Second-Order Asymptotics of Mutual Information

Viacheslav V. Prelov and Sergio Verdú, *Fellow, IEEE*

***Abstract*—A formula for the second-order expansion of the input–output mutual information of multidimensional channels as the signal-to-noise ratio (SNR) goes to zero is obtained. While the additive noise is assumed to be Gaussian, we deal with very general classes of input and channel distributions. As special cases, these channel models include fading channels, channels with random parameters, and channels with almost Gaussian noise. When the channel is unknown at the receiver, the second term in the asymptotic expansion depends not only on the covariance matrix of the input signal but also on the fourth mixed moments of its components. The study of the second-order asymptotics of mutual information finds application in the analysis of the bandwidth–power tradeoff achieved by various signaling strategies in the wideband regime.**

***Index Terms*—Channel capacity, fading channels, low-power communication, mutual information, nonlinear channels.**

I. Introduction

Recent results on the capacity of channels in the wideband regime [1] have revealed the important role played by the second derivative of the Shannon capacity at zero signal-to-noise ratio (SNR). The first derivative of capacity (at zero SNR) is known (e.g., [2]) to yield the minimum energy per bit required to transmit information reliably in the absence of bandwidth constraints. For very general fading channels, [1] shows that the received minimum energy per bit is 1.59 dB below the one-sided spectral noise level. The bandwidth required to send a given data rate with given power is proportional, in the low-power regime, to minus the second derivative at zero SNR [1].

An input distribution is said to be *first-order optimal* [1] if the derivative at zero SNR of its mutual information is equal to the derivative of capacity at zero SNR. A first-order optimal distribution achieves the minimum energy per information bit when bandwidth is allowed to be infinite. Denote by $C(\rho)$ the capacity (per channel use) as a function of the SNR, and let E_b be the transmitted energy per information bit. Then the minimum E_b/N_0 required for reliable communication is [1]

$$\left(E_b/N_0\right)_{\min} = \frac{\log_e 2}{\dot{C}(0)} \tag{1}$$

where $\dot{C}(0)$ denotes the first derivative at 0 of the function $C(\rho)$ in nats. If a distribution is not first-order optimal, the minimum energy per bit that it achieves can be computed through (1) replacing $C(\rho)$ by the mutual information as a function of ρ.

An important measure of the bandwidth–power tradeoff in the low-power regime is the "wideband slope" introduced in [1]. It is shown in [1] that the slope (b/s/Hz/(3 dB)) of the spectral efficiency at the point $(E_b/N_0)_{\min}$ is given by the formula

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

where $\ddot{C}(0)$ denotes the second derivative at 0 of the function $C(\rho)$ in nats. An input distribution is said to be *second-order optimal* [1] if it is first-order optimal and the second derivative at zero SNR of its mutual information is equal to the second derivative of capacity at zero SNR. In the wideband regime, for a given power and rate, a second-order optimal distribution requires the minimum possible bandwidth. An important motivation for the asymptotic expansions found in this paper is the comparison of the minimum energy and wideband slope achieved by various inputs.

For the additive white Gaussian noise (AWGN) channel, the optimum input distribution simply scales with the allowed power. In other words, for a given power constraint, the optimum input distribution does not depend on the noise power. In contrast, for fading channels not fully known at the receiver, the structure of the input distribution depends on the SNR. For example, in the case of the scalar noncoherent Rayleigh channel, it is known [3], [4] that the optimum distribution is discrete and that for sufficiently low SNR, the optimum distribution has a mass at zero and a vanishing mass that migrates to infinity as the SNR decreases. Because the capacity-achieving input signaling may be unknown or because of complexity, peak-to-average, or other constraints, in practice it is common to use a given noncapacity-achieving constellation whose structure is not adapted to the SNR. As we saw above, the power penalty in the wideband regime incurred by a given suboptimal input distribution is equal to the ratio of the first derivatives of the input–output mutual information achieved with the given input distribution and with the capacity-achieving distribution. For two distributions that achieve the same first derivative, the ratio of their second derivatives is equal to the ratio of their respective required bandwidths, for a given rate, in the wideband limit [1].

Motivated by the lack of explicit mutual-information expressions as a function of SNR for many channels and inputs of interest, the mutual information has been studied in the low-power regime in several previous works [1], [2], [5]–[9]. Non-Gaussian additive noise is treated in [6], [9], with not only input power restrictions but also peak-power limitations. With the exception of [1], [7], [8], those works limit themselves to the first-order asymptotic expansion of mutual information, and the scope of

Manuscript received May 10, 2002; revised November 19, 2003. This work was supported in part by DIMACS, by the National Science Foundation under Grant NCR-0074277, and by the Russian Foundation for Basic Research (RFFI) under Grant 03-01-00592.

V. V. Prelov is with the Institute for Problems of Information Transmission, Russian Academy of Sciences, Moscow 101447, Russia.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by İ. E. Telatar, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2004.831784

0018-9448/04$20.00 © 2004 IEEE

[8] is a channel not subject to fading or impairments other than additive noise which is slightly non-Gaussian (see also [10]).

In the present paper, the channel output vector is the sum of a Gaussian vector with independent equal-strength components and an input-dependent vector. The input-dependent signal is extremely general, encompassing stochastic and nonlinear effects. In addition to limiting the weight of the asymptotic tails of the input vector, our sufficient conditions on the channel are technical conditions that preclude anomalous behavior of high-order moments. All these technical conditions are normally satisfied in channels of practical interest.

In Section II, we present the general channel model and special cases of interest. Although in most applications of current interest the channel is linear, our basic framework is a very general, not necessarily linear, setting. Section III introduces several measures of dependence based on the conditional mean and conditional variance. These measures play a key role in the statement of the main results of the paper, which are given in Section IV and proved in Section V.

II. Model

A. Nonlinear Channels

Consider the general communication channel model where the complex-valued output signal vector $\boldsymbol{y} = (\mathsf{y}_1, \ldots, \mathsf{y}_m)^T$[1] is given by

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{n}$ is a complex Gaussian random vector with independent components with independent real and imaginary parts each with zero mean and variance $N_0/2$; the channel input vector $\boldsymbol{x} = (\mathsf{x}_1, \ldots, \mathsf{x}_n)^T$ also has complex-valued components with finite second moments; and the input-dependent received signal is a proper-complex[2] random m-vector $\boldsymbol{g}(\boldsymbol{x})$. We assume throughout that the pair $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))$ and $\boldsymbol{n}$ are mutually independent.

B. Fading Channels

A fading channel

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{4}$$

where $\boldsymbol{H} = (\mathsf{h}_{ij})$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, is an $m \times n$ matrix with random complex-valued components h_{ij}, can be considered a special case of channel (3) where

$$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{H}\boldsymbol{x}. \tag{5}$$

For the fading channel (4), $\boldsymbol{x}$, $\boldsymbol{H}$, and $\boldsymbol{n}$ are assumed to be mutually independent. The channel in (4) encompasses a variety of features arising in fading channels such as frequency selectivity, multiantenna, spread spectrum, and multiuser communication (e.g., [1]).

In this paper, we obtain asymptotic expressions for mutual information with and without knowledge of $\boldsymbol{H}$ at the receiver.

[1] T denotes transposition.

[2] A complex-valued random vector $\boldsymbol{z}$ is proper-complex (in the sense of [11]) if $E[\boldsymbol{z}\boldsymbol{z}^T] = E[\boldsymbol{z}]E[\boldsymbol{z}]^T$.

C. Additive White Gaussian Noise (AWGN) Channels

If $m = n$ and $\boldsymbol{H} = \boldsymbol{I}$ almost surely (a.s.) where $\boldsymbol{I}$ is the $m \times m$ identity matrix, then (4) reduces to the ordinary AWGN channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}. \tag{6}$$

D. Almost-Gaussian Noise Channels

Another special case of (3) is a channel with a random parameter and additive Gaussian noise

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{n} \tag{7}$$

where $\boldsymbol{f}(\cdot, \cdot)$ is a nonrandom function (taking values in $\mathcal{C}^m$) of the input signal $\boldsymbol{x}$ and a random parameter $\boldsymbol{u}$. In this case, it is assumed that $\boldsymbol{x}$, $\boldsymbol{u}$, and $\boldsymbol{n}$ are mutually independent. In particular, if $m = n$ and $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) = \boldsymbol{x} + \boldsymbol{u}$, i.e.,

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{n}. \tag{8}$$

As we let $N_0 \to \infty$, we obtain a channel with almost-Gaussian noise, with contaminating weak noise $\boldsymbol{u}$.

E. Notation

In this paper, we are interested in finding the asymptotic behavior of the mutual information $I(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n})$ up to terms of order $o(N_0^{-2})$, or equivalently, if N_0 is held fixed, the asymptotics of $I(\boldsymbol{x}; \varepsilon\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n})$ as $\varepsilon \to 0$ up to terms of order $o(\varepsilon^4)$. To that end, the following moments will play a crucial role.

Denote the conditional mean vector of $\boldsymbol{g}(\boldsymbol{x})$ by

$$\bar{\boldsymbol{g}}(\boldsymbol{x}) = E[\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x}] \tag{9}$$

and denote by $\operatorname{cov}(\boldsymbol{x})$ and $\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})$ the covariance matrix of $\boldsymbol{x}$ and the conditional covariance matrix of $\boldsymbol{g}(\boldsymbol{x})$ given $\boldsymbol{x}$, respectively, i.e.,[3]

$$\operatorname{cov}(\boldsymbol{x}) = E[(\boldsymbol{x} - E[\boldsymbol{x}])(\boldsymbol{x} - E[\boldsymbol{x}])^\dagger] \tag{10}$$

and

$$\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x}) = E[(\boldsymbol{g}(\boldsymbol{x}) - \bar{\boldsymbol{g}}(\boldsymbol{x}))(\boldsymbol{g}(\boldsymbol{x}) - \bar{\boldsymbol{g}}(\boldsymbol{x}))^\dagger|\boldsymbol{x}]. \tag{11}$$

We will also use the covariance matrix $\operatorname{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x}))$ of the conditional mean vector $\bar{\boldsymbol{g}}(\boldsymbol{x})$ which is defined as

$$\begin{aligned}\operatorname{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x})) &= E[(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\bar{\boldsymbol{g}}(\boldsymbol{x})])(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\bar{\boldsymbol{g}}(\boldsymbol{x})])^\dagger] \\ &= E[(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\boldsymbol{g}(\boldsymbol{x})])(\bar{\boldsymbol{g}}(\boldsymbol{x}) - E[\boldsymbol{g}(\boldsymbol{x})])^\dagger]. \end{aligned} \tag{12}$$

Note also that if

$$E\left[\bar{\boldsymbol{g}}(\boldsymbol{x})\bar{\boldsymbol{g}}^\dagger(\boldsymbol{x})\right] = E\left[\bar{\boldsymbol{g}}(\boldsymbol{x})\right] E\left[\bar{\boldsymbol{g}}^\dagger(\boldsymbol{x})\right] \tag{13}$$

then

$$E[\operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})] = \operatorname{cov}(\boldsymbol{g}(\boldsymbol{x})). \tag{14}$$

Throughout the paper, $\operatorname{trace}(\cdot)$ denotes the trace of the corresponding matrix and the base of $\log(\cdot)$ is the unit in which we measure the mutual information.

[3] $\dagger$ denotes conjugate transposition.

We make frequent use of the Euclidean (or Frobenius) squared norm of a complex matrix, which is equal to the sum of the magnitude squares of all its coefficients. In the special case of a covariance matrix

$$\|\mathrm{cov}(\boldsymbol{z})\|^2 = \mathrm{trace}\{\mathrm{cov}^2(\boldsymbol{z})\}. \tag{15}$$

III. Second-Order Information Measures

Our results suggest the definition of several new measures of the dependence between arbitrary random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ (not necessarily connected through the channels in Section II). Specifically, the measures of dependence depend exclusively on the conditional first and second moments: $E[\boldsymbol{y}|\boldsymbol{x}]$ and $E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]$.

The simplest such measure is

$$\begin{aligned} G(\boldsymbol{x};\boldsymbol{y}) &= E\left[\|E[\boldsymbol{y}|\boldsymbol{x}] - E[\boldsymbol{y}]\|^2\right] & (16)\\ &= E\left[\|E[\boldsymbol{y}|\boldsymbol{x}]\|^2\right] - \|E[\boldsymbol{y}]\|^2 & (17)\\ &= \mathrm{trace}\left\{\mathrm{cov}(E[\boldsymbol{y}|\boldsymbol{x}])\right\} & (18)\end{aligned}$$

which is equal to zero if and only if $E[\boldsymbol{y}|\boldsymbol{x}] = E[\boldsymbol{y}]$ almost surely.

Note that

$$G(\boldsymbol{x};\boldsymbol{x}) = \mathrm{trace}\left\{\mathrm{cov}(\boldsymbol{x})\right\} = \mathrm{var}[\boldsymbol{x}] \tag{19}$$

and if $\boldsymbol{m}$ is independent of $\boldsymbol{x}$ then

$$G(\boldsymbol{x};\boldsymbol{y}) = G(\boldsymbol{x};\boldsymbol{y}+\boldsymbol{m}). \tag{20}$$

It is easy to show that $G(\boldsymbol{x};\boldsymbol{y})$ is equal to the decrease in minimum mean-square error (MMSE) in the estimation of $\boldsymbol{y}$ achievable by observing $\boldsymbol{x}$

$$G(\boldsymbol{x};\boldsymbol{y}) = E\left[\|\boldsymbol{y} - E[\boldsymbol{y}]\|^2\right] - E\left[\|\boldsymbol{y} - E[\boldsymbol{y}|\boldsymbol{x}]\|^2\right]. \tag{21}$$

As we will see in Section IV, this information measure is connected to mutual information (under certain technical conditions (Theorem 1)) through

$$\lim_{N_0\to\infty} N_0 I\left(\boldsymbol{x};\boldsymbol{z}+\sqrt{N_0}\bar{\boldsymbol{n}}\right) = G(\boldsymbol{x};\boldsymbol{z})\log e \tag{22}$$

where the vector $\bar{\boldsymbol{n}}$ is proper-complex Gaussian with independent unit-variance components, independent of $\boldsymbol{x}$ and $\boldsymbol{z}$. Note that (22) relates the mutual information in the low SNR regime to the MMSE of the estimation of the (noiseless) channel output given the input, whereas [2] showed that the derivative of mutual information at zero SNR is given by the MMSE of the estimation of the input given the noisy channel output.[4]

Define

$$K(\boldsymbol{x};\boldsymbol{y}) = E\left[\|E[\boldsymbol{y}|\boldsymbol{x}]\|^4\right] - \|E[\boldsymbol{y}]\|^4. \tag{23}$$

From the convexity of the function $f(\boldsymbol{z}) = \|\boldsymbol{z}\|^4$ it follows that

$$K(\boldsymbol{x};\boldsymbol{y}) \geq 0 \tag{24}$$

with equality if (but not only if) $G(\boldsymbol{x};\boldsymbol{y}) = 0$.

Note that if $\boldsymbol{n}$ has zero mean and is independent of $\boldsymbol{x}$

$$K(\boldsymbol{x};\boldsymbol{y}+\boldsymbol{n}) = K(\boldsymbol{x};\boldsymbol{y}). \tag{25}$$

[4]A result recently generalized to any SNR in [12].

In the case where x and y are jointly Gaussian proper-complex scalars

$$K(\mathsf{x};\mathsf{y}) = 2\rho^4\sigma_{\mathsf{y}}^4 + 4\rho^2\sigma_{\mathsf{y}}^2|\mu_{\mathsf{y}}|^2 \tag{26}$$

where y has mean μ_{y} and variance σ_{y}^2, and ρ is the correlation coefficient between x and y.

If the matrix $\boldsymbol{H}$ has zero mean conditioned on $\boldsymbol{x}$, then

$$K(\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}) = 0. \tag{27}$$

Define now the following measure of dependence of the conditional correlation matrix $E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]$ on $\boldsymbol{x}$:

$$J(\boldsymbol{x};\boldsymbol{y}) = \mathrm{trace}\left\{E\left[E^2[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]\right] - E^2[\boldsymbol{y}\boldsymbol{y}^\dagger]\right\}. \tag{28}$$

Although $I(\boldsymbol{x};\boldsymbol{y}) = 0$ implies $J(\boldsymbol{x};\boldsymbol{y}) = 0$, the converse is not true. In fact, $J(\boldsymbol{x};\boldsymbol{y})$ is a measure of dependence in the sense that

$$J(\boldsymbol{x};\boldsymbol{y}) \geq 0$$

with equality if and only if $E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]$ does not depend on $\boldsymbol{x}$. To check this property note that the convexity of $\|\boldsymbol{M}\boldsymbol{w}\|^2$ in the Hermitian matrix $\boldsymbol{M}$ implies

$$E[\boldsymbol{M}^2] - E^2[\boldsymbol{M}] \geq 0. \tag{29}$$

Furthermore, only when $\boldsymbol{M}$ is deterministic, does the trace of the matrix in the left side of (29) become zero. Letting $\boldsymbol{M} = E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]$ we see that $E\left[E^2[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]\right] - E^2[\boldsymbol{y}\boldsymbol{y}^\dagger]$ is nonnegative definite, with zero trace only if $E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]$ is deterministic.

Note that if $\boldsymbol{n}$ has zero mean and is independent of $\boldsymbol{x}$ and $\boldsymbol{y}$

$$J(\boldsymbol{x};\boldsymbol{y}) = J(\boldsymbol{x};\boldsymbol{y}+\boldsymbol{n}) \tag{30}$$

and if $G(\boldsymbol{x};\boldsymbol{y}) = 0$, then

$$J(\boldsymbol{x};\boldsymbol{y}) = J(\boldsymbol{x};\boldsymbol{y}+\boldsymbol{m}) \tag{31}$$

for any deterministic $\boldsymbol{m}$.

As we will see in Section IV, under the technical conditions in Theorem 1, if $G(\boldsymbol{x};\boldsymbol{y}) = 0$, then

$$\lim_{N_0\to\infty} N_0^2 I\left(\boldsymbol{x};\boldsymbol{z}+\sqrt{N_0}\bar{\boldsymbol{n}}\right) = \frac{1}{2}J(\boldsymbol{x};\boldsymbol{z})\log e \tag{32}$$

where the vector $\bar{\boldsymbol{n}}$ is proper-complex Gaussian with independent unit-variance components, independent of $\boldsymbol{x}$ and $\boldsymbol{z}$. Or, equivalently, we can formulate the same relationship among real-valued vectors in which case the right-hand side of (32) becomes $J(\boldsymbol{x};\boldsymbol{z})\log e$.

Denote now

$$L(\boldsymbol{x};\boldsymbol{y}) = E\left[E[\boldsymbol{y}^\dagger|\boldsymbol{x}]E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]E[\boldsymbol{y}|\boldsymbol{x}]\right] - E[\boldsymbol{y}^\dagger]E[\boldsymbol{y}\boldsymbol{y}^\dagger]E[\boldsymbol{y}]. \tag{33}$$

For example, if x and y are jointly Gaussian proper-complex scalars, then

$$L(\mathsf{x};\mathsf{y}) = \rho^2\sigma_{\mathsf{y}}^2\left(\sigma_{\mathsf{y}}^2(1+\rho^2) + 3|\mu_{\mathsf{y}}|^2\right). \tag{34}$$

Finally, define

$$\begin{aligned}\Delta(\boldsymbol{x};\boldsymbol{y}) &= \mathrm{trace}\left\{E[\mathrm{cov}^2(\boldsymbol{y}|\boldsymbol{x})]\right\} - \mathrm{trace}\left\{\mathrm{cov}^2(\boldsymbol{y})\right\}\\ &= J(\boldsymbol{x};\boldsymbol{y}) + K(\boldsymbol{x};\boldsymbol{y}) - 2L(\boldsymbol{x};\boldsymbol{y}) \end{aligned}\tag{35}$$

where (35) follows from

$$\operatorname{trace}\left\{\operatorname{cov}^2(\boldsymbol{y}|\boldsymbol{x})\right\} = \operatorname{trace}\left\{E^2[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]\right\} + \|E[\boldsymbol{y}|\boldsymbol{x}]\|^4 - 2E[\boldsymbol{y}^\dagger|\boldsymbol{x}]E[\boldsymbol{y}\boldsymbol{y}^\dagger|\boldsymbol{x}]E[\boldsymbol{y}|\boldsymbol{x}]. \quad (36)$$

If $G(\boldsymbol{x};\boldsymbol{y}) = 0$, then $K(\boldsymbol{x};\boldsymbol{y}) = L(\boldsymbol{x};\boldsymbol{y}) = 0$ and

$$\Delta(\boldsymbol{x};\boldsymbol{y}) = J(\boldsymbol{x};\boldsymbol{y}). \quad (37)$$

Furthermore, if $\boldsymbol{y}$ is a deterministic function of $\boldsymbol{x}$, then

$$\Delta(\boldsymbol{x};\boldsymbol{y}) = -\|\operatorname{cov}(\boldsymbol{y})\|^2. \quad (38)$$

Note that all the information measures introduced in this section are invariant to one-to-one transformations of $\boldsymbol{x}$.

IV. Main Results

A. Statements

The following theorem is the main result of the paper.

Theorem 1: Let $\boldsymbol{n}$ be a proper-complex Gaussian random vector with variance N_0. Assume that the probability distribution of the input signal $\boldsymbol{x}$ (possibly dependent on N_0) satisfies the condition

$$P\left\{\|\boldsymbol{x}\| > \delta\right\} \le \exp\left\{-\delta^\nu\right\} \quad (39)$$

for all $\delta > \delta_0$, where $\delta_0 > 0$ and $\nu > 0$ are some positive constants independent on N_0. Assume also that $\boldsymbol{z}$ is a proper complex random vector both unconditionally and conditioned on any input $\boldsymbol{x}_0$. Furthermore, there exist some finite constants $\alpha > 0$, $c > 0$, $k > 0$, and $l > 0$ such that

$$E[\|\boldsymbol{z}\|^{4+\alpha}] < c \quad (40)$$

$$E[\|\boldsymbol{z}\|^{4+\alpha}|\boldsymbol{x}] \le c\|\boldsymbol{x}\|^k \quad \text{a.s.} \quad (41)$$

and

$$\operatorname{trace}\left\{\operatorname{cov}^2(\boldsymbol{z}|\boldsymbol{x})\right\} \le c\|\boldsymbol{x}\|^l \quad \text{a.s.} \quad (42)$$

Then, as $N_0 \to \infty$, uniformly in all $\boldsymbol{x}$ satisfying the above conditions, the following asymptotic expression holds:

$$I(\boldsymbol{x};\boldsymbol{z}+\boldsymbol{n}) = \frac{\log e}{N_0}G(\boldsymbol{x};\boldsymbol{z}) + \frac{\log e}{2N_0^2}\Delta(\boldsymbol{x};\boldsymbol{z}) + o(N_0^{-2}). \quad (43)$$

For fading channels (4), as a special case of channels (3), the asymptotic expansion is given by the following result.

Theorem 2: Assume that the input signal $\boldsymbol{x}$ satisfies condition (39) of Theorem 1, and that $\bar{\boldsymbol{H}}\boldsymbol{x}$ is proper complex where $\bar{\boldsymbol{H}} = E[\boldsymbol{H}]$. Assume that the channel matrix is such that there exists a constant $\alpha > 0$ such that

$$E[\|\boldsymbol{H}\|^{4+\alpha}] < \infty. \quad (44)$$

Furthermore, the real and imaginary parts[5] of the coefficients in $\boldsymbol{H}$ are uncorrelated and have identical covariances, i.e., for all i, j, k, l

$$E[\Re\{\mathsf{h}_{ij}\}\Im\{\mathsf{h}_{kl}\}] = E[\Re\{\mathsf{h}_{ij}\}]E[\Im\{\mathsf{h}_{kl}\}]$$

and

$$E[\Re\{\mathsf{h}_{ij} - \bar{\mathsf{h}}_{ij}\}\Re\{\mathsf{h}_{kl} - \bar{\mathsf{h}}_{kl}\}] = E[\Im\{\mathsf{h}_{ij} - \bar{\mathsf{h}}_{ij}\}\Im\{\mathsf{h}_{kl} - \bar{\mathsf{h}}_{kl}\}]$$

where $\bar{\mathsf{h}}_{ij} = E[\mathsf{h}_{ij}]$. Then, as $N_0 \to \infty$, uniformly in all $\boldsymbol{x}$ satisfying condition (39), we have

$$I(\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}) = \frac{\log e}{N_0}E[\|\bar{\boldsymbol{H}}(\boldsymbol{x} - E[\boldsymbol{x}])\|^2] + \frac{\log e}{2N_0^2}\Delta(\boldsymbol{x},\boldsymbol{H}\boldsymbol{x}) + o(N_0^{-2}). \quad (45)$$

Corollary 1: In particular, if $\bar{\boldsymbol{H}} = 0$, then

$$I(\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}) = \frac{\log e}{2N_0^2}\operatorname{trace}\left\{E\left[\left(E\left[\boldsymbol{H}\boldsymbol{x}\boldsymbol{x}^\dagger\boldsymbol{H}^\dagger|\boldsymbol{x}\right]\right)^2\right] - \left(E\left[\boldsymbol{H}E[\boldsymbol{x}\boldsymbol{x}^\dagger]\boldsymbol{H}^\dagger\right]\right)^2\right\} + o(N_0^{-2}). \quad (46)$$

Corollary 2: Deterministic channel. If $\boldsymbol{H} = \bar{\boldsymbol{H}}$ a.s., then

$$I(\boldsymbol{x};\bar{\boldsymbol{H}}\boldsymbol{x}+\boldsymbol{n}) = \frac{\log e}{N_0}\operatorname{trace}\left\{\bar{\boldsymbol{H}}\operatorname{cov}(\boldsymbol{x})\bar{\boldsymbol{H}}^\dagger\right\} - \frac{\log e}{2N_0^2}\operatorname{trace}\left\{(\bar{\boldsymbol{H}}\operatorname{cov}(\boldsymbol{x})\bar{\boldsymbol{H}}^\dagger)^2\right\} + o(N_0^{-2}). \quad (47)$$

Corollary 3: In the one-dimensional case (where $m = n = 1$) we have

$$I(\mathsf{x};\mathsf{h}\mathsf{x}+\mathsf{n}) = \frac{\log e}{N_0}\operatorname{var}[\mathsf{x}]|\bar{\mathsf{h}}|^2 + \frac{\log e}{2N_0^2}\left[E[|\mathsf{x}|^4]\operatorname{var}^2[\mathsf{h}] - \operatorname{var}^2[\mathsf{h}\mathsf{x}]\right] + o(N_0^{-2}). \quad (48)$$

In particular, if $\bar{\mathsf{h}} = 0$, then

$$I(\mathsf{x};\mathsf{h}\mathsf{x}+\mathsf{n}) = \frac{\log e}{2N_0^2}\left(E[|\mathsf{h}|^2]\right)^2\operatorname{var}[|\mathsf{x}|^2] + o(N_0^{-2}) \quad (49)$$

and for a deterministic channel, we obtain

$$I(\mathsf{x};\bar{\mathsf{h}}\mathsf{x}+\mathsf{n}) = \frac{\log e}{N_0}\operatorname{var}[\mathsf{x}]|\bar{\mathsf{h}}|^2 - \frac{\log e}{2N_0^2}\operatorname{var}^2[\mathsf{x}]|\bar{\mathsf{h}}|^4 + o(N_0^{-2}). \quad (50)$$

Theorem 2 pertains to the case where the receiver does not know the realization of $\boldsymbol{H}$. If it does, then the following result is of interest.

Theorem 3: Assume that the input $\boldsymbol{x}$ is a proper complex random vector and that there exist some finite constants $c > 0$ and $\alpha > 0$ such that

$$E[\|\boldsymbol{x}\|^{4+\alpha}] < c. \quad (51)$$

Assume also that the matrix $\boldsymbol{H}$ satisfies a condition similar to (39), i.e.,

$$P\left\{\|\boldsymbol{H}\| > \delta\right\} \le \exp\left\{-\delta^\nu\right\} \quad (52)$$

for all sufficiently large $\delta > 0$ where $\nu > 0$ is a positive constant. Then, as $N_0 \to \infty$, uniformly in all $\boldsymbol{x}$ satisfying the conditions of the theorem we have

$$I(\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}\,|\boldsymbol{H}) = \frac{\log e}{N_0}\operatorname{trace}\left\{E[\boldsymbol{H}\operatorname{cov}(\boldsymbol{x})\boldsymbol{H}^\dagger]\right\} - \frac{\log e}{2N_0^2}\operatorname{trace}\left\{E[(\boldsymbol{H}\operatorname{cov}(\boldsymbol{x})\boldsymbol{H}^\dagger)^2]\right\} + o(N_0^{-2}). \quad (53)$$

Corollary 4: Under the conditions of Theorem 3, and assuming that the receiver knows the channel but the transmitter

[5]The real and imaginary parts of X are denoted by $\Re\mathsf{X}$ and $\Im\mathsf{X}$, respectively.

has no knowledge of the channel or of its statistics, any proper-complex input distribution with equal-power independent components is second-order optimal.

With receiver knowledge of the channel but lacking knowledge of the channel statistics at the transmitter, the optimum input (for all N_0) consists of independent and identically distributed (i.i.d.) Gaussian complex random variables. The corollary follows immediately upon noting from (53) that the first and second derivatives depend only on the covariance matrix of the input, which is a multiple of the identity in this case. From (53) we can obtain the derivatives obtained in [1] by differentiating the log-determinant capacity formula

$$\dot{C}(0) = \frac{\log e}{2n} \text{trace}\left\{E[\boldsymbol{H}^\dagger \boldsymbol{H}]\right\} \tag{54}$$

and

$$\ddot{C}(0) = -\frac{m \log e}{2n^2} \text{trace}\left\{E[(\boldsymbol{H}^\dagger \boldsymbol{H})^2]\right\}. \tag{55}$$

Corollary 4 was obtained in Theorem 14 [1] by using another reasoning. Note that the simplest proper-complex input distribution is quadrature phase-shift-keying (QPSK).

B. Remarks

Remark 1: Although there is certainly some room for weakening sufficent condition (39), it is definitely not superfluous. According to Theorem 1 in [1], the derivative with respect to SNR of channel capacity at zero SNR is given by the maximum channel gain

$$\mathsf{G} = \sup_{P_{\boldsymbol{x}}} \frac{E[\|\boldsymbol{Hx}\|^2]}{E[\|\boldsymbol{x}\|^2]}. \tag{56}$$

Note that this does not contradict (45) as the asymptotically optimal input distribution for noncoherent channels does not satisfy condition (39). For example, consider the sequence of two-mass distributions proven in [4] to achieve the capacity of the scalar Rayleigh channel when $N_0 \to \infty$. Assuming without loss of generality that $E[\|\boldsymbol{x}\|^2] = 1$, and that $N_0 > \gamma$ for some γ, the optimum input distribution with noise spectral level N_0, $\boldsymbol{x}_{N_0}$, is a two-mass distribution concentrated at points 0 and $\nu(N_0) > 0$ (with sizes $1 - \nu^{-2}$ and ν^{-2}, respectively) and $\nu(N_0) \to \infty$ as $N_0 \to \infty$. Then

$$\sup_{N_0 > \gamma} P[\|\boldsymbol{x}_{N_0}\| > \delta] = \frac{1}{\delta^2}$$

and condition (39) is not satisfied uniformly in all the distributions that are optimum for sufficiently large N_0.

Remark 2: The asymptotic expansions presented in Theorems 1–3 can be applied to optimize (for low SNR) the input distribution over a class which is a proper subset of the energy-constrained inputs. For example, a useful such constraint is a fourth moment or kurtosis constraint. As remarked in [1], under those conditions there is no guarantee that the wideband slope and $\frac{E_b}{N_0}_{\min}$ are given by equations analogous to (2) and (1), where the capacity function therein is replaced by the maximum mutual information over the set of allowed distributions, as that maximal mutual information need not be a concave function of the SNR. In particular, $\frac{E_b}{N_0}_{\min}$ may not be obtained at vanishing SNR.

Remark 3: For the fading channel (4), the mutual information between input and output signal (and therefore its asymptotic expression (45)) depends on the mean of the input signal $\boldsymbol{x}$ when the channel is not known at the receiver. Moreover, the second term of (45) depends, in general, not only on the covariance matrix of $\boldsymbol{x}$ but also on mixed fourth moments of the components of $\boldsymbol{x}$. In contrast, when the channel is known at the receiver, the second-order expansion (53) depends only on the input covariance matrix and not on the input mean. The dependence of the second-order asymptotics on the channel matrix is through its fourth-order mixed moments.

Remark 4: Under the additional assumption that the components of the random vector $\boldsymbol{g}(\boldsymbol{x})$ conditioned on $\boldsymbol{x}$ are jointly Gaussian, it is shown in [1], that the following upper bound for the mutual information holds:

$$I(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}) \leq \frac{\log e}{N_0} \text{trace}\left\{\text{cov}(\overline{\boldsymbol{g}}(\boldsymbol{x}))\right\} + \frac{\log e}{2N_0^2} \text{trace}\left\{E[\text{cov}^2(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})]\right\}. \tag{57}$$

Note that this bound holds for all N_0. But, comparing (57) and (43), we can observe that, in general, the upper bound (57) is not tight asymptotically, although the leading terms on the right-hand sides of (57) and (43) coincide if $N_0 \to \infty$.

Remark 5: The real-valued version of the AWGN channel (6) was earlier investigated in [7] where for any integer s an asymptotic expression for the mutual information $I(\boldsymbol{x}; \boldsymbol{x} + \boldsymbol{n})$ up to terms of order $o(N_0^{-s})$, $N_0 \to \infty$, was obtained under the condition that $E[\|\boldsymbol{x}\|^{s+\alpha}] < \infty$ for some $\alpha > 0$.

Remark 6: In the case where conditions of proper complexity of $\boldsymbol{g}(\boldsymbol{x})$ and/or $\boldsymbol{g}(\boldsymbol{x}_0)$ given $\boldsymbol{x}_0$ are not fulfilled but the other conditions of Theorem 1 are satisfied, the asymptotic expression (43) remains true if we replace $\boldsymbol{x}$ and $\boldsymbol{g}(\boldsymbol{x})$ on the right-hand side of (43) and in condition (42) by real-valued random vectors of lengths $2n$ and $2m$ whose components are real and imaginary parts of the complex-valued components of vectors $\boldsymbol{x}$ and $\boldsymbol{g}(\boldsymbol{x})$, respectively. This assertion follows from the proof of Theorem 1 (see (100)). Similar remarks apply to Theorems 2 and 3.

C. Examples

Example 1 (Ricean Channel):

For the Ricean block-fading channel, the channel matrix $\boldsymbol{H}$ in (4) can be represented in the form

$$\boldsymbol{H} = (\bar{\mathsf{h}} + \mathsf{g})\boldsymbol{I} \tag{58}$$

where $\bar{\mathsf{h}}$ is deterministic and g is zero-mean proper complex Gaussian with variance γ^2 independent of $\boldsymbol{x}$ and $\boldsymbol{n}$. It can easily be verified that the conditions on the channel required in Theorem 2 are satisfied in this case.

By direct calculations one can easily check that the general formula (45) reduces to

$$\begin{aligned} I(\boldsymbol{x}; \boldsymbol{Hx} + \boldsymbol{n}) &= \frac{\log e}{N_0} |\bar{\mathsf{h}}|^2 \text{var}[\boldsymbol{x}] + \frac{\log e}{2N_0^2} \\ &\quad \cdot \left[\gamma^4 E[\|\boldsymbol{x}\|^4] - \text{trace}\left\{\left[\gamma^2 E[\boldsymbol{x}\boldsymbol{x}^\dagger] + |\bar{\mathsf{h}}|^2 \text{cov}(\boldsymbol{x})\right]^2\right\}\right] \\ &\quad + o(N_0^{-2}). \end{aligned} \tag{59}$$

In particular, if $E[\boldsymbol{x}] = 0$, then

$$I(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) = \frac{\log e}{N_0} |\bar{\mathsf{h}}|^2 \mathrm{var}[\boldsymbol{x}] + \frac{\log e}{2N_0^2} \cdot \left[\gamma^4 E[\|\boldsymbol{x}\|^4] - (|\bar{\mathsf{h}}|^2 + \gamma^2)^2 \sum_{i,j} |\mathrm{cov}(\mathsf{x}_i, \mathsf{x}_j)|^2 \right] + o(N_0^{-2}) \tag{60}$$

and if $\bar{\mathsf{h}} = 0$, then

$$I(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) = \frac{\log e}{2N_0^2} \gamma^4 \left[\sum_{i,j} \mathrm{var}(\mathsf{x}_i \mathsf{x}_j^*) \right] + o(N_0^{-2}). \tag{61}$$

In fact, none of the asymptotic expansions given in this example depends on g being Gaussian; they are guaranteed to hold for any proper-complex zero-mean random variable with $E[|\mathsf{g}|^{4+\alpha}] < \infty$]. The result in (60) has been used in [13], [14] to obtain the minimum energy per bit and the wideband slope for Ricean channels under various input constraints.

The particular case $\bar{\mathsf{h}} = 0$, $\mathsf{g} = \sqrt{\gamma} e^{j\theta}$ with θ uniformly distributed on $[0, 2\pi)$ is particularly interesting as it models the noncoherent channel without fading [15]. In this case, (59) reduces to

$$I(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) = \frac{\gamma^4 \log e}{2N_0^2} \left[E[\|\boldsymbol{x}\|^4] - \mathrm{trace}\left\{ \left[E[\boldsymbol{x}\boldsymbol{x}^\dagger] \right]^2 \right\} \right] + o(N_0^{-2}). \tag{62}$$

Maximizing the second-order term in (62) with respect to the input distribution assuming a constraint on the power $E[\|\boldsymbol{x}\|^2] = 1$ and on the kurtosis of each component

$$E[|\mathsf{x}_i|^4] \le \kappa E^2[|\mathsf{x}_i|^2]$$

we obtain that it is best to concentrate the power on one component using the maximum allowable peakiness, yielding

$$E[\|\boldsymbol{x}\|^4] - \mathrm{trace}\left\{ \left[E[\boldsymbol{x}\boldsymbol{x}^\dagger] \right]^2 \right\} = \kappa - 1. \tag{63}$$

Example 2 (A Channel with Almost-Gaussian Noise): Consider the channel (8) with almost-Gaussian noise. In this case, $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{u}$ and it is easy to see that the general formula (43) has the form

$$I(\boldsymbol{x}; \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{n}) = \frac{\log e}{N_0} \mathrm{var}[\boldsymbol{x}] - \frac{\log e}{2N_0^2} \cdot \left[\mathrm{trace}\left\{ \mathrm{cov}^2(\boldsymbol{x}) + 2\mathrm{cov}(\boldsymbol{x})\mathrm{cov}(\boldsymbol{u}) \right\} \right] + o(N_0^{-2}). \tag{64}$$

It should be noted that this formula holds for any proper input signal $\boldsymbol{x}$ and proper contaminating noise $\boldsymbol{u}$ such that

$$E[\|\boldsymbol{x}\|^{4+\alpha}] < \infty, \qquad E[\|\boldsymbol{u}\|^{4+\alpha}] < \infty$$

for a constant $\alpha > 0$. The proof of this assertion easily follows from Theorem 5 (see Section V).

Example 3 (Multiantenna Block-Fading Channel): In the Marzetta–Hochwald channel model [16] with M transmit antennas and N receive antennas, the MN entries of the propagation matrix are assumed to be independent proper complex with zero mean and variance ζ^2, such that time is slotted in blocks of T symbols, and in each block the propagation matrix does not change, and takes independent instantiations in different blocks. It is natural to model inputs to a memoryless channel as $M \times T$ matrices and the corresponding outputs as $N \times T$ matrices. Alternatively, we can stack the T columns of those matrices and obtain a representation that is a special case of the fading channel model in Section II-B. The resulting channel matrix $\boldsymbol{H}$ is $NT \times MT$, block diagonal with each block equal to the same $N \times M$ matrix $\boldsymbol{G}$ of propagation coefficients. The signaling scheme is characterized by the joint distribution of the transmitted vectors across the M antennas at times $t = 1, \ldots, T : \{\boldsymbol{x}_1 \cdots \boldsymbol{x}_T\}$. Denoting by $\boldsymbol{x}$ the stack of those T M-vectors, and using the independence of the propagation coefficients affecting every different pair of transmit–receive antennas, we have that the desired covariance matrices (Corollary 1) are $NT \times NT$ block matrices

$$E\left[\boldsymbol{H}\boldsymbol{x}\boldsymbol{x}^\dagger \boldsymbol{H}^\dagger | \boldsymbol{x} \right] = \zeta^2 \begin{bmatrix} \|\boldsymbol{x}_1\|^2 \boldsymbol{I} & \ldots & \boldsymbol{x}_1^\dagger \boldsymbol{x}_T \boldsymbol{I} \\ \vdots & \ddots & \vdots \\ \boldsymbol{x}_T^\dagger \boldsymbol{x}_1 \boldsymbol{I} & \ldots & \|\boldsymbol{x}_T\|^2 \boldsymbol{I} \end{bmatrix} \tag{65}$$

$$E\left[\boldsymbol{H} E[\boldsymbol{x}\boldsymbol{x}^\dagger] \boldsymbol{H}^\dagger \right] = \zeta^2 \begin{bmatrix} E[\|\boldsymbol{x}_1\|^2] \boldsymbol{I} & \ldots & E[\boldsymbol{x}_1^\dagger \boldsymbol{x}_T] \boldsymbol{I} \\ \vdots & \ddots & \vdots \\ E[\boldsymbol{x}_T^\dagger \boldsymbol{x}_1] \boldsymbol{I} & \ldots & E[\|\boldsymbol{x}_T\|^2] \boldsymbol{I} \end{bmatrix}. \tag{66}$$

Subtracting the traces of the squares of (65) and (66), we obtain that the mutual information (over a block of T symbols) for a receiver which has no side information of the channel satisfies

$$I(\boldsymbol{x}; \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}) = \frac{\zeta^4 N \log e}{2N_0^2} \sum_{j=1}^{T} \sum_{k=1}^{T} E\left[\left(\boldsymbol{x}_j^\dagger \boldsymbol{x}_k \right)^2 \right] - \left| E\left[\left(\boldsymbol{x}_j^\dagger \boldsymbol{x}_k \right) \right] \right|^2 + o\left(N_0^{-2} \right) \tag{67}$$

an expansion previously found in [17], [18].

D. NonGaussian Noise

Consider the additive noise channel

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} \tag{68}$$

where now we assume that the noise components of $\boldsymbol{n}$ are i.i.d. zero-mean random variables (not necessarily Gaussian) with independent real and imaginary parts each with variance $N_0/2$.

Whereas the non-Gaussian nature of the noise increases the first-order derivative (and therefore reduces the $\frac{E_b}{N_0}_{\min}$) it may reduce the wideband slope. For example, in the case in which the components of the noise follow the Laplacian density function

$$p(\mathsf{x}) = \frac{1}{N_0} \exp\left(-\frac{2}{\sqrt{N_0}} (|\Re \mathsf{x}| + |\Im \mathsf{x}|) \right) \tag{69}$$

it is shown in [1] that the first-order derivative is twice that obtained in the Gaussian case, whereas the second-order derivative is $-\infty$ (and therefore $\mathcal{S}_0 = 0$) as opposed to 2 b/s/Hz/(3 dB) obtained in the Gaussian case.

In this subsection, we investigate sufficient conditions on the noise distribution under which the wideband slope is zero, $\mathcal{S}_0 = 0$.

The proof of the equality $\mathcal{S}_0 = 0$ for the Laplacian noise channel given in [1] uses very specific properties of the divergence $D(P_{Y|X=\boldsymbol{x}} \,\|\, P_{Y|X=0})$ which can be explicitly calculated in that case. Here, $P_{Y|X=\boldsymbol{x}}(\cdot)$ stands for the distribution of

$\boldsymbol{y} = \boldsymbol{x}+\boldsymbol{n}$ conditioned on $\boldsymbol{x}$ and, in particular, $P_{Y|X=0}(\cdot)$ stands for the distribution of $\boldsymbol{n}$. The following result gives a class of distributions (which includes the Laplacian law) under which the wideband slope is zero.

Theorem 4: Suppose that the noise distribution is such that the following conditions are satisfied.

i) There exists a finite limit

$$\lim_{\boldsymbol{x}_0\to 0} \frac{D(P_{Y|X=\boldsymbol{x}_0} \,\|\, P_{Y|X=0})}{\|\boldsymbol{x}_0\|^2} = a, \qquad o < a < \infty. \quad (70)$$

ii) There exist some constants $b > 0$, $0 < h < a$, $M > 0$, and $0 < \beta < 2$ such that

$$D(P_{Y|X=\boldsymbol{x}_0} \,\|\, P_{Y|X=0}) \le \left(a\|\boldsymbol{x}_0\|^2 - b\|\boldsymbol{x}_0\|^{2+\beta}\right) \times 1\left\{\|\boldsymbol{x}_0\| \le M\right\} + h\|\boldsymbol{x}_0\|^2 1\left\{\|\boldsymbol{x}_0\| > M\right\}. \quad (71)$$

Then

$$\mathcal{S}_0 = 0. \quad (72)$$

The following result is proved in the Appendix.

Lemma 1: The Laplacian noise channel satisfies conditions i) and ii) of Theorem 4 with $m = 1$.

Remark 7: It should be noted that conditions i) and ii) imply the equality

$$\sup_{\boldsymbol{x}_0} \frac{D(P_{Y|X=\boldsymbol{x}_0} \,\|\, P_{Y|X=0})}{\|\boldsymbol{x}_0\|^2} = \lim_{\boldsymbol{x}_0\to 0} \frac{D(P_{Y|X=\boldsymbol{x}_0} \,\|\, P_{Y|X=0})}{\|\boldsymbol{x}_0\|^2} = a \quad (73)$$

and, hence, $\dot{C}(0) = aN_0$ since it is well known that

$$\dot{C}(0) = N_0 \sup_{\boldsymbol{x}_0} \frac{D(P_{Y|X=\boldsymbol{x}_0} \,\|\, P_{Y|X=0})}{\|\boldsymbol{x}_0\|^2}. \quad (74)$$

V. PROOFS

Proof of Theorem 1: We will make use of the following result for the real-valued AWGN channel.

Theorem 5: Let $\boldsymbol{x}_\varepsilon = (\mathsf{x}_{1\varepsilon}, \dots, \mathsf{x}_{l\varepsilon})$ and $\boldsymbol{w} = (\mathsf{w}_1, \dots, \mathsf{w}_l)$ be l-dimensional independent real-valued random vectors for any given $\varepsilon > 0$, where the components of $\boldsymbol{w}$ are i.i.d. Gaussian random variables with parameters $(0, \sigma^2)$. If there exist some constants $\alpha > 0$ and $\mu > 0$ such that

$$E[\|\boldsymbol{x}_\varepsilon\|^{4+\alpha}] \le \left(\log \frac{1}{\varepsilon}\right)^\mu \quad (75)$$

then

$$\begin{aligned} I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}) &= I(\varepsilon\widehat{\boldsymbol{x}}_\varepsilon; \varepsilon\widehat{\boldsymbol{x}}_\varepsilon + \boldsymbol{w}) + o(\varepsilon^4) \\ &= \frac{\log e}{2\sigma^2}\mathrm{trace}\left\{\mathrm{cov}(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^2 \\ &\quad - \frac{\log e}{4\sigma^4}\mathrm{trace}\left\{\mathrm{cov}^2(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^4 + o(\varepsilon^4), \\ &\qquad \varepsilon \to 0 \end{aligned} \quad (76)$$

where $\widehat{\boldsymbol{x}}_\varepsilon$ is a Gaussian random vector with the same covariance matrix as $\boldsymbol{x}_\varepsilon$ and $\frac{o(\varepsilon^4)}{\varepsilon^4} \to 0$ as $\varepsilon \to 0$ uniformly with respect to all random vectors $\boldsymbol{x}_\varepsilon$ satisfying (75).

Theorem 5 can be viewed as a generalization of Lemma 2.2 in [7] to the case when we are interested in finding an asymptotic expansion for the mutual information up to terms of order $o(\varepsilon^4)$, $\varepsilon \to 0$. We present the proof of Theorem 5 in the Appendix.

Throughout this proof, we represent the noiseless channel output by $\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{x})$ where $\boldsymbol{g}(\cdot)$ is a possibly random function. In order to apply the statement of Theorem 5 to the proof of Theorem 1, we first reduce our complex-valued problem to the real-valued one. By definition we have

$$I(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}) = I\left(\boldsymbol{x}^{(r)}; \boldsymbol{g}^{(r)}(\boldsymbol{x}) + \boldsymbol{n}^{(r)}\right) \quad (77)$$

where

$$\boldsymbol{x}^{(r)} = \begin{bmatrix} \Re\boldsymbol{x} \\ \Im\boldsymbol{x} \end{bmatrix}$$

denotes the real-valued vector of length $2n$ generated by real and imaginary parts of $\boldsymbol{x}$. The real-valued vectors

$$\boldsymbol{g}^{(r)}(\boldsymbol{x}) = \begin{bmatrix} \Re\boldsymbol{g}(\boldsymbol{x}) \\ \Im\boldsymbol{g}(\boldsymbol{x}) \end{bmatrix}$$

and

$$\boldsymbol{n}^{(r)} = \begin{bmatrix} \Re\boldsymbol{n} \\ \Im\boldsymbol{n} \end{bmatrix}$$

of length $2m$ are similarly defined. Therefore, it is clear that

$$I(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}) = I\left(\boldsymbol{g}^{(r)}(\boldsymbol{x}); \boldsymbol{g}^{(r)}(\boldsymbol{x}) + \boldsymbol{n}^{(r)}\right) - I\left(\boldsymbol{g}^{(r)}(\boldsymbol{x}); \boldsymbol{g}^{(r)}(\boldsymbol{x}) + \boldsymbol{n}^{(r)}|\boldsymbol{x}^{(r)}\right). \quad (78)$$

Let

$$\boldsymbol{z}(\boldsymbol{x}^{(r)}) \stackrel{\mathrm{def}}{=} \boldsymbol{g}^{(r)}(\boldsymbol{x}), \quad \boldsymbol{w} \stackrel{\mathrm{def}}{=} \varepsilon\boldsymbol{n}^{(r)} \quad (79)$$

where

$$\varepsilon \stackrel{\mathrm{def}}{=} \sqrt{\frac{2}{N_0}}. \quad (80)$$

Then we can rewrite (78) as

$$I(\boldsymbol{x}; \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}) = I\left(\varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}) + \boldsymbol{w}\right) - I\left(\varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}) + \boldsymbol{w}|\boldsymbol{x}^{(r)}\right) \quad (81)$$

where $\boldsymbol{w}$ is a real-valued Gaussian random vector with independent components each with zero mean and variance 1.

Condition (40) implies that

$$E[\|\boldsymbol{z}(\boldsymbol{x}^{(r)})\|^{4+\alpha}] \le c < \infty$$

and therefore we can apply Theorem 5 to obtain the following asymptotic expression for the mutual information $I(\varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}) + \boldsymbol{w})$:

$$\begin{aligned} I\left(\varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}) + \boldsymbol{w}\right) &= \frac{\log e}{2}\mathrm{trace}\left\{\mathrm{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\}\varepsilon^2 \\ &\quad - \frac{\log e}{4}\mathrm{trace}\left\{\mathrm{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\} \\ &\quad \times \varepsilon^4 + o(\varepsilon^4), \qquad \varepsilon \to 0. \end{aligned} \quad (82)$$

Now note that we cannot directly apply Theorem 5 to write out the asymptotics for the second term of the right-hand side of (78) because, in general, condition (75) is not uniformly satisfied for all values of $\boldsymbol{x}^{(r)}$. Therefore, to obtain the asymptotic expression for the second term of (78), let us represent $I(\boldsymbol{g}(\boldsymbol{x}); \varepsilon\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{n}|\boldsymbol{x})$ as

$$I\left(\varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}^{(r)}) + \boldsymbol{w}|\boldsymbol{x}^{(r)}\right) = I_1 + I_2 \tag{83}$$

where

$$I_1 = \int_{A_\varepsilon} I\left(\varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right); \varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right) + \boldsymbol{w}\right) P_X\left(d\boldsymbol{x}_0^{(r)}\right) \tag{84}$$

and

$$I_2 = \int_{\overline{A}_\varepsilon} I\left(\varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right); \varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right) + \boldsymbol{w}\right) P_X\left(d\boldsymbol{x}_0^{(r)}\right). \tag{85}$$

Here, $P_X(\cdot)$ denotes the probability distribution of $\boldsymbol{x}^{(r)}$

$$A_\varepsilon = \left\{\boldsymbol{x}_0^{(r)} : \left\|\boldsymbol{x}_0^{(r)}\right\| \le \log^s(1/\varepsilon)\right\}$$
$$\overline{A}_\varepsilon = \left\{\boldsymbol{x}_0^{(r)} : \left\|\boldsymbol{x}_0^{(r)}\right\| > \log^s(1/\varepsilon)\right\}$$

and s is a positive constant that will be chosen later.

It follows from condition (41) that for almost all $\boldsymbol{x}_0^{(r)} \in A_\varepsilon$ we have

$$E\left[\left\|\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right\|^{4+\alpha}\right] = E\left[\|\boldsymbol{g}(\boldsymbol{x}_0)\|^{4+\alpha}\right] \le c\|\boldsymbol{x}_0\|^k = c\left\|\boldsymbol{x}_0^{(r)}\right\|^k \le c\log^{ks}(1/\varepsilon).$$

Hence, we can apply Theorem 5 to write I_1 in the form

$$I_1 = \int_{A_\varepsilon}\left[\frac{\log e}{2}\text{trace}\left\{\text{cov}\left(\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right)\right\}\varepsilon^2 - \frac{\log e}{4} \times\text{trace}\left\{\text{cov}^2\left(\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right)\right\}\varepsilon^4 + o(\varepsilon^4)\right] P_X\left(d\boldsymbol{x}_0^{(r)}\right) \tag{86}$$

where $o(\varepsilon^4)/\varepsilon^4 \to 0$ as $\varepsilon \to 0$ uniformly in all $\boldsymbol{x}_0^{(r)} \in A_\varepsilon$. Using (86) we intend to show that

$$I_1 = \frac{\log e}{2}\text{trace}\left\{E[\text{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\}\varepsilon^2 - \frac{\log e}{4}\text{trace}\left\{E[\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\}\varepsilon^4 + o(\varepsilon^4), \qquad \varepsilon \to 0. \tag{87}$$

To this end it suffices to show that

$$\overline{I}_{11} \stackrel{\text{def}}{=} \int_{\overline{A}_\varepsilon} \text{trace}\left\{\text{cov}(\boldsymbol{z}(\boldsymbol{x}_0^{(r)}))\right\} P_X\left(d\boldsymbol{x}_0^{(r)}\right) = o(\varepsilon^2), \qquad \varepsilon \to 0 \tag{88}$$

and

$$\overline{I}_{12} \stackrel{\text{def}}{=} \int_{\overline{A}_\varepsilon} \text{trace}\left\{\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}_0^{(r)}))\right\} P_X\left(d\boldsymbol{x}_0^{(r)}\right) = o(1), \qquad \varepsilon \to 0. \tag{89}$$

First observe that

$$\text{trace}\left\{\text{cov}(\boldsymbol{z}(\boldsymbol{x}_0^{(r)}))\right\} = \text{trace}\left\{\text{cov}(\boldsymbol{g}(\boldsymbol{x}_0))\right\} = E\left[\|\boldsymbol{g}(\boldsymbol{x}_0) - E[\boldsymbol{g}(\boldsymbol{x}_0)]\|^2\right] \tag{90}$$

and therefore condition (41) implies that[6]

$$\text{trace}\left\{\text{cov}\left(\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right)\right\} \le E[\|\boldsymbol{g}(\boldsymbol{x}_0)\|^2] \le \left(E[\|\boldsymbol{g}(\boldsymbol{x}_0)\|^{4+\alpha}]\right)^{\frac{2}{4+\alpha}} \le c_1\|\boldsymbol{x}_0\|^{k_1} = c_1\left\|\boldsymbol{x}_0^{(r)}\right\|^{k_1} \tag{91}$$

for almost all $\boldsymbol{x}_0^{(r)} \in \overline{A}_\varepsilon$, where $k_1 = 2k/(4+\alpha)$. Hence,

$$\overline{I}_{11} \le c_1\int_{\overline{A}_\varepsilon}\left\|\boldsymbol{x}_0^{(r)}\right\|^{k_1} P_X\left(d\boldsymbol{x}_0^{(r)}\right). \tag{92}$$

Furthermore, it is not difficult to check by direct calculation that

$$\text{trace}\left\{\text{cov}^2\left(\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right)\right\} = \frac{1}{2}\text{trace}\left\{\text{cov}^2(\boldsymbol{g}(\boldsymbol{x}_0))\right\} \tag{93}$$

if $\boldsymbol{g}(\boldsymbol{x}_0)$ is a proper complex random vector. Therefore, it follows from (93), condition (42), and equality (89) that

$$\overline{I}_{12} \le c_2\int_{\overline{A}_\varepsilon}\left\|\boldsymbol{x}_0^{(r)}\right\|^{l} P_X\left(d\boldsymbol{x}_0^{(r)}\right). \tag{94}$$

It is well known that condition (39) implies

$$\int_{\left\|\boldsymbol{x}_0^{(r)}\right\|>\delta}\left\|\boldsymbol{x}_0^{(r)}\right\|^{t} P_X\left(d\boldsymbol{x}_0^{(r)}\right) \le \delta^M\exp\left\{-\delta^\nu\right\} \tag{95}$$

for any $t \ge 0$ (where $M > 0$ is a constant depending on t) and for all sufficiently large δ. Using (95) we can easily derive from (92) and (94) that $\overline{I}_{11} = o(\varepsilon^2)$ and $\overline{I}_{12} = o(1)$, $\varepsilon \to 0$, if s is sufficiently large (in fact, the relation $\overline{I}_{12} = o(1)$, $\varepsilon \to 0$, immediately follows even from the fact that the integral $\int\|\boldsymbol{x}_0^{(r)}\|^l P_X(d\boldsymbol{x}_0^{(r)})$ converges). Thus, equalities (88) and (89) and, therefore, (87) are proved.

Now, let us show that

$$I_2 = o(\varepsilon^4), \qquad \varepsilon \to 0 \tag{96}$$

where I_2 is defined in (85). Note that for $\boldsymbol{x}_0^{(r)} \in \overline{A}_\varepsilon$, we cannot apply the statement of Theorem 5 to write out the asymptotic expression for $I(\varepsilon\boldsymbol{z}(\boldsymbol{x}_0^{(r)}); \varepsilon\boldsymbol{z}(\boldsymbol{x}_0^{(r)}) + \boldsymbol{w})$ in form of (82). But we can easily estimate I_2 taking into account that

$$I\left(\varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right); \varepsilon\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right) + \boldsymbol{w}\right) \le I\left(\varepsilon\widehat{\boldsymbol{z}}\left(\boldsymbol{x}_0^{(r)}\right); \varepsilon\widehat{\boldsymbol{z}}\left(\boldsymbol{x}_0^{(r)}\right) + \boldsymbol{w}\right) \tag{97}$$

where $\widehat{\boldsymbol{z}}(\boldsymbol{x}_0^{(r)})$ is a Gaussian random vector having the same covariance function as $\boldsymbol{z}(\boldsymbol{x}_0^{(r)})$. Since

$$I\left(\varepsilon\widehat{\boldsymbol{z}}\left(\boldsymbol{x}_0^{(r)}\right); \varepsilon\widehat{\boldsymbol{z}}\left(\boldsymbol{x}_0^{(r)}\right) + \boldsymbol{w}\right) = \frac{1}{2}\sum_{i=1}^{2m}\log\left(1 + \lambda_i^0\varepsilon^2\right) \le \frac{\log e}{2}\text{trace}\left\{\text{cov}\left(\boldsymbol{z}\left(\boldsymbol{x}_0^{(r)}\right)\right)\right\}\varepsilon^2$$

(where $\lambda_i^0, i = 1, \dots, 2m$, are eigenvalues of $\text{cov}(\boldsymbol{z}(\boldsymbol{x}_0^{(r)}))$), the required equality (96) follows from (97) and the already proved equality (88).

[6]Hereafter, $c, c_1, \dots$ denote finite constants perhaps different in different inequalities.

Thus, from (78), (82), (83), (87), and (96) we obtain

$$I(\boldsymbol{x};\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{n}) = \frac{\log e}{2}\Big[\text{trace}\left\{\text{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\} - \text{trace}\left\{E[\text{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\}\Big]\varepsilon^2 + \frac{\log e}{4}\Big[\text{trace}\left\{E[\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\} - \text{trace}\left\{\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\}\Big]\varepsilon^4 + o(\varepsilon^4), \qquad \varepsilon\to 0. \tag{98}$$

It can easily be seen that

$$\text{trace}\left\{\text{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\} - \text{trace}\left\{E[\text{cov}(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\} = \text{trace}\left\{\text{cov}(\bar{\boldsymbol{z}}(\boldsymbol{x}^{(r)}))\right\} \tag{99}$$

where

$$\bar{\boldsymbol{z}}(\boldsymbol{x}^{(r)}) \stackrel{\text{def}}{=} E[\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)}].$$

Hence,

$$I(\boldsymbol{x};\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{n}) = \frac{\log e}{2}\Big[\text{trace}\left\{\text{cov}(\bar{\boldsymbol{z}}(\boldsymbol{x}^{(r)}))\right\}\Big]\varepsilon^2 + \frac{\log e}{4}\Big[\text{trace}\left\{E[\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\} - \text{trace}\left\{\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\}\Big]\varepsilon^4 + o(\varepsilon^4), \qquad \varepsilon\to 0. \tag{100}$$

To complete the proof of Theorem 1, let us observe that the equality

$$\text{trace}\left\{\text{cov}(\bar{\boldsymbol{z}}(\boldsymbol{x}^{(r)}))\right\} = \text{trace}\left\{\text{cov}(\bar{\boldsymbol{g}}(\boldsymbol{x}))\right\} \tag{101}$$

is always true even without the assumption that $\boldsymbol{g}(\boldsymbol{x})$ is proper. Moreover, by direct calculations one can verify that the equalities

$$\text{trace}\left\{E[\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)})|\boldsymbol{x}^{(r)})]\right\} = \frac{1}{2}\text{trace}\left\{E[\text{cov}^2(\boldsymbol{g}(\boldsymbol{x})|\boldsymbol{x})]\right\} \tag{102}$$

and

$$\text{trace}\left\{\text{cov}^2(\boldsymbol{z}(\boldsymbol{x}^{(r)}))\right\} = \frac{1}{2}\text{trace}\left\{\text{cov}^2(\boldsymbol{g}(\boldsymbol{x}))\right\} \tag{103}$$

are valid if $\boldsymbol{g}(\boldsymbol{x}_0)$ is proper for any fixed $\boldsymbol{x}_0$ and $\boldsymbol{g}(\boldsymbol{x})$ is proper, respectively (inequality (102) was already mentioned above (see (93)). Substituting $\varepsilon^2 = 2/N_0$ into (100) and taking into account equalities (101)–(103), we obtain the required equality (43). □

Proof of Theorem 2: The conditions that $\bar{\boldsymbol{H}}\boldsymbol{x}$ is proper complex and that the real and imaginary parts of the coefficients in $\boldsymbol{H}$ are uncorrelated and have identical covariances guarantee that the random vector $\boldsymbol{H}\boldsymbol{x}$ is proper, and that the random vector $\boldsymbol{H}x_0$ is proper for all $\boldsymbol{x}_0$, as required by Theorem 1. To see the latter condition note that the i, j entry of the matrix $E[(\boldsymbol{H}-\bar{\boldsymbol{H}})\boldsymbol{x}_0\boldsymbol{x}_0^T(\boldsymbol{H}-\bar{\boldsymbol{H}})^T]$ is a weighted sum of terms of the form

$$E[(\boldsymbol{H}-\bar{\boldsymbol{H}})_{ik}(\boldsymbol{H}-\bar{\boldsymbol{H}})_{jn}] = 0.$$

To see that $\boldsymbol{H}\boldsymbol{x}$ is proper complex note that

$$E\left[(\boldsymbol{H}\boldsymbol{x}-\bar{\boldsymbol{H}}\bar{\boldsymbol{x}})(\boldsymbol{H}\boldsymbol{x}-\bar{\boldsymbol{H}}\bar{\boldsymbol{x}})^T\right] = \bar{\boldsymbol{H}}E\left[(\boldsymbol{x}-\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})^T\right]\bar{\boldsymbol{H}}^T + E[(\boldsymbol{H}-\bar{\boldsymbol{H}})\boldsymbol{x}\boldsymbol{x}^T(\boldsymbol{H}-\bar{\boldsymbol{H}})^T]$$

where the first term in the right-hand side is zero by assumption, and the second term is also zero because of the properness of $\boldsymbol{H}\boldsymbol{x}_0$ for arbitrary fixed $\boldsymbol{x}_0$.

In addition, we need to show that for fading channels (4) conditions (44) and (39) imply conditions (40)–(42).

Indeed, first note that, as follows from condition (39), all the moments of $\boldsymbol{x}$ exist and are finite. Therefore, by independence of $\boldsymbol{H}$ and $\boldsymbol{x}$ and by condition (44) we easily obtain

$$E\left[||\boldsymbol{H}\boldsymbol{x}||^{4+\alpha}\right] \le cE\left[||\boldsymbol{H}||^{4+\alpha}\right]E\left[||\boldsymbol{x}||^{4+\alpha}\right] < c_1$$

and hence condition (40) holds. Moreover, it is also follows from the inequality above that

$$E[||\boldsymbol{H}\boldsymbol{x}||^{4+\alpha}|\boldsymbol{x}] \le c||\boldsymbol{x}||^{4+\alpha} \quad \text{a.s.}$$

i.e., condition (41) is also satisfied.

Finally, one can easily verify that

$$\text{trace}\left\{\text{cov}^2(\boldsymbol{H}\boldsymbol{x}|\boldsymbol{x})\right\} \le c||\boldsymbol{x}||^4 \quad \text{a.s.}$$

and, therefore, condition (42) is fulfilled. To complete the proof of the theorem we only need to note that the equality

$$\text{trace}\left\{\text{cov}(E[\boldsymbol{H}\boldsymbol{x}|\boldsymbol{x}])\right\} = E[||\bar{\boldsymbol{H}}(\boldsymbol{x}-E[\boldsymbol{x}])||^2]$$

easily follows from definitions (9)–(11). □

Proof of Theorem 3: First of all, we have

$$I(\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}|\boldsymbol{H}) = I(\boldsymbol{H}\boldsymbol{x};\boldsymbol{H}\boldsymbol{x}+\boldsymbol{n}|\boldsymbol{H}) = I\left(\varepsilon(\boldsymbol{H}\boldsymbol{x})^{(r)};\varepsilon(\boldsymbol{H}\boldsymbol{x})^{(r)}+\boldsymbol{w}|\boldsymbol{H}^{(r)}\right) \tag{104}$$

where the quantities

$$(\boldsymbol{H}\boldsymbol{x})^{(r)} = \begin{bmatrix}\Re(\boldsymbol{H}\boldsymbol{x})\\ \Im(\boldsymbol{H}\boldsymbol{x})\end{bmatrix}, \quad \boldsymbol{H}^{(r)} = \begin{bmatrix}\Re\boldsymbol{H}\\ \Im\boldsymbol{H}\end{bmatrix}$$

$$\boldsymbol{w} = \varepsilon\boldsymbol{n}^{(r)} = \varepsilon\begin{bmatrix}\Re\boldsymbol{n}\\ \Im\boldsymbol{n}\end{bmatrix}, \quad \varepsilon = \sqrt{\frac{2}{N_0}} \tag{105}$$

are defined similar to $\boldsymbol{x}^{(r)}, \boldsymbol{g}^{(r)}(\boldsymbol{x})$, etc., in the proof of Theorem 1 (cf. (79) and (80)). Note now that the right-hand side of (104) has the same form as the second term of the right-hand side of (81) (but now we have $(\boldsymbol{H}\boldsymbol{x})^{(r)}$ and $\boldsymbol{H}^{(r)}$ instead of $\boldsymbol{z}(\boldsymbol{x}^{(r)})$ and $\boldsymbol{x}^{(r)}$, respectively). Therefore, the same reasoning as was used in the proof of Theorem 1 shows that, as $\varepsilon\to 0$, we have

$$I(\varepsilon(\boldsymbol{H}\boldsymbol{x})^{(r)};\varepsilon(\boldsymbol{H}\boldsymbol{x})^{(r)}+\boldsymbol{w}|\boldsymbol{H}^{(r)}) = \frac{\log e}{2}\text{trace}\left\{E[\text{cov}((\boldsymbol{H}\boldsymbol{x})^{(r)}|\boldsymbol{H}^{(r)})]\right\}\varepsilon^2 - \frac{\log e}{4}\text{trace}\left\{E[\text{cov}^2((\boldsymbol{H}\boldsymbol{x})^{(r)}|\boldsymbol{H}^{(r)})]\right\}\varepsilon^4 + o(\varepsilon^4). \tag{106}$$

If $\boldsymbol{x}$ is proper, the matrix $\boldsymbol{H}$ satisfies condition (52), and

$$E[||\boldsymbol{H}\boldsymbol{x}||^{4+\alpha}] < c \tag{107}$$

$$E[||\boldsymbol{H}\boldsymbol{x}||^{4+\alpha}|\boldsymbol{H}] \le c||\boldsymbol{H}||^k \quad \text{a.s.} \tag{108}$$

$$\text{trace}\left\{\text{cov}^2(\boldsymbol{H}\boldsymbol{x}|\boldsymbol{H})\right\} \le c||\boldsymbol{H}||^l \quad \text{a.s.} \tag{109}$$

for some finite constants $\alpha > 0$, $c > 0$, $k > 0$, and $l > 0$. But a reasoning similar to that used in the proof of Theorem 2 shows that conditions (107)–(109) are fulfilled if $\boldsymbol{x}$ and $\boldsymbol{H}$ satisfy conditions (51) and (52).

To complete the proof of Theorem 3, we only need to verify that (106) implies (53). Indeed, one can check by direct calculations that

$$\text{trace}\left\{E[\text{cov}((\boldsymbol{Hx})^{(r)}|\boldsymbol{H}^{(r)})]\right\} = \text{trace}\left\{E[\text{cov}(\boldsymbol{Hx}|\boldsymbol{H})]\right\} = \text{trace}\left\{E[\boldsymbol{H}\text{cov}(\boldsymbol{x})\boldsymbol{H}^\dagger]\right\} \quad (110)$$

and

$$\text{trace}\left\{E[\text{cov}^2((\boldsymbol{Hx})^{(r)}|\boldsymbol{H}^{(r)})]\right\} = \frac{1}{2}\text{trace}\left\{E[\text{cov}^2(\boldsymbol{Hx}|\boldsymbol{H})]\right\} = \frac{1}{2}\text{trace}\left\{E[(\boldsymbol{H}\text{cov}(\boldsymbol{x})\boldsymbol{H}^\dagger)^2]\right\} \quad (111)$$

since $\boldsymbol{x}$ is proper (cf. (101) and (103)). Substituting $\varepsilon^2 = 2/N_0$ and the right-hand side expressions from (110) and (111) instead of the left-hand side ones into (106), we get the required formula (53). □

Proof of Theorem 4: Taking into account expression (2) for $\mathcal{S}_0$, we see that we need to show that $\ddot{C} = -\infty$.

For any input signal $\boldsymbol{x}$ whose second moment $E[||\boldsymbol{x}||^2] = mN_0\rho$ we can write

$$\begin{aligned} I(\boldsymbol{x};\boldsymbol{y}) - m\dot{C}(0)\rho &= D\left(P_{Y|X=\boldsymbol{x}} \parallel P_{Y|X=0}|P_{\boldsymbol{x}}\right) - D\left(P_Y \parallel P_{Y|X=0}\right) - amN_0\rho \\ &\le D\left(P_{Y|X=\boldsymbol{x}} \parallel P_{Y|X=0}|P_{\boldsymbol{x}}\right) - aE[||\boldsymbol{x}||^2] \\ &= \int_{\mathcal{R}^n}\left[D\left(P_{Y|X=\boldsymbol{x}_0} \parallel P_{Y|X=0}\right) - a||\boldsymbol{x}_0||^2\right]P_X(d\boldsymbol{x}_0) \\ &= I' + I'' \end{aligned} \quad (112)$$

where

$$I' \stackrel{\text{def}}{=} \int_{||\boldsymbol{x}_0||\le M}\left[D\left(P_{Y|X=\boldsymbol{x}_0} \parallel P_{Y|X=0}\right) - a||\boldsymbol{x}_0||^2\right]P_X(d\boldsymbol{x}_0)$$

and

$$I'' \stackrel{\text{def}}{=} \int_{||\boldsymbol{x}_0||> M}\left[D\left(P_{Y|X=\boldsymbol{x}_0} \parallel P_{Y|X=0}\right) - a||\boldsymbol{x}_0||^2\right]P_X(d\boldsymbol{x}_0).$$

Using conditions i) and ii) we easily obtain

$$I' \le -b\int_{||\boldsymbol{x}_0||\le M}||\boldsymbol{x}_0||^{2+\beta}P_X(d\boldsymbol{x}_0) \le -b\left[\int_{||\boldsymbol{x}_0||\le M}||\boldsymbol{x}_0||^2 P_X(d\boldsymbol{x}_0)\right]^{\frac{2+\beta}{2}} \quad (113)$$

and

$$I'' \le -(a-h)\int_{||\boldsymbol{x}_0||>M}||\boldsymbol{x}_0||^2 P_X(d\boldsymbol{x}_0). \quad (114)$$

Set

$$q_1^2 \stackrel{\text{def}}{=} \int_{||\boldsymbol{x}_0||\le M}||\boldsymbol{x}_0||^2 P_X(d\boldsymbol{x}_0)$$

and

$$q_2^2 \stackrel{\text{def}}{=} \int_{||\boldsymbol{x}_0||> M}||\boldsymbol{x}_0||^2 P_X(d\boldsymbol{x}_0).$$

It follows from (112)–(114) that

$$\sup_{\boldsymbol{x}:E[||\boldsymbol{x}||^2]\le mN_0\rho} I(\boldsymbol{x};\boldsymbol{y}) - m\dot{C}(0)\rho \le -\inf_{q_1,q_2}\left[bq_1^{2+\beta} + (a-h)q_2^2\right] \quad (115)$$

where the infimum on the right-hand side of (115) is taken over all nonnegative q_1 and q_2 such that $q_1^2 + q_2^2 = mN_0\rho$. It is clear that this infimum is equal to $b(mN_0\rho)^{(2+\beta)/2}$ if ρ is sufficiently small, and hence,

$$\sup_{\boldsymbol{x}:E[||\boldsymbol{x}||^2]\le mN_0\rho} I(\boldsymbol{x};\boldsymbol{y}) - m\dot{C}(0)\rho \le -b(mN_0\rho)^{(2+\beta)/2}. \quad (116)$$

From (116) we immediately arrive at the conclusion that $\ddot{C}(0) = -\infty$ since $1 + \frac{\beta}{2} < 2$ by condition ii). Thus, the theorem is proved. □

APPENDIX

A. *Proof of Theorem 5*

The main idea and method of the proof of Theorem 5 is quite related to that of Lemma 2.2 in [7]. Let $\boldsymbol{Q} = \boldsymbol{Q}_\varepsilon$ be the matrix of an orthogonal transformation in $\mathcal{R}^l$ such that

$$\boldsymbol{Q}\text{cov}(\boldsymbol{x}_\varepsilon)\boldsymbol{Q}^\dagger = \lambda\boldsymbol{I}$$

where $\lambda = \lambda_\varepsilon = (\lambda_{1\varepsilon}, \dots, \lambda_{l\varepsilon})$ is a vector whose components are eigenvalues of the covariance matrix $\text{cov}(\boldsymbol{x}_\varepsilon)$. Note that

$$\boldsymbol{Q}\text{cov}(\boldsymbol{w})\boldsymbol{Q}^\dagger = \text{cov}(\boldsymbol{w})$$

since $\text{cov}(\boldsymbol{w})$ is a diagonal matrix with equal diagonal elements. In other words, the random vector $\boldsymbol{Qx}_\varepsilon$ has uncorrelated components and, moreover

$$E[||\boldsymbol{Qx}_\varepsilon||^{4+\alpha}] \le \left(\log\frac{1}{\varepsilon}\right)^\mu$$

as obviously follows from (75). It is well known that any orthogonal transformation of random vectors does not change the mutual information, i.e.,

$$I(\varepsilon\boldsymbol{Qx}_\varepsilon; \varepsilon\boldsymbol{Qx}_\varepsilon + \boldsymbol{Qw}) = I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}).$$

Therefore, without loss of generality, we will always assume below that the random vector $\boldsymbol{x}_\varepsilon$ has uncorrelated components and, moreover, $E[\boldsymbol{x}_\varepsilon] = 0$ since the mutual information $I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w})$ does not depend on $E[\boldsymbol{x}_\varepsilon]$.

To continue the proof of the theorem we need the following lemma.

Lemma 2: Under the conditions of Theorem 2, the probability density function $q_\varepsilon(\boldsymbol{t})$ of the random vector $\varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}$ can be represented as

$$q_\varepsilon(\boldsymbol{t}) = p(\boldsymbol{t}) + \sum_{j=2}^{4}\frac{(-1)^j\varepsilon^j}{j!} \times E\left[\left(\frac{\partial}{\partial t_1}\mathrm{x}_{1\varepsilon} + \cdots + \frac{\partial}{\partial t_l}\mathrm{x}_{l\varepsilon}\right)^j p(\boldsymbol{t})\right] + r_\varepsilon(\boldsymbol{t}) \quad (117)$$

where $p(\boldsymbol{t})$ is the probability density function of $\boldsymbol{w}$ and $r_\varepsilon(\boldsymbol{t})$ satisfies the condition

$$|r_\varepsilon(\boldsymbol{t})| \le c\varepsilon^{4+\gamma} \text{ for some } \gamma > 0 \tag{118}$$

where $c > 0$ is a constant independent of ε and $\boldsymbol{x}_\varepsilon$.

Proof of Lemma 2: Formula (117) follows from the representation

$$q_\varepsilon(\boldsymbol{t}) = \int_{\mathcal{R}^l} p(\boldsymbol{t} - \varepsilon\boldsymbol{u}) dF_\varepsilon(\boldsymbol{u})$$

and Taylor's formula for $p(\boldsymbol{t} - \varepsilon\boldsymbol{u})$ at the point $\boldsymbol{t}$ where $r_\varepsilon(\boldsymbol{t})$ can be written in the form

$$r_\varepsilon(\boldsymbol{t}) = \frac{\varepsilon^4}{24}\int_{\mathcal{R}^l}\left[\left(\frac{\partial}{\partial t_1}u_1 + \cdots + \frac{\partial}{\partial t_l}u_l\right)^4 p(\boldsymbol{t} - \theta\varepsilon\boldsymbol{u}) - \left(\frac{\partial}{\partial t_1}u_1 + \cdots + \frac{\partial}{\partial t_l}u_l\right)^4 p(\boldsymbol{t})\right] dF_\varepsilon(\boldsymbol{u}). \tag{119}$$

Here $F_\varepsilon(\cdot)$ denotes the distribution function of $\boldsymbol{x}_\varepsilon$. Therefore, to prove the lemma we only need to show that $r_\varepsilon(\boldsymbol{t})$ satisfies inequality (118).

For any $\boldsymbol{i} = (i_1, \ldots, i_4)$ and k, where $i_1, \ldots, i_4$, $k \in \{1, \ldots, l\}$, set

$$\Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u})) \triangleq \left|\frac{\partial^4 p}{\partial t_{i_1}\ldots\partial t_{i_4}}(\boldsymbol{t} - \theta\varepsilon\boldsymbol{u}) - \frac{\partial^4 p}{\partial t_{i_1}\ldots\partial t_{i_4}}(\boldsymbol{t})\right| \tag{120}$$

and upper-bound the integrals

$$I_{\boldsymbol{i}}(k) = \int_{\mathcal{R}^l} u_k^4 \Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u})) dF_\varepsilon(\boldsymbol{u}). \tag{121}$$

It is clear that

$$I_{\boldsymbol{i}}(k) \le \sum_{j=0}^{l} I_{\boldsymbol{i}}(j,k) \tag{122}$$

where

$$I_{\boldsymbol{i}}(0,k) = \int_{\{|u_j| \le \varepsilon^{\beta-1}, j=1,\ldots,l\}} u_k^4 \Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u}) dF_\varepsilon(\boldsymbol{u})$$
$$I_{\boldsymbol{i}}(j,k) = \int_{\{|u_j| \ge \varepsilon^{\beta-1}, |u_k| \le \varepsilon^{\beta-1}\}} u_k^4 \Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u}) dF_\varepsilon(\boldsymbol{u}), \quad \text{if } j \ne k$$

and

$$I_{\boldsymbol{i}}(k,k) = \int_{\{|u_k| \ge \varepsilon^{\beta-1}\}} u_k^4 \Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u}) dF_\varepsilon(\boldsymbol{u})$$

and where β, $0 < \beta < 1$, is a constant which will be chosen later.

Now we have

$$\begin{aligned} I_{\boldsymbol{i}}(0,k) &\le c\varepsilon^\beta \int_{\mathcal{R}^l} u_k^4 dF_\varepsilon(\boldsymbol{u}) \\ &\le c\varepsilon^\beta \left(E[\|\boldsymbol{x}_\varepsilon\|^{4+\alpha}]\right)^{\frac{4}{4+\alpha}} \\ &\le c\varepsilon^\beta \left(\log\frac{1}{\varepsilon}\right)^{\frac{4\mu}{4+\alpha}} \le c\varepsilon^{\beta'} \end{aligned} \tag{123}$$

where $\beta' > 0$ is a constant such that $\beta' < \beta$. Here we have used the fact that $\Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u})) \le c\varepsilon\|\boldsymbol{u}\|$, the well-known inequalities between moments of random variables, and condition (75) of Theorem 2.

For any j, $j \ne k$, we obtain

$$\begin{aligned} I_{\boldsymbol{i}}(j,k) &\le c\varepsilon^{4(\beta-1)} P\left\{\|\boldsymbol{x}_\varepsilon\| > \varepsilon^{\beta-1}\right\} \\ &\le c\varepsilon^{4(\beta-1)} E[\|\boldsymbol{x}_\varepsilon\|^{4+\alpha}]\varepsilon^{-(\beta-1)(4+\alpha)} \\ &\le c\varepsilon^{\alpha-\alpha\beta}\left(\log\frac{1}{\varepsilon}\right)^\mu \le c\varepsilon^{\alpha'} \end{aligned} \tag{124}$$

where α' is a constant such that $0 < \alpha' < \alpha - \alpha\beta$ (note that $\alpha - \alpha\beta > 0$ if $\beta < 1$). To get (124), we have used inequality $\Delta_{\boldsymbol{i}}(\boldsymbol{t}, \theta\varepsilon\boldsymbol{u})) < \text{const} < \infty$ for all $\boldsymbol{t}$ and $\boldsymbol{u}$, the Chebyshev's inequality, and condition (75).

Finally, we have

$$\begin{aligned} I_{\boldsymbol{i}}(k,k) &\le c\int_{\{||u_k| > \varepsilon^{\beta-1}|\}} u_k^4 dF_\varepsilon(\boldsymbol{u}) \\ &= c\varepsilon^{4(\beta-1)}\int_{\{||u_k| > \varepsilon^{\beta-1}|\}} \left(\frac{u_k}{\varepsilon^{\beta-1}}\right)^4 dF_\varepsilon(\boldsymbol{u}) \\ &\le c\varepsilon^{\alpha-\alpha\beta}\int_{\mathcal{R}^l} \|\boldsymbol{u}\|^{4+\alpha} dF_\varepsilon(\boldsymbol{u}) \\ &\le c\varepsilon^{\alpha-\alpha\beta}\left(\log\frac{1}{\varepsilon}\right)^\mu \le c\varepsilon^{\alpha'}. \end{aligned} \tag{125}$$

Now, it follows from (122)–(125) that

$$0 \le I_{\boldsymbol{i}}(k) \le c\varepsilon^\gamma \tag{126}$$

for all $\boldsymbol{i}$ and k where $\gamma > 0$ is a constant. Taking into account (126), (119)–(121), and the fact that $|u_{i_1} \cdots u_{i_4}| \le \sum_{k=1}^{l} u_k^4$ for any $i_1, \ldots, i_4 \in \{1, \ldots, l\}$, we arrive at the conclusion that $r_\varepsilon(\boldsymbol{t})$ satisfies (118). This completes the proof of Lemma 2.[7] □

To complete the proof of Theorem 5, we now follow a reasoning similar to that used in [7, proof of Lemma 2.2]. Set

$$\begin{aligned} G_{\varepsilon,\kappa} &\stackrel{\text{def}}{=} \left\{\boldsymbol{t} \in \mathcal{R}^l : \exp\left[-\frac{\|\boldsymbol{t}\|^2}{2\sigma^2}\right] \ge \varepsilon^{4+\kappa}\right\} \\ \overline{G}_{\varepsilon,\kappa} &\stackrel{\text{def}}{=} \left\{\boldsymbol{t} \in \mathcal{R}^l : \exp\left[-\frac{\|\boldsymbol{t}\|^2}{2\sigma^2}\right] < \varepsilon^{4+\kappa}\right\} \end{aligned} \tag{127}$$

where $\kappa > 0$ is a sufficiently small number which will be chosen later, and

$$a_\varepsilon(\boldsymbol{t}) \triangleq \sum_{j=2}^{4} \frac{(-1)^j\varepsilon^j}{j!} E\left[\left(\frac{\partial}{\partial t_1}\mathsf{x}_{1\varepsilon} + \cdots + \frac{\partial}{\partial t_n}\mathsf{x}_{l\varepsilon}\right)^j p(\boldsymbol{t})\right]. \tag{128}$$

It follows from Lemma 2, definition of $\overline{G}_{\varepsilon,\kappa}$, and condition (75) that

$$q_\varepsilon(\boldsymbol{t}) < 1, \qquad \text{if } \boldsymbol{t} \in \overline{G}_{\varepsilon,\kappa}$$

[7] Although we considered only the case where the noise is Gaussian, under some (weak) additional conditions on the density of the non-Gaussian noise the same method of proof works.

for all sufficiently small ε. Therefore, using formulas (117) and (128), we easily get (later, throughout we will omit arguments $\boldsymbol{t}$ and ε in the notations of $a_\varepsilon(\boldsymbol{t})$, $r_\varepsilon(\boldsymbol{t})$, and $p(\boldsymbol{t})$)

$$\begin{aligned} I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}) &\geq -\int_{G_{\varepsilon,\kappa}} (a+r)\log p d\boldsymbol{t} + \int_{\overline{G}_{\varepsilon,\kappa}} p \log p d\boldsymbol{t} \\ &\quad - \log e \left[\int_{G_{\varepsilon,\kappa}} (p+a+r)\ln\left(1+\frac{a+r}{p}\right) d\boldsymbol{t}\right] \\ &= -\int_{\mathcal{R}^l} a\log p d\boldsymbol{t} + \int_{\overline{G}_{\varepsilon,\kappa}} p\log p d\boldsymbol{t} + \int_{\overline{G}_{\varepsilon,\kappa}} a \log p d\boldsymbol{t} \\ &\quad - \int_{G_{\varepsilon,\kappa}} r\log p d\boldsymbol{t} \\ &\quad - \log e \left[\int_{G_{\varepsilon,\kappa}} (a+r) d\boldsymbol{t} \right. \\ &\qquad \left. + \sum_{j=2}^{\infty} \frac{(-1)^j}{(j-1)j} \int_{G_{\varepsilon,\kappa}} \frac{(a+r)^j}{p^{j-1}} d\boldsymbol{t}\right]. \qquad (129) \end{aligned}$$

Lemma 3: If $\kappa > 0$ is sufficiently small, then

$$\begin{aligned} &\int_{\overline{G}_{\varepsilon,\kappa}} p\log p d\boldsymbol{t} = o(\varepsilon^4) \\ &\int_{\overline{G}_{\varepsilon,\kappa}} a\log p d\boldsymbol{t} = o(\varepsilon^4), \qquad \varepsilon \to 0 \qquad (130) \\ &\int_{G_{\varepsilon,\kappa}} r\log p d\boldsymbol{t} = o(\varepsilon^4) \\ &\int_{G_{\varepsilon,\kappa}} (a+r) d\boldsymbol{t} = o(\varepsilon^4), \qquad \varepsilon \to 0 \qquad (131) \end{aligned}$$

and

$$\sum_{j=3}^{\infty} \frac{(-1)^j}{(j-1)j} \int_{G_{\varepsilon,\kappa}} \frac{(a+r)^j}{p^{j-1}} d\boldsymbol{t} = o(\varepsilon^4), \qquad \varepsilon \to 0. \qquad (132)$$

Proof of Lemma 3: The proof equalities (130)–(132) is based on Lemma 2, properties of the Gaussian density function $p(\boldsymbol{t})$), and the definitions of the sets $G_{\varepsilon,\kappa}$ and $\overline{G}_{\varepsilon,\kappa}$.

The first equality in (130) follows from rather obvious relations

$$\begin{aligned} \left|\int_{G_{\varepsilon,\kappa}} p\log p d\boldsymbol{t}\right| &\leq c\int_{\|\boldsymbol{t}\|>\delta} \|\boldsymbol{t}\|^2 \exp\left(-\|\boldsymbol{t}\|^2/2\right) d\boldsymbol{t} \\ &\leq c\int_{|s|>\delta} s^{l+1}\exp(-s^2/2) ds \\ &\leq c\left(\log\frac{1}{\varepsilon}\right)^{1/2} \varepsilon^{4+\kappa} = o(\varepsilon^4), \qquad \varepsilon \to 0 \end{aligned}$$

where δ is defined by the equality

$$\exp(-\delta^2/2) = \varepsilon^{4+\kappa}.$$

The proof of the second equality in (130) is similar to the first one if we note that

$$\left|\frac{\partial^j p}{\partial t_{i_1}\ldots\partial t_{i_j}}(\boldsymbol{t})\right| \leq c\|\boldsymbol{t}\|^j p(\boldsymbol{t}), \qquad \boldsymbol{t} \in \overline{G}_{\varepsilon,\kappa}.$$

Equality (131) can be proved by the same method as (130) taking into account that

$$\int_{G_{\varepsilon,\kappa}} a d\boldsymbol{t} = -\int_{\overline{G}_{\varepsilon,\kappa}} a d\boldsymbol{t}.$$

The proof of (132) is a little more complicated. First, note that

$$\left|\frac{a}{p^{(j-1)/j}}\right| \leq c\varepsilon^2 \log^\tau \frac{1}{\varepsilon}, \qquad \boldsymbol{t} \in G_{\varepsilon,\kappa}$$

and

$$\left|\frac{r}{p^{(j-1)/j}}\right| \leq c\frac{\varepsilon^{4+\gamma}}{\varepsilon^{(4+\kappa)(j-1)/j}} c\varepsilon^{\frac{4+\kappa}{j}} \varepsilon^{\gamma-\kappa}, \qquad \boldsymbol{t} \in G_{\varepsilon,\kappa}$$

since it is clear that

$$|a| \leq c\varepsilon^2 \left(\log\frac{1}{\varepsilon}\right)^\tau p, \qquad \boldsymbol{t} \in G_{\varepsilon,\kappa},$$

where τ is a finite number. Therefore, we get

$$\begin{aligned} \left|\frac{(a+r)^j}{p^{j-1}}\right| &\leq \left[c\varepsilon^{\frac{4+\kappa}{j}}\left(\varepsilon^{2-\frac{4+\kappa}{j}}\log^\tau\frac{1}{\varepsilon} + \varepsilon^{\gamma-\kappa}\right)\right]^j \\ &\leq \varepsilon^{4+\kappa}\cdot o(1), \qquad \boldsymbol{t} \in G_{\varepsilon,\kappa} \qquad (133) \end{aligned}$$

if $j \geq 3$ and κ is sufficiently small. Now equality (133) implies (132). Lemma 3 is proved. □

Lemma 3 and inequality (129) imply

$$\begin{aligned} I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}) \geq &-\int_{\mathcal{R}^l} a\log p d\boldsymbol{t} \\ &- \frac{\log e}{2}\int_{G_{\varepsilon,\kappa}} \frac{(a+r)^2}{p} d\boldsymbol{t} + o(\varepsilon^4), \\ &\qquad \varepsilon \to 0. \qquad (134) \end{aligned}$$

By direct calculations one can verify that

$$-\int_{\mathcal{R}^l} a\log p d\boldsymbol{t} = \frac{\log e}{2\sigma^2}\operatorname{trace}\{\operatorname{cov}(\boldsymbol{x}_\varepsilon)\}\varepsilon^2. \qquad (135)$$

Furthermore, we have

$$\begin{aligned} \int_{G_{\varepsilon,\kappa}} \frac{(a+r)^2}{p} d\boldsymbol{t} = &\int_{\mathcal{R}^l} \frac{a^2}{p} d\boldsymbol{t} - \int_{\overline{G}_{\varepsilon,\kappa}} \frac{a^2}{p} d\boldsymbol{t} \\ &+ 2\int_{G_{\varepsilon,\kappa}} \frac{ar}{p} d\boldsymbol{t} + \int_{G_{\varepsilon,\kappa}} \frac{r^2}{p} d\boldsymbol{t}. \qquad (136) \end{aligned}$$

Using a reasoning similar to that used in the proof of Lemma 3, it can be checked that

$$\begin{aligned} &\int_{\overline{G}_{\varepsilon,\kappa}} \frac{a^2}{p} d\boldsymbol{t} = o(\varepsilon^4) \\ &\int_{G_{\varepsilon,\kappa}} \frac{ar}{p} d\boldsymbol{t} = o(\varepsilon^4) \end{aligned}$$

$$\int_{G_{\varepsilon,\kappa}} \frac{r^2}{p} d\boldsymbol{t} = o(\varepsilon^4), \qquad \varepsilon \to 0. \tag{137}$$

It is not difficult to show that

$$\int_{\mathcal{R}^l} \frac{a^2}{p} d\boldsymbol{t} = \frac{1}{2\sigma^4} \text{trace}\left\{\text{cov}^2(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^4 + o(\varepsilon^4), \qquad \varepsilon \to 0. \tag{138}$$

Thus, relations (134)–(138) yield

$$\begin{aligned} I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}) \geq \frac{\log e}{2\sigma^2} &\text{trace}\left\{\text{cov}(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^2 \\ &- \frac{\log e}{4\sigma^4}\text{trace}\left\{\text{cov}^2(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^4 + o(\varepsilon^4), \\ &\qquad \varepsilon \to 0. \end{aligned} \tag{139}$$

On the other hand, we have

$$\begin{aligned} I(\varepsilon\boldsymbol{x}_\varepsilon; \varepsilon\boldsymbol{x}_\varepsilon + \boldsymbol{w}) &\leq I(\varepsilon\widehat{\boldsymbol{x}}_\varepsilon; \varepsilon\widehat{\boldsymbol{x}}_\varepsilon + \boldsymbol{w}) \\ &= \frac{\log e}{2\sigma^2}\text{trace}\left\{\text{cov}(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^2 \\ &\quad - \frac{\log e}{4\sigma^4}\text{trace}\left\{\text{cov}^2(\boldsymbol{x}_\varepsilon)\right\}\varepsilon^4 + o(\varepsilon^4), \\ &\qquad \varepsilon \to 0 \end{aligned} \tag{140}$$

where $\widehat{\boldsymbol{x}}_\varepsilon$ is a Gaussian random vector with the same covariance matrix as $\boldsymbol{x}_\varepsilon$. Theorem 5 follows from (139) and (140). □

Proof of Lemma 1: Since it is assumed that for the Laplacian noise channel the noise components are i.i.d. random variables (and hence the channel is memoryless) each with Laplacian density function (69), we can take $m = 1$ without loss of generality. In this case, we have

$$\begin{aligned} D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) &= -2 + \frac{2}{\sqrt{N_0}}(|\Re\mathsf{x}_0| + |\Im\mathsf{x}_0|) \\ &\quad + \exp\left[-\frac{2|\Re\mathsf{x}_0|}{\sqrt{N_0}}\right] + \exp\left[-\frac{2|\Im\mathsf{x}_0|}{\sqrt{N_0}}\right]. \end{aligned} \tag{141}$$

We intend to show that for this divergence, conditions i) and ii) are satisfied with

$$a = \frac{2}{N_0}, \quad b = \frac{1}{3\sqrt{2}(N_0)^{3/2}}, \quad h = \frac{1.9}{N_0}$$
$$M = \frac{\sqrt{N_0}}{2}, \text{ and } \beta = 1.$$

Denote for simplicity

$$\Re\mathsf{x}_0 = x, \qquad \Im\mathsf{x}_0 = y$$

and let

$$D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) = D_1(x) + D_1(y) \tag{142}$$

where

$$D_1(x) \stackrel{\text{def}}{=} 1 + \frac{2}{\sqrt{N_0}}|x| + \exp\left[-\frac{2|x|}{\sqrt{N_0}}\right]. \tag{143}$$

It is clear that

$$\lim_{\mathsf{x}_0 \to 0} \frac{D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right)}{\|\mathsf{x}_0\|^2} = \frac{2}{N_0} \tag{144}$$

and, therefore, $a = 2/N_0$.

Now, let us show that

$$D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) \leq \frac{2}{N_0}\|\mathsf{x}_0\|^2 - \frac{1}{3\sqrt{2}(N_0)^{3/2}}\|\mathsf{x}_0\|^3$$

for

$$\|\mathsf{x}_0\| \leq \frac{\sqrt{N_0}}{2}. \tag{145}$$

Indeed, for $|x| \leq \sqrt{N_0}/2$, we have

$$\begin{aligned} D_1(x) &\leq \frac{2}{N_0}|x|^2 - \frac{4}{3(N_0)^{3/2}}|x|^3 + \frac{2}{3(N_0)^2}|x|^4 \\ &\leq \frac{2}{N_0}|x|^2 - \frac{2}{3(N_0)^{3/2}}|x|^3. \end{aligned} \tag{146}$$

Here, we have used the well-known property of the Leibnitz power series: $a_0 - a_1 + a_2 - a_3 + \cdots \leq a_0$ if the coefficients a_i vanish to zero monotonically. Thus,

$$\begin{aligned} D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) &\leq \frac{2}{N_0}(|x|^2 + |y|^2) \\ &\quad - \frac{2}{3(N_0)^{3/2}}(|x|^3 + |y|^3) \\ &\leq \frac{2}{N_0}\|\mathsf{x}_0\|^2 - \frac{1}{3\sqrt{2}(N_0)^{3/2}}\|\mathsf{x}_0\|^3 \end{aligned} \tag{147}$$

for $\|\mathsf{x}_0\| \leq \sqrt{N_0}/2$, i.e., (145) is proved.

It remains to show that

$$D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) \leq \frac{1.9}{N_0}\|\mathsf{x}_0\|^2 \text{ for } \|\mathsf{x}_0\| > \frac{\sqrt{N_0}}{2}. \tag{148}$$

To this end, note that

$$-1 + 2t + e^{-2t} < 1.7t^2, \qquad \text{for } t > \frac{1}{2\sqrt{2}} \tag{149}$$

since, as can easily be seen

$$f\left(\frac{1}{2\sqrt{2}}\right) > 0, \quad f'\left(\frac{1}{2\sqrt{2}}\right) > 0$$

and

$$f''(t) > 0, \qquad \text{for } t > \frac{1}{2\sqrt{2}}$$

where $f(t) = 1.7t^2 + 1 - 2t - e^{-2t}$. It follows from (149) that

$$D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) \leq \frac{1.7}{N_0}\|\mathsf{x}_0\|^2, \qquad \text{if } |x| > \frac{\sqrt{N_0}}{2\sqrt{2}}$$

and

$$|y| > \frac{\sqrt{N_0}}{2\sqrt{2}}. \tag{150}$$

Assume for definiteness that $|x|^2 = \lambda\|\mathsf{x}_0\|^2$, $0 \leq \lambda \leq 1/2$. Then, taking into account (150), (146), and (147), we obtain

$$\begin{aligned} &D\left(P_{Y|X=\mathsf{x}_0} \,\|\, P_{Y|X=0}\right) \\ &\quad \leq \max_{0\leq\lambda\leq 1/2}\left[\frac{2}{N_0}\lambda\|\mathsf{x}_0\|^2 + \frac{1.7}{N_0}(1-\lambda)\|\mathsf{x}_0\|^2\right] \\ &\quad < \frac{1.9}{N_0}\|\mathsf{x}_0\|^2 \end{aligned}$$

for all $\|\mathsf{x}_0\| > \sqrt{N_0}/2$, i.e., inequality (147) is proved. This completes the proof of the lemma. □

References

[1] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

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

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

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

[5] V. V. Prelov, "Asymptotic behavior of the capacity of a continuous channel with a large amount of noise," *Probl. Pered. Inform.*, vol. 6, pp. 40–57, Apr.–June 1970.

[6] I. A. Ibragimov and R. Z. Khasminsky, "Weak signal transmission in a memoryless channel," *Probl. Inform. Transm.*, vol. 8, no. 4, pp. 28–39, Oct. 1972.

[7] V. V. Prelov, "Communication channel capacity with almost Gaussian noise," *Theory Probab. Its Applic.*, vol. 33, no. 3, pp. 405–422, July 1988.

[8] ——, "Asymptotic expansions for the mutual information and for the capacity of continuous memoryless channels with weak input signal," *Probl. Control Inform. Theory*, vol. 18, no. 2, pp. 91–106, 1989.

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

[10] M. Pinsker, V. V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1877–1888, Nov. 1995.

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

[12] D. Guo, S. Shamai, and S. Verdú, "Mutual information and MMSE in Gaussian channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, June/July 2004, p. 347.

[13] M. C. Gursoy, H. V. Poor, and S. Verdú, "Efficient signaling for low-power Rician fading channels," in *Proc. 40th Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2–4, 2002.

[14] ——, "On the capacity-achieving distribution of the noncoherent Rician fading channel," in *Proc. 2003 Canadian Workshop on Information Theory*, Waterloo, ON, Canada, May 18–21, 2003.

[15] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent additive white Gaussian noise channel," in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, July 2002, p. 165.

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

[17] B. Hajek and V. Subramanian, "Capacity and reliability function for small peak signal constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp. 828–839, Apr. 2002.

[18] C. Rao and B. Hassibi, "Analysis of multiple antenna wireless links at low signal-to-noise ratio," in *Proc. 2003 IEEE Int. Symp. Information Theory*, Yokohama, Japan, June/July 2003, p. 468.