Derivative of Mutual Information at Zero SNR: The Gaussian-Noise Case

Yihong Wu, *Student Member, IEEE*, Dongning Guo, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—Assuming additive Gaussian noise, a general sufficient condition on the input distribution is established to guarantee that the ratio of mutual information to signal-to-noise ratio (SNR) goes to one half nat as SNR vanishes. The result allows SNR-dependent input distribution and side information.**

***Index Terms*—Gaussian noise, low-power regime, minimum mean-square error (MMSE), mutual information, signal-to-noise ratio (SNR).**

I. Introduction

THE asymptotics of input-output mutual information of a class of channels with weak input has been investigated in the past (see [1] and references therein). This paper studies the mutual information between a random variable and its observation in Gaussian noise with low signal-to-noise ratio (SNR). The fundamental role of the derivative of mutual information at zero SNR was recognized by Shannon, who observed in [2] that a binary antipodal input is first-order optimal for the Gaussian channel in the low-SNR regime in the sense that it achieves the derivative of the capacity at zero SNR. In [3] and [4] it is shown that the zero-SNR derivative of mutual information determines the minimum energy per information bit required for reliable communication in the wideband regime. In the Gaussian noise case, this derivative[1] is equal to one half nat[2] [6], [7]. This result has proven to be useful in the treatment of capacity and spectral efficiency of fading channels (see, e.g., [6], [8]). It has also been used to obtain the derivative of the mutual information of a Gaussian model with respect to its SNR by translating the problem at any positive SNR to that of zero SNR with side information [7].

The asymptotic expansion of mutual information of additive Gaussian noise models with weak input has been addressed in the literature under various assumptions. The proof due to Lapidoth and Shamai [6, Lemma 5.2.1] makes use of a truncation argument and then applies [1, Theorem 1] to the peak-limited input signal. The peak-limited result is in turn a special case of Prelov's work [9, Lemma 2.2], which requires existence of input moments of αth order for some $\alpha > 2$. Furthermore, a simplified argument was given in [7] without delving into the technicalities required to justify the interchange of expectation and limit.

The goal of this paper is to present a rigorous, self-contained proof of the asymptotic result allowing a more general setup than those considered before, in which the input distribution can depend on the SNR, and an SNR-dependent side information may be available to the receiver. Unlike the proof in [10] for [7, Lemma 1], the proof here uses basic inequalities in lieu of convergence theorems of integrals, and in fact yields nonasymptotic bounds on the first-order expansion of the mutual information. The generalized result is also useful in the context of the SNR-incremental channel [7], where the side information is dependent on the SNR, so that the conditional input distribution is also SNR-dependent.

II. Main Result

Theorem 1: Let $\{Z_\delta\}_{\delta>0}$ be a collection of random variables which satisfy

$$\lim_{\delta\to 0} \mathsf{var}\{Z_\delta\} = \sigma^2 < \infty. \tag{1}$$

Suppose also that $\{Z_\delta^2\}_{\delta>0}$ is uniformly integrable, i.e.

$$\lim_{t\to\infty} \sup_{\delta>0} \mathsf{E}\left\{Z_\delta^2 \cdot 1_{\{Z_\delta^2>t\}}\right\} = 0. \tag{2}$$

Then, as $\delta \to 0$

$$I(Z_\delta; \sqrt{\delta}Z_\delta + W) = \frac{\sigma^2}{2}\delta + o(\delta) \tag{3}$$

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

Remark 1: The first-order approximation of the mutual information in (3) depends only on the SNR, and is independent of the input distribution otherwise.

Remark 2: For Gaussian noise, all the first-order expansions shown in the literature, in particular, [9], [6], [7], [11]–[13], [4] are special cases of Theorem 1. For instance:

- In [7, Lemma 1], there is a special case where the input $Z_\delta = Z$ is independent of the gain δ. In this case, the uniform integrability reduces to $\mathsf{E}\{Z^2\} < \infty$.
- In [9, Theorem 2.2] requires $\mathsf{E}\{|X|^\alpha\} < \infty$ for some $\alpha > 2$ while Theorem 1 implies that $\alpha = 2$ is sufficient.
- The first-order expansion in [12, Theorem 1] assumes that the input distribution has a density that is differentiable at zero.

Manuscript received December 19, 2010; revised April 22, 2011; accepted June 06, 2011. Date of current version November 11, 2011. This work was supported in part by the NSF under grants CCF-0644344 and CCF-0635154.

Y. Wu and S. Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

D. Guo is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA.

Communicated by D. Palomar, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2011.2161752

[1]A finer Taylor series expansion is found in [5].

[2]Throughout the paper, natural logarithms are adopted and information units are nats.

- Reference [14, Lemma III-1] deals with more general transformations subject to certain regularity conditions, which includes Gaussian channel as a particular case; however, it only applies to input distributions with *bounded* support. In this special case, our derivation reveals the stronger result that the $o(\delta)$ term in (3) is in fact $O(\delta^{\frac{3}{2}})$.

Higher-order expansion of mutual information are obtained in [15] for Gaussian noise, [16] for non-Gaussian noise, and [17] for general models under various technical conditions.

Remark 3: The uniform integrability condition in Theorem 1 is not superfluous. For example, consider the following input:

$$Z_\delta = \begin{cases} \delta^{-1} & \text{with probability } \delta^2, \\ 0 & \text{with probability } 1-\delta^2 \end{cases} \tag{4}$$

whose square is not uniformly integrable. Then $\mathsf{var}\{Z_\delta\} \to 1$ as $\delta \to 0$. However, as shown in Appendix A, [3]

$$I(Z_\delta; \sqrt{\delta} Z_\delta + W) = \Theta\left(\delta^2 \log \frac{1}{\delta}\right) \tag{5}$$

hence (3) does not hold for this sequence of input distributions.

Theorem 1 can be generalized to Gaussian models when side information about the input is available to the receiver. Denote by $\mathsf{var}\{Z|V\}$ the conditional variance of Z given the random variable V, which is a function of V. Its expectation is the MMSE of estimating Z from the side information V

$$\mathsf{mmse}(Z|V) = \mathsf{E}\{(Z - \mathsf{E}\{Z|V\})^2\} \tag{6}$$
$$= \mathsf{E}\{\mathsf{var}\{Z|V\}\}. \tag{7}$$

The following result generalizes Theorem 1.

Theorem 2: Let $\{Z_\delta, V_\delta\}_{\delta>0}$ be a collection of jointly distributed random variables satisfying:

$$\lim_{\delta\to 0} \mathsf{mmse}(Z_\delta|V_\delta) = \sigma^2 < \infty. \tag{8}$$

Suppose that the uniform integrability condition (2) is satisfied. Let $W \sim \mathcal{N}(0,1)$ be independent of $\{Z_\delta, V_\delta\}_{\delta>0}$. Then, as $\delta \to 0$

$$I(Z_\delta; \sqrt{\delta} Z_\delta + W | V_\delta) = \frac{\sigma^2}{2}\delta + o(\delta). \tag{9}$$

The remainder of this paper is devoted to a simple, rigorous proof of Theorem 2, which also establishes Theorem 1 and [7, Lemma 1] as special case. Peak-limited inputs are first treated in Section III. In Section IV, we extend the result to arbitrarily distributed inputs by sending the peak limit to infinity.

III. Peak-Limited Input

Lemma 1: Fix $0 \le \delta \le 1$. Let (Z, V) be any jointly distributed random variables where V takes values on an arbitrary alphabet and the essential supremum of Z is finite

$$\|Z\|_\infty = M < \infty. \tag{10}$$

Then

$$\left| I(Y; Z|V) - \frac{\delta}{2}\mathsf{mmse}(Z|V) \right| \le 5\delta^{\frac{3}{2}} e^{10M^2} \tag{11}$$

where $Y = \sqrt{\delta} Z + W$ with $W \sim \mathcal{N}(0,1)$ independent of (Z, V).

Proof: To establish the upper bound (11), which depends only on δ and the peak amplitude of the input Z, we need the following basic inequalities proved in Appendix B:

1) For any $x \in \mathbb{R}$

$$e^x \ge 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3. \tag{12}$$

2) Let $A > 0$. Then for any $x \le A$

$$e^x \le 1 + x + \frac{e^A}{A^2}x^2 \tag{13}$$

and

$$e^x \le 1 + x + \frac{1}{2}x^2 + \frac{e^A}{A^3}(x^3)^+ \tag{14}$$

where $x^+ \triangleq \max\{x, 0\}$.

3) For any $x > -1$

$$\log(1+x) \ge x - \frac{x^2}{1+x}. \tag{15}$$

Without loss of generality, we assume that $\mathsf{E}\{Z\} = 0$. Since the right-hand side (RHS) of (11) does not depend on the side information V, it suffices to consider the case where the side information is absent, i.e., V is independent of Z.

Denote the standard normal density by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{16}$$

The input-output conditional distribution is Gaussian

$$p_{Y|Z}(y|z) = \varphi(y - \sqrt{\delta} z) \tag{17}$$

whose average over Z is the output density

$$p_Y(y) = \mathsf{E}\{\varphi(y - \sqrt{\delta}\, Z)\} \tag{18}$$
$$= \varphi(y)\mathsf{E}\{\mathrm{e}^{U(y)}\} \tag{19}$$

where

$$U(y) \triangleq \sqrt{\delta} y Z - \frac{1}{2}\delta Z^2. \tag{20}$$

Clearly

$$\mathsf{E}\{U(y)\} = -\frac{1}{2}\delta \mathsf{var}\{Z\}. \tag{21}$$

Since

$$\mathsf{E}\{Y\} = 0 \tag{22}$$
$$\mathsf{var}\{Y\} = \delta \mathsf{var}\{Z\} + 1 \tag{23}$$

the mutual information can be expressed as

$$I(Y; Z) = \mathsf{E}\left\{\log \frac{p_{Y|Z}(Y|Z)}{p_Y(Y)}\right\} \tag{24}$$
$$= \frac{\delta}{2}\mathsf{var}\{Z\} - \mathsf{E}\{J(Y)\} \tag{25}$$

[3]We use the standard asymptotic notations: $f(x) = O(g(x))$ if $\limsup \frac{|f(x)|}{|g(x)|} < \infty$, $f(x) = \Omega(g(x))$ if $g(x) = O(f(x))$, $f(x) = \Theta(g(x))$ if $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$, $f(x) = o(g(x))$ if $\lim \frac{|f(x)|}{|g(x)|} = 0$.

where we have defined

$$J(y) \triangleq \log \mathsf{E}\{e^{U(y)}\} \tag{26}$$

to decouple the expectation over Z (or U) and the expectation over Y.

The remainder of the proof is devoted to showing

$$|\mathsf{E}\{J(Y)\}| \leq 5\delta^{\frac{3}{2}} e^{10M^2} \tag{27}$$

which, in view of (25), implies the desired result

$$\left| I(Y;Z) - \frac{\delta}{2}\mathsf{var}\{Z\} \right| \leq 5\delta^{\frac{3}{2}} e^{10M^2}. \tag{28}$$

Upper Bound: In view of (10) and (20), we have

$$U \leq |y|\sqrt{\delta}M \leq |y|M. \tag{29}$$

Using the fact that $\log x \leq x - 1$ and then (14) with $A = |y|M$, we can write

$$\begin{aligned} J(y) &\leq \mathsf{E}\{e^U\} - 1 &(30)\\ &\leq \mathsf{E}\left\{U + \frac{1}{2}U^2\right\} + e^{M|y|}\delta^{\frac{3}{2}} &(31)\end{aligned}$$

where the last step uses the first inequality in (29). Plugging (20) and (29) into (31), we obtain

$$J(y) \leq \frac{1}{2}(y^2-1)\delta M^2 + \left(e^{M|y|} - \frac{y}{2}\mathsf{E}\{Z^3\}\right)\delta^{\frac{3}{2}} + \frac{M^4\delta^2}{8}. \tag{32}$$

Taking expectation of (32) yields the upper bound:[4]

$$\begin{aligned} \mathsf{E}\{J(Y)\} &\leq \left(2e^{\frac{3}{2}M^2} + \frac{5}{8}M^4\sqrt{\delta}\right)\delta^{\frac{3}{2}} &(33)\\ &\leq 3\delta^{\frac{3}{2}} e^{2M^2} &(34)\end{aligned}$$

where we have used (22)

$$\mathsf{E}\{e^{M|Y|}\} \leq e^{M^2}\mathsf{E}\{e^{M|W|}\} \leq 2e^{\frac{3}{2}M^2} \tag{35}$$

and

$$M^2 \leq e^{M^2-1}. \tag{36}$$

Lower Bound: Applying (15) yields[5]

$$J(y) \geq \mathsf{E}\{e^U - 1\} - \frac{|\mathsf{E}\{e^U - 1\}|^2}{\mathsf{E}\{e^U\}}. \tag{37}$$

In view of (13) with $A = |y|M$ and (29), we have

$$\mathsf{E}\{e^U - 1\} \leq \mathsf{E}\{U\} + e^{|y|M}\delta. \tag{38}$$

On the other hand

$$\mathsf{E}\{e^U - 1\} \geq \mathsf{E}\{U\}. \tag{39}$$

[4]Note that directly applying Hoeffding's inequality [18, Lemma 1] to (26) only gives an $O(\delta)$ upper bound.

[5]Although the RHS of (37) is equal to $1 - \frac{1}{\mathsf{E}\{e^U\}}$, we prefer to use (37) because, as we shall see in the remainder of Section III, a much looser bound for the denominator will be enough.

Combining (20), (38), and (39) yields

$$\begin{aligned} |\mathsf{E}\{e^U - 1\}| &\leq |\mathsf{E}\{U\}| + e^{|y|M}\delta &(40)\\ &\leq \left(e^{|y|M} + \frac{1}{2}M^2\right)\delta. &(41)\end{aligned}$$

In view of (20) and (10), we have $U \geq -M|y| - \frac{1}{2}M^2$, which implies

$$\mathsf{E}\{e^U\} \geq e^{-M|y| - \frac{1}{2}M^2}. \tag{42}$$

Plugging (41) and (42) into (37) gives

$$\begin{aligned} J(y) &\geq \mathsf{E}\{e^U - 1\} - \frac{\delta^2}{4}(2e^{|y|M} + M^2)^2 e^{M|y| + \frac{M^2}{2}} &(43)\\ &\geq \mathsf{E}\{e^U - 1\} - \frac{9\delta^2}{4}e^{3M|y| + \frac{5}{2}M^2} &(44)\end{aligned}$$

where we have used (36) repeatedly. By (12) and (20)

$$\begin{aligned} &\mathsf{E}\{e^U - 1\}\\ &\geq \mathsf{E}\left\{U + \frac{U^2}{2} + \frac{U^3}{6}\right\} &(45)\\ &= \frac{\delta}{2}(y^2-1)\mathsf{E}\{Z^2\} + \frac{\delta^{3/2}}{6}(y^3 - 3y)\mathsf{E}\{Z^3\}\\ &+ \frac{\delta^2}{8}(1-2y^2)\mathsf{E}\{Z^4\} + \frac{\delta^{5/2}}{8}y\mathsf{E}\{Z^5\} - \frac{\delta^3}{48}\mathsf{E}\{Z^6\}. &(46)\end{aligned}$$

Similar to (35), we have $\mathsf{E}\left\{e^{3M|Y|}\right\} \leq e^{3M^2}\mathsf{E}\left\{e^{3M|W|}\right\} \leq 2e^{\frac{15}{2}M^2}$. Substituting (46) into (44) and taking expectation yield

$$\begin{aligned} \mathsf{E}\{J(Y)\} &\geq \frac{\delta^2}{2}(\mathsf{E}\{Z^2\})^2 + \frac{\delta^3}{6}(\mathsf{E}\{Z^3\})^2 - \frac{\delta^3}{48}\mathsf{E}\{Z^6\}\\ &\quad - \frac{\delta^2}{8}(2\delta\mathsf{E}\{Z^2\} - 1)\mathsf{E}\{Z^4\} - \frac{9\delta^2}{2}e^{10M^2} &(47)\\ &\geq -\frac{\delta^3}{48}M^6 - \frac{\delta^3}{4}M^6 - \frac{9\delta^2}{2}e^{10M^2} &(48)\\ &\geq -5\delta^2 e^{10M^2} &(49)\end{aligned}$$

where we have used (22), (23), (10), and (36). ■

IV. Proof of Theorem 2

To prove Theorem 2, we find upper and lower bounds on the conditional mutual information $I(Z_\delta; \sqrt{\delta}Z_\delta + W \big| V_\delta)$ and show that they coincide. Conditioned on $V_\delta = v$, the mutual information cannot be greater than if the input were Gaussian with the same conditional variance: for all $\delta > 0$

$$I(Z_\delta; Y_\delta | V_\delta = v) \leq \frac{1}{2}\log(1 + \delta\mathsf{var}\left\{Z_\delta \big| V_\delta = v\right\}). \tag{50}$$

By Jensen's inequality and (8)

$$\begin{aligned} I(Z_\delta; Y_\delta | V_\delta) &\leq \frac{1}{2}\log(1 + \delta\mathsf{mmse}(Z_\delta | V_\delta)) &(51)\\ &= \frac{\delta}{2}\sigma^2 + o(\delta) &(52)\end{aligned}$$

as $\delta \to 0$. It remains to show the lower bound

$$\liminf_{\delta \downarrow 0} \frac{I(Z_\delta; Y_\delta | V_\delta)}{\delta} \geq \frac{\sigma^2}{2}. \tag{53}$$

By Lemma 1 and (8), (53) holds for peak-limited inputs, since

$$I(Z_\delta; Y_\delta | V_\delta) \geq \frac{\delta}{2} \mathsf{mmse}(Z_\delta | V_\delta) - 5\delta^{\frac{3}{2}} e^{10M^2} \tag{54}$$

where $M = \sup_\delta \|Z_\delta\|_\infty$. Next we prove that truncating Z_δ beyond a threshold t has vanishing impact on the conditional mutual information and the MMSE as $t \to \infty$.

To streamline notations, define

$$K(t) = \sup_{\delta > 0} \mathsf{E}\left\{ Z_\delta^2 \mathbf{1}_{\{|Z_\delta| > t\}} \right\}. \tag{55}$$

By the uniform integrability of $\{Z_\delta^2\}_{\delta > 0}$, we have

$$\lim_{t \to \infty} K(t) = 0. \tag{56}$$

Next define Z_δ^t to be a random variable distributed according to the distribution of Z_δ conditioned on the event $\{|Z_\delta| \leq t\}$, i.e., define the joint distribution of (Z_δ^t, V_δ) according to

$$\mathsf{P}(V_\delta \in A, Z_\delta^t \in B) = \frac{\mathsf{P}(V_\delta \in A, Z_\delta \in B \cap [-t, t])}{\mathsf{P}(Z_\delta \in [-t, t])} \tag{57}$$

for all measurable subsets A and B. Note that the denominator in (57) is positive for all sufficiently large t. This can be seen from (56) and Chebyshev's inequality

$$\mathsf{P}(|Z_\delta| > t) \leq t^{-2} K(t). \tag{58}$$

We proceed by noting that

$$\begin{aligned} I&\left(Z_\delta; Y_\delta \middle| V_\delta\right) \\ &= I\left(Z_\delta, 1_{\{|Z_\delta| \leq t\}}; \sqrt{\delta} Z_\delta + W \middle| V_\delta\right) && (59) \\ &\geq I\left(Z_\delta; \sqrt{\delta} Z_\delta + W \middle| V_\delta, 1_{\{|Z_\delta| \leq t\}}\right) && (60) \\ &\geq \mathsf{P}(|Z_\delta| \leq t) I\left(Z_\delta; \sqrt{\delta} Z_\delta + W \middle| V_\delta, |Z_\delta| \leq t\right) && (61) \\ &= \mathsf{P}(|Z_\delta| \leq t) I\left(Z_\delta^t; \sqrt{\delta} Z_\delta^t + W \middle| V_\delta\right) && (62) \end{aligned}$$

where (60) follows from the chain rule of mutual information and (62) follows from the definition of Z_δ^t. In view of Lemma 1

$$\frac{I\left(Z_\delta^t; \sqrt{\delta} Z_\delta^t + W \middle| V_\delta\right)}{\delta} \geq \frac{1}{2} \mathsf{mmse}\left(Z_\delta^t | V_\delta\right) - 5\delta^{\frac{1}{2}} e^{10t^2}. \tag{63}$$

Next we show that $\mathsf{mmse}\left(Z_\delta^t | V_\delta\right)$ and $\mathsf{mmse}(Z_\delta | V_\delta)$ are close when the threshold t is large. Since

$$\hat{Z}_\delta^t(v) \triangleq \mathsf{E}\left\{ Z_\delta^t \middle| V_\delta = v \right\} \tag{64}$$

is a suboptimal estimator of Z_δ, we have

$$\begin{aligned} \mathsf{mmse}(Z_\delta | V_\delta) &\leq \mathsf{E}\left\{ (Z_\delta - \hat{Z}_\delta^t(V_\delta))^2 \right\} && (65) \\ &= \mathsf{E}\left\{ (Z_\delta - \hat{Z}_\delta^t(V_\delta))^2 \middle| |Z_\delta| \leq t \right\} \mathsf{P}\left\{|Z_\delta| \leq t\right\} \\ &\quad + \mathsf{E}\left\{ (Z_\delta - \hat{Z}_\delta^t(V_\delta))^2 \mathbf{1}_{\{|Z_\delta| > t\}} \right\} && (66) \\ &\leq \mathsf{E}\left\{ (Z_\delta^t - \hat{Z}_\delta^t(V_\delta))^2 \right\} \mathsf{P}\left\{|Z_\delta| \leq t\right\} \\ &\quad + 4\, \mathsf{E}\left\{ |Z_\delta|^2 \mathbf{1}_{\{|Z_\delta| > t\}} \right\} && (67) \\ &\leq \mathsf{mmse}(Z_\delta^t | V_\delta) \mathsf{P}\left\{|Z_\delta| \leq t\right\} + 4K(t) && (68) \end{aligned}$$

where (67) is because, conditioned on the event $\{|Z_\delta| \leq t\}$, Z_δ has the same distribution as Z_δ^t, and (68) is due to (55). Substituting (63) and (68) into (62) yields

$$\frac{I(Z_\delta; Y_\delta | V_\delta)}{\delta} \geq \frac{1}{2} \mathsf{mmse}(Z_\delta | V_\delta) - 2K(t) - 5\delta^{\frac{1}{2}} e^{10t^2} \tag{69}$$

for all $0 \leq \delta \leq 1$. In view of (8), we have

$$\liminf_{\delta \downarrow 0} \frac{I(Z_\delta; Y_\delta | V_\delta)}{\delta} \geq \frac{\sigma^2}{2} - 2K(t). \tag{70}$$

By the arbitrariness of t and (56), we obtain the desired result in (53). Hence the proof of Theorem 2 is complete.

V. Concluding Remarks

This paper uses elementary inequalities to establish rigorously the first-order approximation of the input-output mutual information of Gaussian channels with weak input. In the special case of peak-limited inputs, the bounds are non-asymptotic in the signal-to-noise ratio.

The original incremental-SNR channel described in [7] leads directly to the right derivative with respect to nonzero SNR. To treat the left derivative, we let the stronger channel have SNR equal to γ and let the slightly degraded channel have SNR equal to $\gamma - \delta$. If we consider the mutual information or estimation error conditioned on the weaker observation, care has to be taken because the weaker observation is made at an SNR dependent on the small decrement δ. Nonetheless, the new theorems in this paper address such derivatives rigorously.

The key results in this paper also apply to some other information measures. Consider the same model and assumptions as in Theorem 2 with an additional condition that Z_δ has finite essential supremum. In view of Lemma 1, we have

$$I(Z_\delta; Y_\delta | V_\delta) = \frac{\delta}{2} \mathsf{mmse}(Z_\delta | V_\delta) + O\left(\delta^{\frac{3}{2}}\right). \tag{71}$$

Let Φ_{Y_δ} be Gaussian with the same mean and variance as Y_δ. Using the following decomposition of mutual information:

$$\begin{aligned} I&(Z_\delta; Y_\delta | V_\delta) \\ &= D(P_{Y_\delta | Z_\delta, V_\delta} \| P_{\Phi_{Y_\delta}} | P_{Z_\delta V_\delta}) - D(P_{Y_\delta | V_\delta} \| P_{\Phi_{Y_\delta}} | P_{V_\delta}) && (72) \\ &= \frac{\delta}{2} \mathsf{mmse}(Z_\delta | V_\delta) - D(P_{Y_\delta | V_\delta} \| P_{\Phi_{Y_\delta}} | P_{V_\delta}) && (73) \end{aligned}$$

we arrive at the following estimate of the conditional non-Gaussianness of Y_δ:

$$D(P_{Y_\delta} \| P_{\Phi_{Y_\delta}} | P_{V_\delta}) = O\left(\delta^{\frac{3}{2}}\right) \tag{74}$$

which slightly improves the $O(\delta)$ result in [14, Lemma III-1].

APPENDIX A
PROOF OF (5)

Denote by $\hat{Z}_\delta$ the estimate of the binary input Z_δ based on Y_δ with the smallest error probability

$$\mathsf{P}\left\{\hat{Z}_\delta \neq Z_\delta\right\} \leq Q\left(\frac{1}{2\sqrt{\delta}}\right) \leq e^{-\frac{1}{8\delta}} \tag{75}$$

where $Q(x) \triangleq \int_x^\infty \varphi(z)\mathrm{d}z$ and $Q(x) \leq e^{-\frac{x^2}{2}}$. Let $h(\cdot)$ denote the binary entropy function. Then

$$\begin{aligned}
h(\delta^2) &= H(Z_\delta) && (76)\\
&\geq I(Z_\delta; Y_\delta) && (77)\\
&\geq I(Z_\delta; \hat{Z}_\delta) && (78)\\
&= H(Z_\delta) - H(Z_\delta|\hat{Z}_\delta) && (79)\\
&\geq h(\delta^2) - h\left(\mathsf{P}\left\{Z_\delta \neq \hat{Z}_\delta\right\}\right) && (80)\\
&\geq h(\delta^2) - h\left(e^{-\frac{1}{8\delta}}\right) && (81)\\
&\geq h(\delta^2)\left(1 - \frac{e^{-\frac{1}{8\delta}}}{\delta^2}\right) && (82)
\end{aligned}$$

where (82) follows from

$$qh(p) \leq h(qp),\ (p,q) \in [0,1]^2. \tag{83}$$

Then (5) follows from the fact that $h(\delta^2) = \Theta(\delta^2 \log \frac{1}{\delta})$ and the fraction in (82) vanishes.

APPENDIX B
PROOF OF (12)–(15)

By Taylor's theorem with remainder of Lagrange form [19], there exists ξ between 0 and x, such that

$$\mathrm{e}^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{\mathrm{e}^\xi}{24}x^4 \tag{84}$$

which implies (12). Similarly, we have the following:

$$\begin{aligned}
\mathrm{e}^x &\leq 1 + x + \frac{1}{2}x^2, \quad x \leq 0 && (85)\\
\mathrm{e}^x &> 1 + x + \frac{1}{2}x^2, \quad x > 0. && (86)
\end{aligned}$$

In view of (85), it suffices to show (14) for $x \in (0, A)$. Let

$$f(x) = \frac{\mathrm{e}^x - 1 - x - \frac{x^2}{2}}{x^3}. \tag{87}$$

Furthermore, by Taylor's theorem with remainder of integral form [19]

$$f(x) = \frac{1}{2}\int_0^1 (1-t)^2 \mathrm{e}^{tx}\mathrm{d}t. \tag{88}$$

Therefore f is increasing on $\mathbb{R}$. Then for all $x \leq A$

$$f(x) \leq f(A) < \frac{\mathrm{e}^A}{A^3} \tag{89}$$

which implies (14). Inequality (13) follows analogously as

$$\frac{\mathrm{e}^x - 1 - x}{x^2} = \int_0^1 (1-t)\mathrm{e}^{tx}\mathrm{d}t \tag{90}$$

is an increasing function in x on $\mathbb{R}$.

To prove (15), define

$$g(x) = \frac{\log(1+x) - x}{x^2} \tag{91}$$

which is increasing on $(-1, \infty)$ because

$$g(x) = -\int_0^1 \frac{1-t}{(1+tx)^2}\mathrm{d}t. \tag{92}$$

Then

$$g(x) \geq -\int_0^1 \frac{1}{(1+tx)^2}\mathrm{d}t = -\frac{1}{1+x} \tag{93}$$

which implies (15).

ACKNOWLEDGMENT

One of the authors (D. Guo) would like to thank Dr. J. Luo for useful discussions during the early stage of this study. The paper has benefited from thorough suggestions by the anonymous reviewers.

REFERENCES

[1] V. V. Prelov and E. C. van der Meulen, "An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals," *IEEE Trans. Inf. Theory*, vol. 39, pp. 1728–1735, Sep. 1993.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Oct. 1948.

[3] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[4] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory, Special Issue on Shannon Theory: Perspective, Trends Appl.*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[5] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

[6] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[7] 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.

[8] M. C. Gursoy, H. V. Poor, and S. Verdú, "Noncoherent Rician fading channel—Part II: Spectral efficiency in the low-power regime," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2193–2221, 2005.

[9] V. V. Prelov, "Communication channel capacity with almost Gaussian noise," *Theory Probab. Appl.*, vol. 33, pp. 405–422, Apr. 1987.

[10] D. Guo and J. Luo, *On the Technicality of the Mutual Information-MMSE Formula.* Technical Report, Northwestern University, Evanston, IL, 2008.

[11] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[12] C. Rao and B. Hassibi, "Analysis of multiple-antenna wireless links at low SNR," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2123–2130, 2004.

[13] M. S. Pinsker, V. V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pt. 2, pp. 1877–1888, Nov. 1995.

[14] B. Hajek and V. G. Subramanian, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 828–839, 2002.

[15] V. V. Prelov, "Communication channel capacity with almost Gaussian noise," *Theory Probab. Appl.*, vol. 33, pp. 405–405, 1988.

[16] V. V. Prelov, "Asymptotic expansions for the mutual information and for the capacity of continuous memoryless channels with weak input signal," *Problems of Contr. Inf. Theory*, vol. 18, no. 2, pp. 91–106, 1989.

[17] V. V. Prelov and E. C. van der Meulen, "Higher-order asymptotics of mutual information for nonlinear channels with non-Gaussian noise," *Problems Inf. Transmiss.*, vol. 39, no. 4, pp. 324–340, 2003.

[18] S. Boucheron, G. Lugosi, and O. Bousquet, "Concentration inequalities," in *Advanced Lectures on Machine Learning*, O. Bousquet, U. von Luxburg, and G. Rätsch, Eds. New York: Springer, 2004, pp. 208–240.

[19] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

Yihong Wu (S'10) received the B.E. and M.A. degrees from Tsinghua University, Beijing, China, in 2006 and Princeton University, Princeton, NJ, in 2008, respectively, both in electrical engineering.

He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, Princeton University. 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.

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 with 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, 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. 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 corecipient 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.

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 Eurasip. 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*.