes quadratic exponentially fast as either  $x$  or  $y$  becomes large with the other variable bounded. In particular,  $h_0(y; \sqrt{\text{snr}})$  is nothing but the marginal distribution of the observation  $Y$ ,

which is always strictly positive. The conditional mean estimate can be expressed as [1], [4]:

$$\mathbb{E}\{X|Y = y\} = \frac{h_1(y; \sqrt{\text{snr}})}{h_0(y; \sqrt{\text{snr}})} \quad (15)$$

and the MMSE can be calculated as [4]

$$\text{mmse}(X, \text{snr}) = \int \int_{\mathbb{R}} \left( x - \frac{h_1(y; \sqrt{\text{snr}})}{h_0(y; \sqrt{\text{snr}})} \right)^2 \varphi(y - \sqrt{\text{snr}} x) dy dP_X(x) \quad (16)$$

which can be simplified if  $\mathbb{E}\{X^2\} < \infty$

$$\text{mmse}(X, \text{snr}) = \mathbb{E}\{X^2\} - \int_{-\infty}^{\infty} \frac{h_1^2(y; \sqrt{\text{snr}})}{h_0(y; \sqrt{\text{snr}})} dy. \quad (17)$$

Note that the estimation error  $X - \mathbb{E}\{X|Y\}$  remains the same if  $X$  is subject to a constant shift. Hence, the following well-known fact:

*Proposition 1:* For every random variable  $X$  and  $a \in \mathbb{R}$

$$\text{mmse}(X + a, \text{snr}) = \text{mmse}(X, \text{snr}). \quad (18)$$

The following scaling property is also straightforward from the definition of MMSE.

*Proposition 2:* For every random variable  $X$  and  $a \in \mathbb{R}$ ,

$$\text{mmse}(aX, \text{snr}) = a^2 \text{mmse}(X, a^2 \text{snr}). \quad (19)$$

### B. The Conditional MMSE and SNR Increment

For any pair of jointly distributed variables  $(X, U)$ , the conditional MMSE of estimating  $X$  at SNR  $\gamma \geq 0$  given  $U$  is defined as

$$\text{mmse}(X, \gamma|U) = \mathbb{E}\{(X - \mathbb{E}\{X|\sqrt{\gamma}X + N, U\})^2\} \quad (20)$$

where  $N \sim \mathcal{N}(0, 1)$  is independent of  $(X, U)$ . It can be regarded as the MMSE achieved with side information  $U$  available to the estimator. For every  $u$ , let  $X_u$  denote a random variable indexed by  $u$  with distribution  $P_{X|U=u}$ . Then the conditional MMSE can be seen as an average

$$\text{mmse}(X, \text{snr}|U) = \int \text{mmse}(X_u, \text{snr}) P_U(du). \quad (21)$$

A special type of conditional MMSE is obtained when the side information is itself a noisy observation of  $X$  through an independent additive Gaussian noise channel. It has long been noticed that two independent looks through Gaussian channels is equivalent to a single look at the sum SNR, e.g., in the context of maximum-ratio combining. As far as the MMSE is concerned, the SNRs of the direct observation and the side information simply add up.

*Proposition 3:* For every  $X$  and every  $\text{snr}, \gamma \geq 0$

$$\text{mmse}(X, \gamma|\sqrt{\text{snr}}X + N) = \text{mmse}(X, \text{snr} + \gamma) \quad (22)$$

where  $N \sim \mathcal{N}(0, 1)$  is independent of  $X$ .

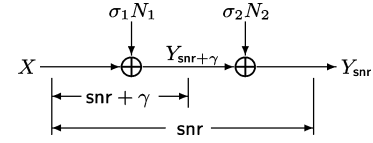


Fig. 2. Incremental Gaussian channel.

Proposition 3 enables translation of the MMSE at any given SNR to a conditional MMSE at a smaller SNR. This result was shown in [1] using the *incremental channel* technique, and has been instrumental in the proof of information-estimation relationships such as (8). Proposition 3 is also the key to the regularity properties and the derivatives of the MMSE presented in subsequent sections. A brief proof of the result is included here for completeness.

*Proof of Proposition 3:* Consider a cascade of two Gaussian channels as depicted in Fig. 2

$$Y_{\text{snr}+\gamma} = X + \sigma_1 N_1 \quad (23a)$$

$$Y_{\text{snr}} = Y_{\text{snr}+\gamma} + \sigma_2 N_2 \quad (23b)$$

where  $X$  is the input, and  $N_1, N_2 \sim \mathcal{N}(0, 1)$  are independent. A subscript is used to explicitly denote the SNR at which each observation is made. Let  $\sigma_1, \sigma_2 > 0$  satisfy  $\sigma_1^2 = 1/(\text{snr} + \gamma)$  and  $\sigma_1^2 + \sigma_2^2 = 1/\text{snr}$  so that the SNR of the first channel (23a) is  $\text{snr} + \gamma$  and that of the composite channel is  $\text{snr}$ . A linear combination of (23a) and (23b) yields

$$(\text{snr} + \gamma) Y_{\text{snr}+\gamma} = \text{snr} Y_{\text{snr}} + \gamma X + \sqrt{\gamma} W \quad (24)$$

where we have defined  $W = (\gamma \sigma_1 N_1 - \text{snr} \sigma_2 N_2)/\sqrt{\gamma}$ . Clearly, the input-output relationship defined by the incremental channel (23) is equivalently described by (24) paired with (23b). Due to mutual independence of  $(X, N_1, N_2)$ , it is easy to see that  $W \sim \mathcal{N}(0, 1)$  and  $(X, W, \sigma_1 N_1 + \sigma_2 N_2)$  are mutually independent. Thus  $W$  is independent of  $(X, Y_{\text{snr}})$  by (23). Based on the above observations, the relationship of  $X$  and  $Y_{\text{snr}+\gamma}$  conditioned on  $Y_{\text{snr}} = y$  is exactly the input-output relationship of a Gaussian channel with SNR equal to  $\gamma$  described by (24) with  $Y_{\text{snr}} = y$ . Because  $Y_{\text{snr}}$  is a physical degradation of  $Y_{\text{snr}+\gamma}$ , providing  $Y_{\text{snr}}$  as the side information does not change the overall MMSE, that is,  $\text{mmse}(X|Y_{\text{snr}+\gamma}) = \text{mmse}(X, \gamma|Y_{\text{snr}})$ , which proves (22). ■

### C. Bounds

The input to a Gaussian model with nonzero SNR can always be estimated with finite mean-square error based on the output, regardless of the input distribution. In fact,  $\tilde{X} = Y/\sqrt{\text{snr}}$  achieves mean-square error of  $1/\text{snr}$ , even if  $\mathbb{E}\{X\}$  does not exist. Moreover, the trivial zero estimate achieves mean-square error of  $\mathbb{E}\{X^2\}$ .

*Proposition 4:* For every input  $X$

$$\text{mmse}(X, \text{snr}) \leq \text{snr}^{-1} \quad (25)$$

and if the input variance  $\text{var}\{X\}$  is finite,

$$\text{mmse}(X, \text{snr}) \leq \min\{\text{var}\{X\}, \text{snr}^{-1}\}. \quad (26)$$

Proposition 4 can also be established using the fact that  $\text{snr} \cdot \text{mmse}(X, \text{snr}) = \text{mmse}(N|\sqrt{\text{snr}}X + N) \leq 1$ , which is simply because the estimation error of the input is proportional to the estimation error of the noise [7]

$$\sqrt{\text{snr}}(X - \mathbb{E}\{X|Y\}) = \mathbb{E}\{N|Y\} - N. \quad (27)$$

Using (27) and the moments of the Gaussian density, the higher moments of the estimation errors can also be bounded as shown in Appendix A:

*Proposition 5:* For every random variable  $X$  and  $\text{snr} > 0$

$$\mathbb{E}\{|X - \mathbb{E}\{X|\sqrt{\text{snr}}X + N\}|^n\} \leq \left(\frac{2}{\sqrt{\text{snr}}}\right)^n \sqrt{n!} \quad (28)$$

for  $n = 0, 1, \dots$ , where  $N \sim \mathcal{N}(0, 1)$  is independent of  $X$ .

#### D. Sub-Gaussian Distributions

In order to show some useful characteristics of the posterior input distribution, it is instructive to introduce the notion of *sub-Gaussianity*. A random variable  $X$  is called sub-Gaussian if the tail of its distribution is dominated by that of some Gaussian random variable, i.e.,

$$\mathbb{P}(|X| > \lambda) \leq Ce^{-c\lambda^2} \quad (29)$$

for some  $c, C > 0$  and all  $\lambda > 0$ . Sub-Gaussianity is equivalent to the growth of moments or moment generating functions not exceeding those of some Gaussian [16, Theorem 2].

*Lemma 1:* The following statements are equivalent:

- 1)  $X$  is sub-Gaussian;
- 2) there exists  $C > 0$  such that for every  $k = 1, 2, \dots$ ,

$$\mathbb{E}\{|X|^k\} \leq C^k \sqrt{k!}; \quad (30)$$

- 3) there exist  $c, C > 0$  such that for all  $t > 0$ ,

$$\mathbb{E}\{e^{tX}\} \leq Ce^{ct^2}. \quad (31)$$

Regardless of the prior input distribution, the posterior distribution of the input given the noisy observation through a Gaussian channel with nonzero SNR is always sub-Gaussian, and the posterior moments can be upper bounded. This is formalized in the following result proved in Appendix B:

*Proposition 6:* Let  $X_y$  be distributed according to  $P_{X|Y=y}$  where  $Y = aX + N$ ,  $N \sim \mathcal{N}(0, 1)$  is independent of  $X$ , and  $a \neq 0$ . Then  $X_y$  is sub-Gaussian for every  $y \in \mathbb{R}$ . Moreover

$$\mathbb{P}\{|X_y| \geq x\} \leq \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{x^2}{2}}}{h_0(y; a)} e^{-\frac{a^2 x^2}{4}} \quad (32)$$

and, for every  $n = 1, 2, \dots$

$$\mathbb{E}\{|X_y|^n\} \leq \frac{ne^{-\frac{y^2}{2}}}{h_0(y; a)} \left(\frac{\sqrt{2}}{|a|}\right)^n \sqrt{(n-1)!} \quad (33)$$

and

$$\mathbb{E}\{|X_y - \mathbb{E}\{X_y\}|^n\} \leq 2^n \mathbb{E}\{|X_y|^n\}. \quad (34)$$

### III. SMOOTHNESS AND ANALYTICITY

This section studies the regularity of the MMSE as a function of the SNR, where the input distribution is arbitrary but fixed. In particular, it is shown that  $\text{mmse}(X, \text{snr})$  is a smooth function of  $\text{snr}$  on  $(0, \infty)$  for every  $P_X$ . This conclusion clears the way towards calculating its derivatives in Section IV. Under certain technical conditions, the MMSE is also found to be real analytic in  $\text{snr}$ . This implies that the MMSE can be reconstructed from its local derivatives. As we shall see, the regularity of the MMSE at the point of zero SNR requires additional conditions.

#### A. Smoothness

*Proposition 7:* For every  $X$ ,  $\text{mmse}(X, \text{snr})$  is infinitely differentiable at every  $\text{snr} > 0$ . If  $\mathbb{E}\{X^{2k+2}\} < \infty$ , then  $\text{mmse}(X, \text{snr})$  is  $k$  right-differentiable at  $\text{snr} = 0$ . Consequently,  $\text{mmse}(X, \text{snr})$  is infinitely right differentiable at  $\text{snr} = 0$  if all moments of  $X$  are finite.

*Proof:* The proof is divided into two parts. In the first part we first establish the smoothness assuming that all input moments are finite, i.e.,  $\mathbb{E}\{X^k\} < \infty$  for all  $k = 1, 2, \dots$

For convenience, let  $Y = aX + N$  where  $a^2 = \text{snr}$ . For every  $i = 0, 1, \dots$ , denote

$$g_i(y; a) = \frac{\partial^i}{\partial a^i} \left(\frac{h_1^2}{h_0}\right)(y; a) \quad (35)$$

and

$$m_i(a) = \int_{-\infty}^{\infty} g_i(y; a) dy \quad (36)$$

where  $h_i$  is given by (14). By (17), we have

$$\text{mmse}(X, a^2) = \mathbb{E}\{X^2\} - m_0(a). \quad (37)$$

Let  $H_n$  denote the  $n$ -th Hermite polynomial [17, Section 5.5]

$$H_n(x) = \frac{(-1)^n}{\varphi(x)} \frac{d^n \varphi(x)}{dx^n} \quad (38)$$

$$= n! \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^k}{k!(n-2k)!} (2x)^{n-2k}. \quad (39)$$

Denote  $h_i^{(n)}(y; a) = \partial^n h_i(y; a) / \partial a^n$  throughout the paper. Then

$$\frac{h_i^{(n)}(y; a)}{h_0(y; a)} = \frac{\mathbb{E}\{X^{i+n} H_n(y - aX) \varphi(y - aX)\}}{h_0(y; a)} \quad (40)$$

$$= \mathbb{E}\{X^{i+n} H_n(N) | Y = y\} \quad (41)$$

where the derivative and expectation can be exchanged to obtain (40) because the product of any polynomial and the Gaussian density is bounded.

The following lemma is established in Appendix C:

*Lemma 2:* For every  $i = 0, 1, \dots$  and all  $w > v$ ,  $(y, a) \mapsto g_i(y; a)$  is integrable on  $\mathbb{R} \times [v, w]$ .

Using Lemma 2 and (36), we have

$$\int_v^w m_{i+1}(a) da = \int_v^w \int_{-\infty}^{\infty} g_{i+1}(y; a) dy da \quad (42)$$

$$= \int_{-\infty}^{\infty} g_i(y; w) - g_i(y; v) dy \quad (43)$$

$$= m_i(w) - m_i(v) \quad (44)$$

where (43) is due to (35) and Fubini's theorem. Therefore for every  $i \geq 0$ ,  $m_i$  is continuous. Hence for each  $a \in \mathbb{R}$ ,

$$\frac{dm_i(a)}{da} = m_{i+1}(a) \quad (45)$$

follows from the fundamental theorem of calculus [18, p. 97]. In view of (37), we have

$$\frac{d^i \text{mmse}(X, a^2)}{da^i} = -m_i(a). \quad (46)$$

This proves that  $a \mapsto \text{mmse}(X, a^2) \in C^\infty(\mathbb{R})$ , which implies that  $\text{mmse}(X, \text{snr})$  is infinitely differentiable in  $\text{snr}$  on  $(0, \infty)$ .

In the second part of this proof, we eliminate the requirement that all moments of the input exist by resorting to the incremental-SNR result, Proposition 3. Fix arbitrary  $\gamma > 0$  and let  $Y_\gamma = \sqrt{\gamma} X + N$ . For every  $u \in \mathbb{R}$ , let  $X_{u;\gamma} \sim P_{X|Y_\gamma=u}$ . By (17), (21) and Proposition 3, we have

$$\text{mmse}(X, \gamma + a^2) = \int \text{mmse}(X_{u;\gamma}, a^2) P_{Y_\gamma}(du) \quad (47)$$

$$= \mathbb{E}\{X^2\} - \tilde{m}_0(a) \quad (48)$$

where

$$h_i(y; a|u; \gamma) = \mathbb{E}\{X^i \varphi(y - aX) | Y_\gamma = u\} \quad (49)$$

$$g_i(y; a|u; \gamma) = \frac{\partial^i}{\partial a^i} \left( \frac{h_1^2}{h_0} \right) (y; a|u; \gamma) \quad (50)$$

and

$$\tilde{m}_i(a) = \int_{\mathbb{R}} \int_{\mathbb{R}} g_i(y; a|u; \gamma) dy h_0(u; \gamma) du \quad (51)$$

for  $i = 0, 1, \dots$ . By Proposition 5, for each  $u$ , all moments of  $X_{u;\gamma}$  are finite. Each  $\tilde{m}_i$  is a well-defined real-valued function on  $\mathbb{R}$ . Repeating the first part of this proof with  $h_i(y; a)$  replaced by  $h_i(y; a|u; \gamma)$ , we conclude that  $a \mapsto \text{mmse}(X, \gamma + a^2) \in C^\infty$  in  $a$  at least on  $|a| \geq \sqrt{\gamma}$ , which further implies that  $a \mapsto \text{mmse}(X, a^2) \in C^\infty(\mathbb{R} \setminus [-\sqrt{2\gamma}, \sqrt{2\gamma}])$  because  $a \mapsto \sqrt{a^2 - \gamma}$  has bounded derivatives of all order when  $|a| > \sqrt{2\gamma}$ . By the arbitrariness of  $\gamma$ , we have  $a \mapsto \text{mmse}(X, a^2) \in C^\infty(\mathbb{R} \setminus \{0\})$ , hence  $\text{mmse}(X, \cdot) \in C^\infty((0, \infty))$ .

Finally, we address the case of zero SNR. It follows from (41) and the independence of  $X$  and  $Y$  at zero SNR that

$$\frac{1}{h_0} \frac{\partial^n h_i}{\partial a^n} (y; 0) = \mathbb{E}\{X^{i+n}\} H_n(y). \quad (52)$$

Since  $\mathbb{E}\{|H_n(N)|\} \leq \sqrt{\mathbb{E}\{H_n^2(N)\}} = \sqrt{n!}$  is always finite, induction reveals that the  $n$ -th derivative of  $m_0$  with respect to  $a$  at 0 depends on the first  $n + 1$  moments of  $X$ . By Taylor's

theorem and the fact that  $m_0(a)$  is an even function of  $a$ , we have

$$m_0(a) = \sum_{j=0}^i \frac{m_{2j}(0)}{(2j)!} a^{2j} + O(|a|^{2i+2}) \quad (53)$$

in the vicinity of  $a = 0$ , which implies that  $m_0$  is  $i$  differentiable with respect to  $a^2$  at 0, with  $d^i m_0(0+)/d(a^2)^i = m_{2i}(0)$ , as long as  $\mathbb{E}\{X^{2i+2}\} < \infty$ . ■

### B. Real Analyticity

A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is said to be *real analytic* at  $x_0$  if it can be represented by a convergent power series in some neighborhood of  $x_0$ , i.e., there exists  $\delta > 0$  such that

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n \quad (54)$$

for every  $x \in (x_0 - \delta, x_0 + \delta)$ . One necessary and sufficient condition for  $f$  to be real analytic is that  $f$  can be extended to some open disk  $D(x_0, \delta) \triangleq \{z \in \mathbb{C} : |z - x_0| < \delta\}$  in the complex plane by the power series (54) [19]. If  $f$  is only defined on  $[0, \infty)$ , its analyticity at zero is defined via (54) for all  $x \in [0, \delta)$  in lieu of  $(-\delta, \delta)$ .

The analyticity of  $\text{snr} \mapsto \text{mmse}(X, \text{snr})$  defined on  $[0, \infty)$  is equivalent to the analyticity of  $a \mapsto \text{mmse}(X, a^2)$  on  $\mathbb{R}$ . This is immediate for  $\text{snr} > 0$  because  $a \mapsto a^2$  and  $\text{snr} \mapsto \sqrt{\text{snr}}$  are both analytic on  $(0, \infty)$  and composition of analytic functions is analytic [20]. To see the equivalence at  $\text{snr} = 0$ , note that the Taylor series expansion of  $\text{mmse}(X, a^2)$  at  $a = 0$  contains only even terms because it is an even function.

The following result provides sufficient conditions for the MMSE to be real analytic. The proof is relegated to Appendix D.

*Proposition 8:* As a function of  $a$ ,  $\text{mmse}(X, a^2)$  is real analytic at  $a_0 \in \mathbb{R}$  if either one of the following two sets of conditions holds:

- 1)  $X$  is sub-Gaussian, and there exist  $c > 0$  and  $r > 0$  such that for every  $y \in \mathbb{R}$

$$\inf_{z \in D(a_0, r)} |h_0(y; z)| > 0 \quad (55)$$

and

$$\liminf_{|y| \rightarrow \infty} \inf_{z \in D(a_0, r)} \frac{|h_0(y; z)|}{h_0(y; \text{Re}(z))} > c; \quad (56)$$

- 2)  $a_0 \neq 0$ , and there exist  $c > 0$ ,  $r > 0$  and  $\delta \in (0, a_0^2)$  such that for every  $y, u \in \mathbb{R}$

$$\inf_{z \in D(a_0, r)} |h_0(y; z|u, \delta)| > 0 \quad (57)$$

and

$$\liminf_{|y| \rightarrow \infty} \inf_{z \in D(a_0, r)} \frac{|h_0(y; z|u, \delta)|}{h_0(y; \text{Re}(z)|u, \delta)} > c. \quad (58)$$

Moreover, whenever  $\text{mmse}(X, a^2)$  is real analytic at  $a \in \mathbb{R}$ , the function  $\text{mmse}(X, \text{snr})$  is also analytic at  $\text{snr} = a^2$ .

Conditions (55) and (56) can be understood as follows. Recall that  $h_0(y; a)$  denotes the density of  $Y = aX + N$ . The function  $h_0(y; a)$  stays positive for all  $a \in \mathbb{R}$ , and decays no faster than the Gaussian density. However,  $h_0(y; a)$  may vanish for some  $a \in \mathbb{C}$ , so that the MMSE may not be extendable to the complex plane. Hence the purpose of (55) and (56) is to ensure that the imaginary part of  $a$  has limited impact on  $|h_0|$ .

As an example, consider the case where  $X$  is equiprobable on  $\{\pm 1\}$ . Then

$$h_0(y; a) = \varphi(y) \exp(-a^2/2) \cosh(ay). \quad (59)$$

Letting  $a = jt$  yields  $h_0(y; jt) = \varphi(\sqrt{y^2 - t^2}) \cos(ty)$ , which has infinitely many zeros. In fact, in this case the MMSE is given by (7), or in an equivalent form:

$$\text{mmse}(X, a^2) = 1 - \int_{-\infty}^{\infty} \varphi(y) \tanh(a^2 - ay) dy. \quad (60)$$

Then for any  $r > 0$ , there exists  $|a_0| < r$  and  $y_0 \in \mathbb{R}$ , such that  $a_0^2 - a_0 y_0 = j\frac{\pi}{2}$  and the integral in (60) diverges near  $y_0$ . Therefore,  $\text{mmse}(X, a^2)$  cannot be extended to any point on the imaginary axis, hence it is not real analytic at  $a = 0$ , and thus the corresponding MMSE is not real analytic at  $\text{snr} = 0$ . Nevertheless, when  $\text{Re}(a) \neq 0$ , condition (56) is satisfied. Hence  $\text{mmse}(X, a^2)$  is real analytic on the real line except zero, which can be shown from (60) directly. Similarly, for any finite-alphabet, exponential or Gaussian distributed  $X$ , (57) and (58) can be verified for all  $a \neq 0$ , hence in those cases the corresponding MMSE is real analytic at all positive SNR.

#### IV. DERIVATIVES

##### A. Derivatives of the MMSE

With the smoothness of the MMSE established in Proposition 7, its first few derivatives with respect to the SNR are explicitly calculated in this section. Consider the third-order Taylor series expansion of the MMSE around  $\text{snr} = 0^+$ :

$$\begin{aligned} \text{mmse}(X, \text{snr}) &= 1 - \text{snr} + [2 - (\mathbb{E}X^3)^2] \frac{\text{snr}^2}{2} \\ &\quad - [15 - 12(\mathbb{E}X^3)^2 - 6\mathbb{E}X^4 + (\mathbb{E}X^4)^2] \frac{\text{snr}^3}{6} + \mathcal{O}(\text{snr}^4) \end{aligned} \quad (61)$$

where  $X$  is assumed to have zero mean and unit variance. The first three derivatives of the MMSE at  $\text{snr} = 0^+$  are thus evident from (61). The technique for obtaining (61) is to expand (12) in terms of the small signal  $\sqrt{\text{snr}}X$ , evaluate  $h_i(y; \sqrt{\text{snr}})$  given by (14) at the vicinity of  $\text{snr} = 0$  using the moments of  $X$  (see [1, Eq. (90)]), and then calculate (16), where the integral over  $y$  can be evaluated as a Gaussian integral.

The preceding expansion (61) at zero SNR can be lifted to arbitrary SNR using Proposition 3. Finiteness of the input moments is not required for the expansion to be valid for  $\text{snr} > 0$

<sup>2</sup>The previous result for the expansion of  $\text{mmse}(\text{snr})$  around  $\text{snr}=0^+$ , given by (91) in [1] is mistaken in the coefficient corresponding to  $\text{snr}^2$ . The expansion of the mutual information given by (92) in [1] should also be corrected accordingly. The second derivative of the MMSE is mistaken in [21] and corrected in Proposition 9 in this paper. The function  $\text{mmse}(X, \text{snr})$  is not always convex in  $\text{snr}$  as claimed in [21], as illustrated using an example in Fig. 1.

because the conditional moments are always finite due to Proposition 5.

Let  $Y = \sqrt{\text{snr}}X + N$  and  $\text{snr} \geq 0$ . We define the following random variables (whose explicit dependence on  $Y$  and  $\text{snr}$  is omitted for convenience):

$$M_i = \mathbb{E} \left\{ (X - \mathbb{E}\{X|Y\})^i | Y \right\}, \quad i = 1, 2, \dots \quad (62)$$

The variables  $M_i$ , according to Proposition 5, are well-defined in case  $\text{snr} > 0$ , and reduce to the unconditional moments of  $X$  when  $\text{snr} = 0$ . Evidently,  $M_1 = 0$

$$M_2 = \text{var}\{X|\sqrt{\text{snr}}X + N\} \quad (63)$$

and

$$\mathbb{E}\{M_2\} = \text{mmse}(X, \text{snr}). \quad (64)$$

If the input distribution  $P_X$  is symmetric, then the distribution of  $M_i$  is also symmetric for all odd  $i$ .

The derivatives of the MMSE are found to be the expected value of polynomials of  $M_i$ , whose existence is guaranteed by Proposition 5.

*Proposition 9:* For every random variable  $X$  and every  $\text{snr} > 0$

$$\frac{d\text{mmse}(X, \text{snr})}{d\text{snr}} = -\mathbb{E}\{M_2^2\} \quad (65)$$

$$\frac{d^2\text{mmse}(X, \text{snr})}{d\text{snr}^2} = \mathbb{E}\{2M_2^3 - M_3^2\} \quad (66)$$

and

$$\begin{aligned} \frac{d^3\text{mmse}(X, \text{snr})}{d\text{snr}^3} \\ = \mathbb{E}\{6M_4M_2^2 - M_4^2 + 12M_3^2M_2 - 15M_2^4\}. \end{aligned} \quad (67)$$

The derivatives in (65) and (66) and (67) are also valid at  $\text{snr} = 0^+$  if  $X$  has finite fourth, sixth, and eighth moments, respectively.

We relegate the proof of Proposition 9 to Appendix E. It is easy to check that the derivatives found in Proposition 9 are consistent with the Taylor series expansion (61) at zero SNR.

In light of the proof of Proposition 7 (and (46)), the Taylor series expansion of the MMSE can be carried out to arbitrary orders, so that all derivatives of the MMSE can be obtained as the expectation of some polynomials of the conditional moments, although the resulting expressions become increasingly complicated.

Proposition 9 is easily verified in the special case of standard Gaussian input ( $X \sim \mathcal{N}(0, 1)$ ), where conditioned on  $Y = y$ , the input is Gaussian distributed

$$X \sim \mathcal{N}\left(\frac{\sqrt{\text{snr}}}{1 + \text{snr}}y, \frac{1}{1 + \text{snr}}\right). \quad (68)$$

In this case  $M_2 = (1 + \text{snr})^{-1}$ ,  $M_3 = 0$  and  $M_4 = 3(1 + \text{snr})^{-2}$  do not depend on  $Y$ , and (65) and (66) and (67) are straightforward.

### B. Derivatives of the Mutual Information

Based on Proposition 8 and 9, the following derivatives of the mutual information are extensions of the key information-estimation relationship (8).

*Corollary 1:* For every distribution  $P_X$  and  $\text{snr} > 0$ ,

$$\frac{d}{d\text{snr}} I(X; \sqrt{\text{snr}} X + N) = \frac{1}{2} \mathbb{E}\{M_2\} \quad (69)$$

$$\frac{d^2}{d\text{snr}^2} I(X; \sqrt{\text{snr}} X + N) = -\frac{1}{2} \mathbb{E}\{M_2^2\} \quad (70)$$

$$\frac{d^3}{d\text{snr}^3} I(X; \sqrt{\text{snr}} X + N) = \mathbb{E}\left\{M_2^3 - \frac{1}{2}M_3^2\right\} \quad (71)$$

$$\begin{aligned} & \frac{d^4}{d\text{snr}^4} I(X; \sqrt{\text{snr}} X + N) \\ &= \frac{1}{2} \mathbb{E}\left\{-M_4^2 + 6M_4M_2^2 + 2M_3^2M_2 - 15M_2^4\right\} \end{aligned} \quad (72)$$

as long as the corresponding expectation on the right hand side exists. In case one of the two set of conditions in Proposition 8 holds,  $\sqrt{\text{snr}} \mapsto I(\sqrt{\text{snr}} X + N; X)$  is also real analytic.

Corollary 1 is a generalization of previous results on the small SNR expansion of the mutual information such as in [22]. Note that (69) is exactly the original relationship of the mutual information and the MMSE given by (8) in light of (64).

### C. Derivatives of the Conditional MMSE

The derivatives in Proposition 9 can be generalized to the conditional MMSE defined in (20). The following is a straightforward extension of (65).

*Corollary 2:* For jointly distributed  $(X, U)$  and  $\text{snr} > 0$

$$\frac{d}{d\text{snr}} \text{mmse}(X, \text{snr}|U) = -\mathbb{E}\{M_2^2(U)\} \quad (73)$$

where for every  $u$  and  $i = 1, 2, \dots$

$$M_i(u) = \mathbb{E}\left\{[X_u - \mathbb{E}\{X_u|Y\}]^i \middle| Y = \sqrt{\text{snr}}X_u + N\right\} \quad (74)$$

is a random variable dependent on  $u$ .

## V. PROPERTIES OF THE MMSE FUNCTIONAL

For any fixed  $\text{snr}$ ,  $\text{mmse}(X, \text{snr})$  can be regarded as a functional of the input distribution  $P_X$ , whose various properties have been studied in [23] and [24]. Moreover, the MMSE curve,  $\{\text{mmse}(X, \text{snr}), \text{snr} \in [0, \infty)\}$ , can be regarded as a ‘‘transform’’ of the input distribution.

### A. Concavity in Input Distribution

*Proposition 10:* The functional  $\text{mmse}(X, \text{snr})$  is concave in  $P_X$  for every  $\text{snr} \geq 0$ .

*Proof:* Let  $X_0$  and  $X_1$  be arbitrary random variables. For any  $\alpha \in [0, 1]$ , let  $B$  be an independent Bernoulli variable with  $\mathbb{P}(B = 0) = 1 - \mathbb{P}(B = 1) = \alpha$ . The random variable  $X_B$  has distribution  $\alpha P_{X_0} + (1 - \alpha)P_{X_1}$ . As far as estimating  $X_B$  is concerned, revealing  $B$  allows one to choose either the optimal estimator for  $X_0$  or  $X_1$ , so that the average MMSE is improved. That is

$$\text{mmse}(X_B, \text{snr}) \geq \text{mmse}(X_B, \text{snr}|B) \quad (75)$$

$$\begin{aligned} &= \alpha \text{mmse}(X_0, \text{snr}) \\ &+ (1 - \alpha) \text{mmse}(X_1, \text{snr}) \end{aligned} \quad (76)$$

which proves the desired concavity.<sup>3</sup> ■

### B. Conditioning Reduces MMSE

As a fundamental measure of uncertainty, the MMSE decreases with additional side information available to the estimator. This is because an informed optimal estimator performs no worse than any uninformed estimator by simply discarding the side information.

*Proposition 11:* For any jointly distributed  $(X, U)$

$$\text{mmse}(X, \text{snr}|U) \leq \text{mmse}(X, \text{snr}) \quad (77)$$

for every  $\text{snr} \geq 0$ . For fixed  $\text{snr} > 0$ , equality holds in (77) if and only if  $X$  is independent of  $U$ .

*Proof:* The inequality (77) is straightforward by the concavity established in Proposition 10. In case the equality holds,  $P_{X|U=u}$  must be identical for  $P_U$ -almost every  $u$  due to strict concavity, that is,  $X$  and  $U$  are independent. ■

### C. Monotonicity

Propositions 10 and 11 are simple facts included here for completeness. They suggest that a mixture of random variables is harder to estimate than the individual variables on average. The same can be said about linear combinations:

*Proposition 12 ([12]):* For every  $\text{snr} \geq 0$  and  $\alpha \in [0, 2\pi]$

$$\begin{aligned} & \text{mmse}(\cos \alpha X_1 + \sin \alpha X_2, \text{snr}) \\ & \geq \cos^2 \alpha \text{mmse}(X_1, \text{snr}) + \sin^2 \alpha \text{mmse}(X_2, \text{snr}). \end{aligned} \quad (78)$$

A generalization of Proposition 12 concerns the MMSE of estimating a normalized sum of independent random variables:

*Proposition 13:* Let  $X_1, X_2, \dots$  be i.i.d. with finite variance. Let  $S_n = (X_1 + \dots + X_n)/\sqrt{n}$ . Then for every  $\text{snr} \geq 0$

$$\text{mmse}(S_{n+1}, \text{snr}) \geq \text{mmse}(S_n, \text{snr}). \quad (79)$$

Because of the central limit theorem, as  $n \rightarrow \infty$  the MMSE converges to the MMSE of estimating a Gaussian random variable with the same variance as that of  $X$ .

Proposition 13 is a simple corollary of the following general result in [8].

*Proposition 14 ([18]):* Let  $X_1, \dots, X_n$  be independent. For any  $\lambda_1, \dots, \lambda_n \geq 0$  which sum up to one and any  $\gamma \geq 0$

$$\text{mmse}\left(\sum_{i=1}^n X_i, \gamma\right) \geq \sum_{i=1}^n \lambda_i \text{mmse}\left(\frac{X_i}{\sqrt{(n-1)\lambda_i}}, \gamma\right) \quad (80)$$

where  $X_{\setminus i} = \sum_{j=1, j \neq i}^n X_j$ .

Setting  $\lambda_i = 1/n$  in (80) yields Proposition 13.

In view of the representation of the entropy or differential entropy using the MMSE in Section I, integrating both sides of (79) proves a monotonicity result of the entropy or differential entropy of  $S_n$  whichever is well-defined. More generally, [8] applies (11) and Proposition 14 to prove a more general result (see also [26]), originally given in [27].

<sup>3</sup>Strict concavity is shown in [23].

#### D. Gaussian Inputs are the Hardest to Estimate

Any non-Gaussian input achieves strictly smaller MMSE than Gaussian input of the same variance. This well-known result is illustrated in Fig. 1 and stated as follows.

*Proposition 15:* For every  $\text{snr} \geq 0$  and random variable  $X$  with variance no greater than  $\sigma^2$

$$\text{mmse}(X, \text{snr}) \leq \frac{\sigma^2}{1 + \text{snr}\sigma^2}. \quad (81)$$

The equality of (81) is achieved for all  $\text{snr}$  if and only if the distribution of  $X$  is Gaussian with variance  $\sigma^2$ .

*Proof:* The right hand side of (81) is achieved by the optimal linear estimator regardless of the input. The inequality (81) is evident due to the suboptimality of the linearity restriction on the estimator. The strict inequality for non-Gaussian inputs can be established using the uniqueness in the orthogonality principle. ■

Note that in case the variance of  $X$  is infinity, (81) reduces to (25).

#### E. The Single-Crossing Property

In view of Proposition 15 and the scaling property of the MMSE, at any given SNR, the MMSE of a non-Gaussian input is equal to the MMSE of some Gaussian input with reduced variance. The following result suggests that there is some additional simple ordering of the MMSEs due to Gaussian and non-Gaussian inputs.

*Proposition 16 (Single-Crossing Property):* For any given random variable  $X$  (with not necessarily unit variance), the curve of  $\text{mmse}(X, \gamma)$  crosses the curve of  $(1 + \gamma)^{-1}$ , which is the MMSE function of the standard Gaussian distribution, at most once on  $(0, \infty)$ . Precisely, define

$$f(\gamma) = (1 + \gamma)^{-1} - \text{mmse}(X, \gamma) \quad (82)$$

on  $[0, \infty)$ . Then:

- 1)  $f(\gamma)$  is strictly increasing at every  $\gamma$  with  $f(\gamma) < 0$ ;
- 2) if  $f(\text{snr}_0) = 0$ , then  $f(\gamma) \geq 0$  at every  $\gamma > \text{snr}_0$ ;
- 3)  $\lim_{\gamma \rightarrow \infty} f(\gamma) = 0$ .

Furthermore, all three statements hold if the term  $(1 + \gamma)^{-1}$  in (82) is replaced by  $\sigma^2/(1 + \sigma^2\gamma)$  with any  $\sigma$ , which is the MMSE function of a Gaussian variable with variance  $\sigma^2$ .

*Proof:* The last of the three statements,  $\lim_{\gamma \rightarrow \infty} f(\gamma) = 0$  always holds because of Proposition 4.

If  $\text{var}\{X\} \leq 1$ , then  $f(\gamma) \geq 0$  at all  $\gamma$  due to Proposition 15, so that the proposition holds. We suppose in the following  $\text{var}\{X\} > 1$ . The instance of the function  $f(\gamma)$  with  $X$  equally likely to be  $\pm\sqrt{2}$  is shown in Fig. 3. Evidently  $f(0) = 1 - \text{var}\{X\} < 0$ . Consider the derivative of the difference (82) at any  $\gamma$  with  $f(\gamma) < 0$ , which by Proposition 9, can be written as

$$f'(\gamma) = \text{E}\{M_2^2\} - (1 + \gamma)^{-2} \quad (83)$$

$$> \text{E}\{M_2^2\} - (\text{mmse}(X, \gamma))^2 \quad (84)$$

$$= \text{E}\{M_2^2\} - (\text{E}M_2)^2 \quad (85)$$

$$\geq 0 \quad (86)$$

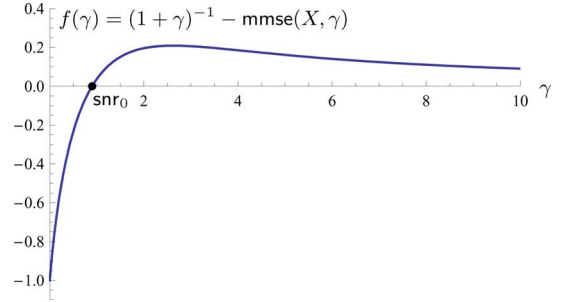


Fig. 3. Example of the difference between the MMSE for standard Gaussian input and that of a binary input equally likely to be  $\pm\sqrt{2}$  (i.e., with variance 2). The difference crosses the horizontal axis only once.

where (85) is due to (64), and (86) is due to Jensen's inequality. That is,  $f'(\gamma) > 0$  as long as  $f(\gamma) < 0$ , i.e., the function  $f$  can only be strictly increasing at every point it is strictly negative. This further implies that if  $f(\text{snr}_0) = 0$  for some  $\text{snr}_0$ , the function  $f$ , which is smooth, cannot dip to below zero for any  $\gamma > \text{snr}_0$ . Therefore, the function  $f$  has no more than one zero crossing.

For any  $\sigma$ , the above arguments can be repeated with  $\sigma^2\gamma$  treated as the SNR. It is straightforward to show that the proposition holds with the standard Gaussian MMSE replaced by the MMSE of a Gaussian variable with variance  $\sigma^2$ . ■

The single-crossing property can be generalized to the conditional MMSE defined in (20).<sup>4</sup>

*Proposition 17:* Let  $X$  and  $U$  be jointly distributed variables. All statements in Proposition 16 hold literally if the function  $f(\cdot)$  is replaced by

$$f(\gamma) = (1 + \gamma)^{-1} - \text{mmse}(X, \gamma|U). \quad (87)$$

*Proof:* For every  $u$ , let  $X_u$  denote a random variable indexed by  $u$  with distribution  $P_{X|U=u}$ . Define also a random variable for every  $u$

$$M(u, \gamma) = M_2(X_u, \gamma) \quad (88)$$

$$= \text{var}\{X_u|\sqrt{\text{snr}}X_u + N\} \quad (89)$$

where  $N \sim \mathcal{N}(0, 1)$ . Evidently,  $\text{E}\{M(u, \gamma)\} = \text{mmse}(X_u, \gamma)$  and hence

$$f(\gamma) = (1 + \gamma)^{-1} - \text{E}\{\text{E}\{M(U, \gamma)|U\}\} \quad (90)$$

$$= (1 + \gamma)^{-1} - \text{E}\{M(U, \gamma)\}. \quad (91)$$

In view of (91), for all  $\gamma$  such that  $f(\gamma) < 0$

$$f'(\gamma) = -\frac{1}{(1 + \gamma)^2} - \text{E}\left\{\frac{d}{d\gamma}M(U, \gamma)\right\} \quad (92)$$

$$= \text{E}\{M^2(U, \gamma)\} - (1 + \gamma)^{-2} \quad (93)$$

$$> \text{E}\{M^2(U, \gamma)\} - (\text{E}\{M(U, \gamma)\})^2 \quad (94)$$

$$\geq 0 \quad (95)$$

where (93) is by Proposition 9 and (95) is Jensen's inequality. The remaining argument is essentially the same as in the proof of Proposition 16. ■

<sup>4</sup>The single-crossing property has also been extended to the parallel degraded MIMO scenario [28].

*F. The High-SNR Asymptotics*

The asymptotics of  $\text{mmse}(X, \gamma)$  as  $\gamma \rightarrow \infty$  can be further characterized as follows. It is upper bounded by  $1/\gamma$  due to Propositions 4 and 15. Moreover, the MMSE can vanish faster than exponentially in  $\gamma$  with arbitrary rate, under, for instance, a sufficiently skewed binary input [29].<sup>5</sup> On the other hand, the decay of the MMSE of a non-Gaussian random variable need not be faster than the MMSE of a Gaussian variable. For example, let  $X = Z + \sqrt{\sigma_X^2 - 1} B$  where  $\sigma_X > 1$ ,  $Z \sim \mathcal{N}(0, 1)$  and the Bernoulli variable  $B$  are independent. Clearly,  $X$  is harder to estimate than  $Z$  but no harder than  $\sigma_X Z$ , i.e.,

$$\frac{1}{1 + \gamma} < \text{mmse}(X, \gamma) < \frac{\sigma_X^2}{1 + \sigma_X^2 \gamma} \tag{96}$$

where the difference between the upper and lower bounds is  $\mathcal{O}(\gamma^{-2})$ . As a consequence, the function  $f$  defined in (82) may not have any zero even if  $f(0) = 1 - \sigma_X^2 < 0$  and  $\lim_{\gamma \rightarrow \infty} f(\gamma) = 0$ . A thorough study of the high-SNR asymptotics of the MMSE can be found in [25], where the limit of the product  $\text{snr} \cdot \text{mmse}(X, \text{snr})$ , called the *MMSE dimension*, is determined for input distributions without singular components.

VI. APPLICATIONS TO CHANNEL CAPACITY

*A. Secrecy Capacity of the Gaussian Wiretap Channel*

This section makes use of the MMSE as an instrument to show that the secrecy capacity of the Gaussian wiretap channel is achieved by Gaussian inputs. The wiretap channel was introduced by Wyner in [30] in the context of discrete memoryless channels. Let  $X$  denote the input, and let  $Y$  and  $Z$  denote the outputs of the main channel and the wiretapper’s channel respectively. The problem is to find the rate at which reliable communication is possible through the main channel, while keeping the mutual information between the message and the wiretapper’s observation as small as possible. Assuming that the wiretapper sees a degraded output of the main channel, Wyner showed that secure communication can achieve any rate up to the secrecy capacity

$$C_s = \max_X \{I(X; Y) - I(X; Z)\} \tag{97}$$

where the supremum is taken over all admissible choices of the input distribution. Wyner also derived the achievable rate-equivocation region.

Consider the Gaussian wiretap channel studied in [31]

$$Y = \sqrt{\text{snr}_1} X + N_1 \tag{98a}$$

$$Z = \sqrt{\text{snr}_2} X + N_2 \tag{98b}$$

where  $\text{snr}_1 \geq \text{snr}_2$  and  $N_1, N_2 \sim \mathcal{N}(0, 1)$  are independent. Let the energy of every codeword  $(x_1, \dots, x_n)$  be constrained by  $\frac{1}{n} \sum_{i=1}^n x_i^2 \leq 1$ . [31] showed that the optimal input which achieves the supremum in (97) is standard Gaussian and that the secrecy capacity is

$$C_s = \frac{1}{2} \log \left( \frac{1 + \text{snr}_1}{1 + \text{snr}_2} \right). \tag{99}$$

<sup>5</sup>In case the input is equally likely to be  $\pm 1$ , the MMSE decays as  $e^{-\frac{1}{2}\text{snr}}$ , not  $e^{-2\text{snr}}$  as stated in [1], [29].

In contrast to [31] which appeals to Shannon’s EPI, we proceed to give a simple proof of the same result using (9), which enables us to write for any  $X$ :

$$I(X; Y) - I(X; Z) = \frac{1}{2} \int_{\text{snr}_2}^{\text{snr}_1} \text{mmse}(X, \gamma) d\gamma. \tag{100}$$

Under the constraint  $E\{X^2\} \leq 1$ , the maximum of (100) over  $X$  is achieved by standard Gaussian input because it maximizes not just the integral but the MMSE for every SNR under the power constraint. Plugging  $\text{mmse}(X, \gamma) = (1 + \gamma)^{-1}$  into (100) yields the secrecy capacity given in (99). In fact the whole rate-equivocation region can be obtained using the same techniques. The MIMO wiretap channel can be treated similarly [11].

*B. The Gaussian Broadcast Channel*

In this section, we use the single-crossing property to show that Gaussian inputs achieve the capacity region of scalar Gaussian broadcast channels. Consider a degraded Gaussian broadcast channel also described by the same model (98). The formulation of the Gaussian broadcast channel is statistically identical to that of the Gaussian wiretap channel, except for a different goal: A convex combination of the rates between the sender and both receivers is to be maximized, rather than minimizing the rate between the sender and the (degraded) wiretapper. The capacity region of degraded broadcast channels under a unit input power constraint is given by [32]:

$$\bigcup_{P_{UX}: E\{X^2\} \leq 1} \left\{ \begin{array}{l} R_1 \leq I(X; Y|U) \\ R_2 \leq I(U; Z) \end{array} \right\} \tag{101}$$

where  $U$  is an auxiliary random variable with  $U - X - (Y, Z)$  being a Markov chain. It has long been recognized that Gaussian  $P_{UX}$  with standard Gaussian marginals and correlation coefficient  $E\{UX\} = \sqrt{1 - \alpha}$  achieves the capacity [33]. The resulting capacity region is

$$\bigcup_{\alpha \in [0, 1]} \left\{ \begin{array}{l} R_1 \leq \frac{1}{2} \log(1 + \alpha \text{snr}_1) \\ R_2 \leq \frac{1}{2} \log \left( \frac{1 + \text{snr}_2}{1 + \alpha \text{snr}_2} \right) \end{array} \right\}. \tag{102}$$

The conventional proof of the optimality of Gaussian inputs relies on the EPI in conjunction with Fano’s inequality [34]. The converse can also be proved directly from (101) using only the EPI [35], [36]. In the following we show a simple alternative proof using the single-crossing property of MMSE.

Due to the power constraint on  $X$ , there must exist  $\alpha \in [0, 1]$  (dependent on the distribution of  $X$ ) such that

$$I(X; Z|U) = \frac{1}{2} \log(1 + \alpha \text{snr}_2) \tag{103}$$

$$= \frac{1}{2} \int_0^{\text{snr}_2} \frac{\alpha}{\alpha\gamma + 1} d\gamma. \tag{104}$$

By the chain rule

$$I(U; Z) = I(U, X; Z) - I(X; Z|U) \tag{105}$$

$$= I(X; Z) - I(X; Z|U). \tag{106}$$

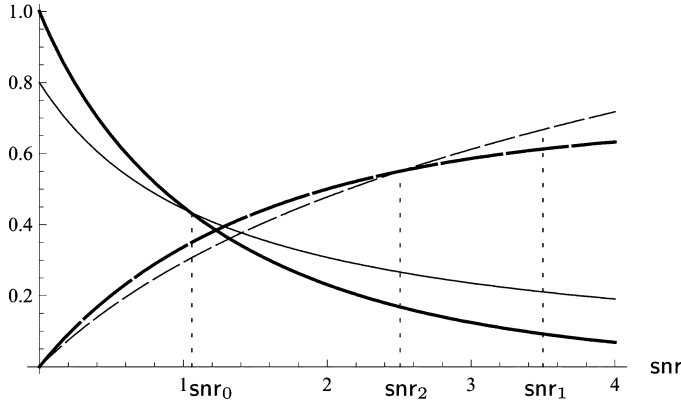


Fig. 4. Thin curves show the MMSE (solid line) and mutual information (dashed line) of a Gaussian input. The thick curves show the MMSE (solid) and mutual information (dashed) of binary input. The two mutual informations are identical at  $\text{snr}_2$ , which must be greater than  $\text{snr}_0$  where the two MMSE curves cross.

By (101) and (103), the desired bound on  $R_2$  is established

$$R_2 \leq \frac{1}{2} \log(1 + \text{snr}_2) - \frac{1}{2} \log(1 + \alpha, \text{snr}_2) \quad (107)$$

$$= \frac{1}{2} \log \left( \frac{1 + \text{snr}_2}{1 + \alpha \text{snr}_2} \right). \quad (108)$$

It remains to establish the desired bound for  $R_1$ . The idea is illustrated in Fig. 4, where crossing of the MMSE curves imply some ordering of the corresponding mutual informations. Note that

$$I(X; Z|U = u) = \frac{1}{2} \int_0^{\text{snr}_2} \text{mmse}(X_u, \gamma) d\gamma \quad (109)$$

and hence

$$I(X; Z|U) = \frac{1}{2} \int_0^{\text{snr}_2} \mathbb{E}\{\text{mmse}(X_U, \gamma|U)\} d\gamma. \quad (110)$$

Comparing (110) with (104), there must exist  $0 \leq \text{snr}_0 \leq \text{snr}_2$  such that

$$\mathbb{E}\{\text{mmse}(X_U, \text{snr}_0|U)\} = \frac{\alpha}{\alpha \text{snr}_0 + 1}. \quad (111)$$

By Proposition 17, this implies that for all  $\gamma \geq \text{snr}_2 \geq \text{snr}_0$

$$\mathbb{E}\{\text{mmse}(X_U, \gamma|U)\} \leq \frac{\alpha}{\alpha\gamma + 1}. \quad (112)$$

Consequently

$$R_1 \leq I(X; Y|U) \quad (113)$$

$$= \frac{1}{2} \int_0^{\text{snr}_1} \mathbb{E}\{\text{mmse}(X_U, \gamma|U)\} d\gamma \quad (114)$$

$$= \frac{1}{2} \left( \int_0^{\text{snr}_2} + \int_{\text{snr}_2}^{\text{snr}_1} \right) \mathbb{E}\{\text{mmse}(X_U, \gamma|U)\} d\gamma \quad (115)$$

$$\leq \frac{1}{2} \log(1 + \alpha \text{snr}_2) + \frac{1}{2} \int_{\text{snr}_2}^{\text{snr}_1} \frac{\alpha}{\alpha\gamma + 1} d\gamma \quad (116)$$

$$= \frac{1}{2} \log(1 + \alpha \text{snr}_1) \quad (117)$$

where the inequality (116) is due to (103), (110), and (112).

### C. Proof of a Special Case of the EPI

A simple proof of the general EPI using the properties of the MMSE was given in [2]. As another simple application of the single-crossing property, we show in this subsection that

$$e^{2h(X+Z)} \geq e^{2h(X)} + 2\pi e\sigma_Z^2 \quad (118)$$

for any independent  $X$  and  $Z$  as long as the differential entropy of  $X$  is not equal to  $-\infty$  and  $Z$  is Gaussian with variance  $\sigma_Z^2$ . This is in fact a special case of Shannon's entropy power inequality. Let  $W \sim \mathcal{N}(0, 1)$  and  $a^2$  be the ratio of the entropy powers of  $X$  and  $W$ , so that

$$h(X) = h(aW) = \frac{1}{2} \log(2\pi e a^2). \quad (119)$$

Consider the difference

$$\begin{aligned} & h(\sqrt{\text{snr}}X + N) - h(\sqrt{\text{snr}}aW + N) \\ &= \frac{1}{2} \int_0^{\text{snr}} \text{mmse}(X, \gamma) - \text{mmse}(aW, \gamma) d\gamma \end{aligned} \quad (120)$$

where  $N$  is standard Gaussian independent of  $X$  and  $W$ . In the limit of  $\text{snr} \rightarrow \infty$ , the left hand side of (120) vanishes due to (119). By Proposition 16, the integrand in (120) as a function of  $\gamma$  crosses zero only once, which implies that the integrand is initially positive, and then becomes negative after the zero crossing (cf. Fig. 3). Consequently, the integral (120) is positive and increasing for small  $\text{snr}$ , and starts to monotonically decrease after the zero crossing. If the integral crosses zero it will not be able to cross zero again. Hence, the integral in (120) must remain positive for all  $\text{snr}$  (otherwise it has to be strictly negative as  $\text{snr} \rightarrow \infty$ ). Therefore

$$\exp(2h(\sqrt{\text{snr}}X + N)) \geq \exp(h(\sqrt{\text{snr}}W + N)) \quad (121)$$

$$= 2\pi e(a^2 \text{snr} + 1) \quad (122)$$

$$= \exp(2h(\sqrt{\text{snr}}X)) + 2\pi e \quad (123)$$

which is equivalent to (118) by choosing  $\text{snr} = \sigma_Z^{-2}$  and appropriate scaling.

The preceding proof technique also applies to conditional EPI, which concerns  $h(X|U)$  and  $h(X + Z|U)$ , where  $Z$  is Gaussian independent of  $U$ . The conditional EPI can be used to establish the capacity region of the scalar broadcast channel in [34], [35].

## VII. CONCLUDING REMARKS

This paper has established a number of basic properties of the MMSE in Gaussian noise as a transform of the input distribution and function of the SNR. Because of the intimate relationship MMSE has with information measures, its properties find direct use in a number of problems in information theory. In a very recent work [37], the smoothness of the MMSE has proven to be instrumental in the analysis of Gaussian channel capacity with finite constellation constraint.

The MMSE can be viewed as a transform from the input distribution to a function of the SNR:  $P_X \mapsto \{\text{mmse}(P_X, \gamma), \gamma \in [0, \infty)\}$ . An interesting question remains to be answered: Is this transform one-to-one? We have the following conjecture:

*Conjecture 1:* For any zero-mean random variables  $X$  and  $Z$ ,  $\text{mmse}(X, \text{snr}) \equiv \text{mmse}(Z, \text{snr})$  for all  $\text{snr} \in [0, \infty)$  if and only if  $X$  is identically distributed as either  $Z$  or  $-Z$ .

There is a strong relationship between the real analyticity of MMSE and Conjecture 1. In particular, MMSE being real-analytic at zero SNR for all input and MMSE being an injective transform on the set of all random variables (with shift and reflection identified) cannot both hold. This is because given the real analyticity at zero SNR, MMSE can be extended to an open disk  $D$  centered at zero via the power series expansion, where the coefficients depend only on the moments of  $X$ . Since the solution to the Hamburger moment problem is not unique in general, there may exist different  $X$  and  $X'$  with the same moments, and hence their MMSE function coincide in  $D$  [38]. By the identity theorem of analytic functions, they coincide everywhere, hence on the real line. Nonetheless, if one is restricted to the class of sub-Gaussian random variables, the moments determine the distribution uniquely by Carleman's condition [39].

#### APPENDIX A PROOF OF PROPOSITION 5

*Proof:* Let  $Y = \sqrt{\text{snr}}X + N$  with  $\text{snr} > 0$ . Using (27) and then Jensen's inequality twice, we have

$$\begin{aligned} & \mathbb{E}\{|X - \mathbb{E}\{X|Y\}|^n\} \\ &= \text{snr}^{-\frac{n}{2}} 2^n \mathbb{E}\{2^{-n} |\mathbb{E}\{N|Y\} - N|^n\} \end{aligned} \quad (124)$$

$$\leq \text{snr}^{-\frac{n}{2}} 2^{n-1} \mathbb{E}\{|\mathbb{E}\{N|Y\}|^n + |N|^n\} \quad (125)$$

$$\leq \text{snr}^{-\frac{n}{2}} 2^n \mathbb{E}\{|N|^n\} \quad (126)$$

which leads to (28) because

$$\mathbb{E}\{|N|^n\} = \sqrt{\frac{2^n}{\pi}} \Gamma\left(\frac{n+1}{2}\right) \quad (127)$$

$$\leq \sqrt{n!}. \quad (128)$$

#### APPENDIX B PROOF OF PROPOSITION 6

*Proof:* We use the characterization by moment generating function in Lemma 1:

$$\mathbb{E}\{e^{tX_y}\} = \frac{1}{h_0(y; a)} \mathbb{E}\{e^{tX} \varphi(y - aX)\} \quad (129)$$

$$= \frac{\varphi(y)}{h_0(y; a)} \mathbb{E} \exp\left((t + ay)X - \frac{a^2 X^2}{2}\right) \quad (130)$$

$$\leq \frac{\varphi(y)}{h_0(y; a)} \exp\left(\frac{(t + ay)^2}{2a^2}\right) \quad (131)$$

$$\leq \frac{\varphi(y)}{h_0(y; a)} \exp\left(\frac{t^2}{a^2} + y^2\right) \quad (132)$$

where (131) and (132) are due to elementary inequalities. Using Chernoff's bound and (132), we have

$$\mathbb{P}\{X_y \geq x\} \leq \mathbb{E}\left\{e^{t(X_y - x)}\right\} \quad (133)$$

$$\leq \frac{\varphi(y)e^{y^2}}{h_0(y; a)} \exp\left(\frac{t^2}{a^2} - tx\right) \quad (134)$$

for all  $x, t > 0$ . Choosing  $t = \frac{a^2 x}{2}$  yields

$$\mathbb{P}\{X_y \geq x\} \leq \frac{e^{\frac{y^2}{2}}}{h_0(y; a)} \varphi\left(\frac{ax}{\sqrt{2}}\right). \quad (135)$$

Similarly,  $\mathbb{P}\{X_y \leq -x\}$  admits the same bound as above, and (32) follows from the union bound. Then, using an alternative formula for moments [40, p. 319]

$$\mathbb{E}\{|X_y|^n\} = n \int_0^\infty x^{n-1} \mathbb{P}\{|X_y| \geq x\} dx \quad (136)$$

$$\leq \frac{2ne^{\frac{y^2}{2}}}{h_0(y; a)} \int_0^\infty x^{n-1} \varphi\left(\frac{ax}{\sqrt{2}}\right) dx \quad (137)$$

$$\leq \frac{ne^{\frac{y^2}{2}}}{h_0(y; a)} \left(\frac{\sqrt{2}}{|a|}\right)^n \mathbb{E}\{|N|^{n-1}\} \quad (138)$$

where  $N \sim \mathcal{N}(0, 1)$  and (137) is due to (32). The inequality (33) is thus established by also noting (128).

Conditioned on  $Y = y$ , using similar techniques leading to (126), we have

$$\begin{aligned} & \mathbb{E}\{|X - \mathbb{E}\{X|Y\}|^n | Y = y\} \\ & \leq 2^{n-1} (\mathbb{E}\{|X|^n | Y = y\} + |\mathbb{E}\{X|Y = y\}|^n) \end{aligned} \quad (139)$$

$$\leq 2^n \mathbb{E}\{|X|^n | Y = y\} \quad (140)$$

which is (34).  $\blacksquare$

#### APPENDIX C PROOF OF LEMMA 2

We first make the following observation:

*Lemma 3:* For every  $i = 0, 1, \dots$ , the function  $g_i$  is a finite weighted sum of functions of the following form:

$$\frac{1}{h_0^{k-1}} \prod_{j=1}^k h_{n_j}^{(m_j)} \quad (141)$$

where  $n_j, m_j, k = 0, 1, \dots$

*Proof:* We proceed by induction on  $i$ : The lemma holds for  $i = 0$  by definition of  $g_0$ . Assume the induction hypothesis holds for  $i$ . Then

$$\begin{aligned} \frac{\partial}{\partial a} \left( \frac{1}{h_0^{k-1}} \prod_{j=1}^k h_{n_j}^{(m_j)} \right) &= \frac{-(k-1)}{h_0^k} h_0^{(1)} \prod_{j=1}^k h_{n_j}^{(m_j)} \\ &+ \frac{1}{h_0^{k-1}} \sum_{l=1}^k h_{n_l}^{(m_l+1)} \prod_{j \neq l} h_{n_j}^{(m_j)} \end{aligned} \quad (142)$$

which proves the lemma.  $\blacksquare$

To show the absolutely integrability of  $g_i$ , it suffices to show the function in (141) is integrable

$$\int_{-\infty}^{\infty} \left| \frac{1}{h_0^{k-1}(y; a)} \prod_{j=1}^k \frac{\partial^{m_j} h_{n_j}(y; a)}{\partial a^{m_j}} \right| dy$$

$$= \mathbb{E} \left\{ \prod_{j=1}^k \left| \frac{1}{h_0(Y; a)} \frac{\partial^{m_j} h_{n_j}(Y; a)}{\partial a^{m_j}} \right| \right\} \quad (143)$$

$$= \mathbb{E} \left\{ \prod_{j=1}^k |\mathbb{E}\{X^{n_j+m_j} H_{m_j}(Y-aX) | Y\}| \right\} \quad (144)$$

$$\leq \prod_{j=1}^k \left[ \mathbb{E} \left\{ (\mathbb{E}\{|X^{n_j+m_j} H_{m_j}(Y-aX)| | Y\})^k \right\} \right]^{\frac{1}{k}} \quad (145)$$

$$\leq \prod_{j=1}^k \left[ \mathbb{E} \left\{ |X|^{k(n_j+m_j)} \right\} \mathbb{E} \left\{ |H_{m_j}(N)|^k \right\} \right]^{\frac{1}{k}} \quad (146)$$

$$< \infty \quad (147)$$

where (144) is by (41), (145) is by the generalized Hölder inequality [41, p. 46] and (146) is due to Jensen's inequality and the independence of  $X$  and  $N = Y - aX$ .

#### APPENDIX D

##### PROOF OF PROPOSITION 8 ON THE ANALYTICITY

We first assume that  $X$  is sub-Gaussian.

Note that  $\varphi$  is real analytic everywhere with infinite radius of convergence, because  $\varphi^{(n)}(y) = (-1)^n H_n(y)\varphi(y)$  and Hermite polynomials admit the following bound [42, p. 997]:

$$|H_n(y)| \leq \kappa \sqrt{n!} e^{\frac{y^2}{4}} \quad (148)$$

where  $\kappa$  is an absolute constant. Hence

$$\lim_{n \rightarrow \infty} \left| \frac{\varphi^{(n)}(y)}{n!} \right|^{\frac{1}{n}} = 0 \quad (149)$$

and the radius of convergence is infinite at all  $y$ . Then

$$\varphi(y - a'x) = \sum_{n=0}^{\infty} \frac{H_n(y - ax)\varphi(y - ax)x^n}{n!} (a' - a)^n \quad (150)$$

holds for all  $a, x \in \mathbb{R}$ . By Lemma 1, there exists  $c > 0$ , such that  $\mathbb{E}\{|X|^n\} \leq c^n \sqrt{n!}$  for all  $n = 1, 2, \dots$ . By (148), it is easy to see that  $|H_n(y)\varphi(y)| \leq \kappa \sqrt{n!}$  for every  $y$ . Hence

$$\mathbb{E}\{|H_n(y - aX)\varphi(y - aX)X^n\}| \leq \kappa c^n n!. \quad (151)$$

Thus, for every  $|a' - a| < R \triangleq \frac{1}{c}$ ,

$$\sum_{n=0}^{\infty} \frac{|a' - a|^n}{n!} \mathbb{E}\{|(H_n \cdot \varphi)(y - aX)X^n\}| < \infty. \quad (152)$$

Applying Fubini's theorem to (150) yields

$$h_0(y; a') = \sum_{n=0}^{\infty} \frac{(a' - a)^n}{n!} \mathbb{E}\{(H_n \cdot \varphi)(y - aX)X^n\}. \quad (153)$$

Therefore,  $h_0(y; a)$  is real analytic at  $a$  and the radius of convergence is lower bounded by  $R$  independent of  $y$ . Similar conclusions also apply to  $h_1(y; a)$  and

$$h_1(y; a') = \sum_{n=0}^{\infty} \frac{(a' - a)^n}{n!} \mathbb{E}\{(H_n \cdot \varphi)(y - aX)X^{n+1}\} \quad (154)$$

holds for all  $y \in \mathbb{R}$  and all  $|a' - a| < R$ . Extend  $h_0(y; a)$  and  $h_1(y; a)$  to the complex disk  $D(a, R)$  by the power series (153) and (154). By (55), there exists  $0 < r < R/2$ , such that  $h_0(y; z)$  does not vanish on the disk  $D(a, r)$ . By [20, Proposition 1.1.5], for all  $y \in \mathbb{R}$

$$g_0(y; z) = \frac{h_1^2(y; z)}{h_0(y; z)} \quad (155)$$

is analytic in  $z$  on  $D(a, r)$ .

By assumption (56), there exist  $B, c > 0$ , such that

$$|h_0(y; z)| \geq c h_0(y; \operatorname{Re}(z)) \quad (156)$$

for all  $z \in D(a, r)$  and all  $|y| \geq B$ . Define

$$m_0^B(z) = \int_{-B}^B g_0(y; z) dy. \quad (157)$$

Since  $(y, z) \mapsto g_0(y; z)$  is continuous, for every closed curve  $\gamma$  in  $D(a, r)$ , we have  $\oint_{\gamma} \int_{-B}^B |g_0(y; z)| dy dz < \infty$ . By Fubini's theorem

$$\oint_{\gamma} \int_{-B}^B g_0(y; a) dy dz = \int_{-B}^B \oint_{\gamma} g_0(y; a) dz dy = 0 \quad (158)$$

where the last equality follows from the analyticity of  $g_0(y; \cdot)$ . By Morera's theorem [43, Theorem 3.1.4],  $m_0^B$  is analytic on  $D(a, r)$ .

Next, we show that as  $B \rightarrow \infty$ ,  $m_0^B$  tends to  $m_0$  uniformly in  $z \in D(a, r)$ . Since uniform limit of analytic functions is analytic [44, p. 156], we obtain the analyticity of  $m_0$ . To this end, it is sufficient to show that  $\{|g_0(\cdot; z)| : z \in D(a, r)\}$  is uniformly integrable. Let  $z = s + it$ . Then

$$|h_1(y; z)| = |\mathbb{E}\{X\varphi(y - zX)\}| \quad (159)$$

$$\leq \mathbb{E}\{|X|\varphi(y - zX)\}| \quad (160)$$

$$= \mathbb{E}\left\{|X|\varphi(y - sX)e^{\frac{1}{2}t^2 X^2}\right\}. \quad (161)$$

Therefore, for all  $z \in D(a, r)$

$$\int_{\mathbb{R}} |g_0(y; z)|^2 dy - \int_{-K}^K |g_0(y; z)|^2 dy$$

$$\leq \frac{1}{c^2} \int_{\mathbb{R}} \left| \frac{h_1(y; z)}{h_0(y; s)} \right|^4 h_0^2(y; s) dy \quad (162)$$

$$\leq \frac{1}{c^2} \int_{\mathbb{R}} \left| \frac{\mathbb{E} \left\{ |X| e^{\frac{1}{2} t^2 X^2} \varphi(y - sX) \right\}}{h_0(y; s)} \right|^4 h_0(y; s) dy \quad (163)$$

$$\leq \frac{1}{c^2} \mathbb{E} \left\{ \left( \mathbb{E} \left\{ |X| e^{\frac{t^2 X^2}{2}} \middle| Y_{s^2} \right\} \right)^4 \right\} \quad (164)$$

$$\leq \frac{1}{c^2} \mathbb{E} \{ X^4 e^{2r^2 X^2} \} \quad (165)$$

where (162) is by (56), (163) is by  $|h_0(y; s)| \leq 1$ , (164) is by (161), and (165) is due to Jensen's inequality and  $|t| \leq r$ . Since  $X$  is sub-Gaussian satisfying (29) and  $r < R/2 = 1/(2c)$

$$\mathbb{E} \{ X^4 e^{2r^2 X^2} \} \leq \sum_{n=0}^{\infty} \frac{(2r^2)^n}{n!} \mathbb{E} \{ |X|^{2n+4} \} \quad (166)$$

$$\leq \sum_{n=0}^{\infty} \frac{(2r^2)^n}{n!} \sqrt{(2n+4)!} c^{2n+4} \quad (167)$$

$$\leq 4c^4 \sum_{n=0}^{\infty} (n^2 + 3n + 2)(2rc)^{2n} \quad (168)$$

$$< \infty. \quad (169)$$

Therefore,  $\{|g_0(\cdot; z)| : z \in D(a, r)\}$  is  $L^2$ -bounded, hence uniformly integrable. We have thus shown that  $m_0(a)$ , i.e., the MMSE, is real analytic in  $a$  on  $\mathbb{R}$ .

We next consider positive SNR and drop the assumption of sub-Gaussianity of  $X$ . Let  $a_0 > 0$  and fix  $\delta$  with  $0 < \sqrt{\delta} < a_0/2$ . We use the incremental-SNR representation for MMSE in (48). Define  $\bar{X}_u$  to be distributed according to  $X - \mathbb{E}\{X|Y_\delta = u\}$  conditioned on  $Y_\delta = u$  and recall the definition of and  $h_i(y; a|u; \delta)$  in (49). In view of Proposition 6,  $\bar{X}_u$  is sub-Gaussian whose growth of moments only depends on  $\delta$  (the bounds depend on  $u$  but the terms varying with  $n$  do not depend on  $u$ ). Repeating the arguments from (148) to (154) with  $c = \sqrt{2/\delta}$ , we conclude that  $h_0(y; a|u; \delta)$  and  $h_1(y; a|u; \delta)$  are analytic in  $a$  and the radius of convergence is lower bounded by  $R = \sqrt{\delta}/2$ , independent of  $u$  and  $y$ .

Let  $r < \sqrt{\delta}/4$ . The remaining argument follows as in the first part of this proof, except that (162)–(169) are replaced by the following estimates: Let  $\tau = t^2/2$ , then

$$\mathbb{E} \left\{ \left( \mathbb{E} \left\{ |\bar{X}| e^{\tau X^2} \middle| Y_{s^2}, Y_\delta \right\} \right)^4 \right\} \leq \mathbb{E} \left\{ \left( \mathbb{E} \left\{ |\bar{X}| e^{\tau X^2} \middle| Y_\delta \right\} \right)^4 \right\} \quad (170)$$

$$= \mathbb{E} \left\{ \prod_{i=1}^4 \sum_{n_i=0}^{\infty} \frac{\tau^{n_i}}{n_i!} \mathbb{E} \left\{ |\bar{X}|^{2n_i+1} \middle| Y_\delta \right\} \right\} \quad (171)$$

$$\leq \sum_{n_1, n_2, n_3, n_4=0}^{\infty} \left( \frac{8\tau}{\delta} \right)^{\sum_i n_i+1} \binom{\sum_i n_i}{n_1, n_2, n_3, n_4} \quad (172)$$

$$\leq \left( \frac{8\tau}{\delta} \right) \sum_{n_1, n_2, n_3, n_4=0}^{\infty} \left( \frac{32\tau}{\delta^2} \right)^{\sum_i n_i} \quad (173)$$

$$= \left( \frac{8\tau}{\delta} \right) \left( \sum_{n=0}^{\infty} \left( \frac{32\tau}{\delta^2} \right)^n \right)^4 \quad (174)$$

$$< \infty \quad (175)$$

where (170) is by Jensen's inequality, (171) is by Fubini's theorem, (175) is because  $\tau \leq r^2/2 < \delta^2/32$ , and (172) is by Lemma 4, to be established next.

Let  $M_i$  be defined as in Section IV-A. The following lemma bounds the expectation of products of  $|M_i|$ :

*Lemma 4:* For any  $\text{snr} > 0$ ,  $k, i_j, n_j \in \mathbb{N}$

$$\mathbb{E} \left\{ \prod_{j=1}^k |M_{i_j}|^{n_j} \right\} \leq \text{snr}^{-\frac{\alpha}{2}} 2^n \sqrt{n!} \quad (176)$$

where  $n = \sum_{j=1}^k i_j n_j$ .

*Proof:* In view of Proposition 5, it suffices to establish

$$\mathbb{E} \left\{ \prod_{j=1}^k |M_{i_j}|^{n_j} \right\} = \mathbb{E} \left\{ \prod_{j=1}^k \prod_{l=1}^{n_j} |M_{i_j}| \right\} \quad (177)$$

$$\leq \prod_{j=1}^k \prod_{l=1}^{n_j} \left( \mathbb{E} \left\{ |M_{i_j}|^{\frac{n_j}{l}} \right\} \right)^{\frac{i_j}{n}} \quad (178)$$

$$\leq \prod_{j=1}^k \prod_{l=1}^{n_j} \left( \mathbb{E} \{ |X - \mathbb{E}\{X|Y\}|^n \} \right)^{\frac{i_j}{n}} \quad (179)$$

$$= \mathbb{E} \{ |X - \mathbb{E}\{X|Y\}|^n \} \quad (180)$$

where (178) and (179) are due to the generalized Hölder's inequality and Jensen's inequality, respectively. ■

## APPENDIX E

### PROOF OF PROPOSITION 9 ON THE DERIVATIVES

The first derivative of the mutual information with respect to the SNR is derived in [1] using the incremental channel technique. The same technique is adequate for the analysis of the derivatives of various other information theoretic and estimation theoretic quantities.

The MMSE of estimating an input with zero mean, unit variance and finite higher-order moments admits the Taylor series expansion at the vicinity of zero SNR given by (61). In general, given a random variable  $X$  with arbitrary mean and variance, we denote its central moments by

$$m_i = \mathbb{E} \{ (X - \mathbb{E}\{X\})^i \}, \quad i = 1, 2, \dots \quad (181)$$

Suppose all moments of  $X$  are finite, the random variable can be represented as  $X = \mathbb{E}\{X\} + \sqrt{m_2} Z$  where  $Z$  has zero mean and unit variance. Clearly,  $\mathbb{E} Z^i = m_2^{-\frac{i}{2}} m_i$ . By (61) and Proposition 2

$$\text{mmse}(X, \text{snr}) = m_2 \text{mmse}(Z, \text{snr} m_2) \quad (182)$$

$$\begin{aligned} &= m_2 - m_2^2 \text{snr} + (2m_2^3 - m_3^2) \frac{\text{snr}^2}{2} \\ &\quad - (m_4^2 - 6m_4 m_2^2 - 12m_3^2 m_2 + 15m_2^4) \frac{\text{snr}^3}{6} \\ &\quad + \mathcal{O}(\text{snr}^4). \end{aligned} \quad (183)$$

In general, taking into account the input variance, we have

$$\text{mmse}'(X, 0) = -m_2^2 \quad (184)$$

$$\text{mmse}''(X, 0) = 2m_2^3 - m_3^2 \quad (185)$$

$$\begin{aligned} \text{mmse}'''(X, 0) &= -m_4^4 + 6m_4m_2^2 + 12m_3^2m_2 \\ &\quad - 15m_2^4, \end{aligned} \quad (186)$$

Now that the MMSE at an arbitrary SNR is rewritten as the expectation of MMSEs at zero SNR, we can make use of known derivatives at zero SNR to obtain derivatives at any SNR. Let  $X_{y;\text{snr}} \sim P_{X|Y_{\text{snr}}=y}$ . Because of (184)

$$\left. \frac{d\text{mmse}(X_{y;\text{snr}}, \gamma)}{d\gamma} \right|_{\gamma=0^+} = -(\text{var}\{X|Y_{\text{snr}}=y\})^2. \quad (187)$$

Thus

$$\frac{d\text{mmse}(X, \text{snr})}{d\text{snr}} = \frac{d}{d\gamma} \text{mmse}(X, \text{snr} + \gamma) \Big|_{\gamma=0^+} \quad (188)$$

$$= \frac{d}{d\gamma} \text{mmse}(X, \gamma|Y_{\text{snr}}) \Big|_{\gamma=0^+} \quad (189)$$

$$= -E\left\{(\text{var}\{X|Y_{\text{snr}}\})^2\right\} \quad (190)$$

$$= -E\{M_2^2\} \quad (191)$$

where (189) is due to Proposition 3 and the fact that the distribution of  $Y_{\text{snr}}$  is not dependent on  $\gamma$ , and (190) is due to (187) and averaging over  $y$  according to the distribution of  $Y_{\text{snr}} = \sqrt{\text{snr}}X + N$ . Hence, (65) is proved. Moreover, because of (185)

$$\begin{aligned} \left. \frac{d^2\text{mmse}(X_{y;\text{snr}}, \gamma)}{d\gamma^2} \right|_{\gamma=0} &= 2(\text{var}\{X|Y_{\text{snr}}=y\})^3 \\ &\quad - (E\{(X - E\{X|Y_{\text{snr}}\})^3|Y_{\text{snr}}=y\})^2 \end{aligned} \quad (192)$$

which leads to (66) after averaging over the distribution of  $Y_{\text{snr}}$ . Similar arguments, together with (186), lead to the third derivative of the MMSE which is obtained as(67).

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments, which have helped to improve the paper noticeably. The authors would also like to thank M. Payaró, D. Palomar, and R. Bustin for their comments.

#### REFERENCES

- [1] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
- [2] S. Verdú and D. Guo, "A simple proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, no. 5, pp. 2165–2166, May 2006.
- [3] T. Weissman, "The relationship between causal and noncausal mismatched estimation in continuous-time AWGN channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4256–4273, Sep. 2010.
- [4] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1982–2010, Jun. 2005.
- [5] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.
- [6] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "Mimo Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.
- [7] D. Guo, S. Shamai, and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, WA, Jul. 2006, pp. 1011–1015.
- [8] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-Gaussianity of the sum of independent random variables: A simple proof," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4295–4297, Sep. 2006.
- [9] R. D. Yates and D. N. C. Tse, "K user fading broadcast channels with CSI at the receivers," in *Proc. Information Theory Application Workshop*, La Jolla, CA, 2011.
- [10] Y. Zhu and D. Guo, "Ergodic fading Z-interference channels without state information at transmitters," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, May 2011.
- [11] R. Bustin, R. Liu, H. V. Poor, and S. S. Shitz, "A MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel," *EURASIP J. Wireless Commun. and Netw., Special Issue on Wireless Physical Security*, Nov. 2009.
- [12] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, pp. 3712–3720, Aug. 2010.
- [13] D. Guo, "Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, 2009.
- [14] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
- [15] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3613–3628, Aug. 2009.
- [16] V. V. Buldygin and Y. V. Kozachenko, "Sub-Gaussian random variables," *Ukrainian Math. J.*, vol. 32, pp. 483–489, 1980.
- [17] G. Szegő, *Orthogonal Polynomials*, 4th ed. Providence, RI: American Mathematical Society, 1975.
- [18] H. L. Royden, *Real Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [19] V. Karunakaran, *Complex Analysis*, 2nd ed. Oxford, U.K.: Alpha Science International Ltd., 2005.
- [20] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Boston, MA: Birkhäuser, 2002.
- [21] D. Guo, S. Shamai (Shitz), and S. Verdú, "Estimation of non-Gaussian random variables in Gaussian noise: Properties of the MMSE," in *Proc. IEEE Int. Symp. Information Theory*, Toronto, ON, Canada, 2008.
- [22] V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.
- [23] Y. Wu and S. Verdú, "Functional properties of MMSE," in *Proc. Symp. Information Theory*, Austin, TX, Jun. 2010.
- [24] Y. Wu and S. Verdú, "Functional properties of MMSE and mutual information," *IEEE Trans. Inf. Theory*, submitted for publication.
- [25] Y. Wu and S. Verdú, "MMSE dimension," *IEEE Trans. Inf. Theory*, to be published.
- [26] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2317–2329, Jul. 2007.
- [27] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *J. Amer. Math. Soc.*, vol. 17, pp. 975–982, 2004.
- [28] R. Bustin, M. Payaró, D. P. Palomar, and S. S. Shitz, "On MMSE properties and I-mmse implications in parallel MIMO Gaussian channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, 2010, pp. 535–539.
- [29] D. Guo, "Gaussian Channels: Information, Estimation and Multiuser Detection," Ph.D.dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, 2004.
- [30] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, Oct. 1975.
- [31] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 451–456, 1978.
- [32] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, pp. 3014–3014, Jul.–Sep. 1974.
- [33] G. Kramer, "Topics in multi-user information theory," *Found. and Trends in Commun. Inf. Theory*, vol. 4, 2008.
- [34] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.
- [35] A. El Gamal and Y.-H. Kim, Lecture notes on network information theory [Online]. Available: <http://arxiv.org/abs/1001.3404>, 2010
- [36] D. Tuninetti, S. Shamai, and G. Caire, "Scalar fading Gaussian broadcast channels with perfect receiver CSI: Is Gaussian input optimal?," in *Proc. Workshop Inform. Theory Applications*, San Diego, CA, Jan. 2007.

- [37] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. Allerton Conf. Commun., Control, & Computing*, Monticello, IL, 2010.
- [38] M. Reed and B. Simon, *Methods of Mathematical Physics*. New York: Academic Press, 1975, vol. II.
- [39] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: Wiley, 1971, vol. II.
- [40] D. Stirzaker, *Elementary Probability*. New York: Cambridge Univ. Press, 2003.
- [41] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. Providence, RI: American Mathematical Society, 2001.
- [42] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. New York: Academic, 2007.
- [43] R. E. Greene and S. G. Krantz, *Function Theory of One Complex Variable*, 3rd ed. Providence, RI: American Mathematical Society, 2006.
- [44] S. Lang, *Complex Analysis*. Berlin, Germany: Springer-Verlag, 1999.

**Dongning Guo** (S'97–M'05) received the B.Eng. degree from the University of Science and Technology of China, the M.Eng. degree from the National University of Singapore, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ.

He was an R&D Engineer in the Center for Wireless Communications (now the Institute for Infocom Research), Singapore, from 1998 to 1999. He joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, in 2004, where he is currently an Associate Professor. He has held visiting positions at Norwegian University of Science and Technology in summer 2006, and at the Chinese University of Hong Kong in 2010–2011. His research interests are in information theory, communications, and networking.

Dr. Guo received the Huber and Suhner Best Student Paper Award in the International Zurich Seminar on Broadband Communications in 2000 and is a co-recipient of the 2010 IEEE Marconi Prize Paper Award in Wireless Communications. He is also a recipient of the National Science Foundation Faculty Early Career Development (CAREER) Award in 2007. He is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY in the area of Shannon Theory.

**Yihong Wu** (S'10) received the B.E. and M.A. degrees from Tsinghua University, Beijing, China, in 2006 and Princeton University in 2008, respectively, both in electrical engineering. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, Princeton University, Princeton, NJ.

His research interests are in information theory, signal processing, mathematical statistics, optimization, and distributed algorithms.

Mr. Wu is a recipient of the Princeton University Wallace Memorial Honorific Fellowship in 2010.

**Shlomo Shamai (Shitz)** (S'80–M'82–SM'88–F'94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 1975, 1981, and 1986, respectively.

From 1975 to 1985, he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986, he has been with the Department of Electrical Engineering, Technion, where he is now the William Fondiller Professor of Telecommunications. His research interests encompass a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is the recipient of the 2011 Claude E. Shannon Award. He is a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003 and the 2004 joint IT/COM societies paper award, and the 2007 IEEE Information Theory Society Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY and also serves on the Board of Governors of the Information Theory Society.

**Sergio Verdú** (S'80–M'84–SM'88–F'93) received the telecommunications engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980, and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, in 1984.

Since 1984, he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering.

Dr. Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from the IEEE Communications Society. He has also received paper awards from the Japanese Telecommunications Advancement Foundation and from Eurasp. He received the 2000 Frederick E. Terman Award from the American Society for Engineering Education for his book *Multiuser Detection*. He served as President of the IEEE Information Theory Society in 1997 and is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*.