DIMACS Technical Report 2002-16
April 2002

Second-order Asymptotics of Mutual Information[1]

by

Viacheslav Prelov
Inst. for Information Transmission Problems
Russian Academy of Sciences
19 Bol'shoi Karetny, Moscow 101447, Russia

Sergio Verdú
Dept. Electrical Engineering
Princeton University
Princeton, New Jersey, USA

[1]This work was carried out with the partial support of DIMACS and NSF.

ABSTRACT

A formula for the second-order expansion of the input-output mutual information of multi-dimensional channels as the signal-to-noise ratio 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 second-order asymptotics of mutual information finds application in the analysis of the bandwidth-power tradeoff achieved by specific (not necessarily optimum) input signaling in the wideband regime.

Keywords: Mutual Information, Channel Capacity, Fading Channels, Nonlinear Channels, Low-power communication.

1 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. The first derivative of capacity (at zero signal-to-noise ratio) 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. For a given first derivative, 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 signal-to-noise ratio [1].

For the additive white Gaussian noise 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, in particular when they are not fully known at the receiver, the structure of the input distribution depends on the signal-to-noise-ratio. For example, in the case of the scalar Rayleigh channel unknown at the receiver, it is known [3, 4] that the optimum distribution is discrete and that for sufficiently low signal-to-noise-ratio, the optimum distribution has a mass at zero and a vanishing mass that migrates to infinity as the signal-to-noise-ratio 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 non-capacity-achieving constellation whose structure is not adapted to the signal-to-noise-ratio. 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 [1]. 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].

An input distribution is said to be *first-order optimal* [1] if the derivative at zero signal-to-noise ratio of its mutual information is equal to the derivative of capacity at zero signal-to-noise ratio. A first-order optimal distribution achieves the minimum energy per information bit when bandwidth is allowed to be infinite. An input distribution is said to be *second-order optimal* [1] if it is first-order optimal and the second derivative at zero signal-to-noise ratio of its mutual information is equal to the second derivative of capacity at zero signal-to-noise ratio. In the wideband regime, for a given power and rate, a second-order optimal distribution requires the minimum possible bandwidth.

An important measure of the bandwidth-power tradeoff in the low power regime the fading is the "wideband slope" introduced in [1]. Denote by $C(\rho)$ the capacity (per channel use) as a function of the signal-to-noise ratio. It is shown in [1] that the slope (b/s/Hz/(3 dB) of spectral efficiency $\mathcal{S}_0$ at the point $(E_b/N_0)_{\min}$ is given by the formula

$$\mathcal{S}_0 = \frac{2[\dot{C}(0)]^2}{-\ddot{C}(0)}, \tag{1}$$

where $\dot{C}(0)$ and $\ddot{C}(0)$ denote the first and second derivative at 0, respectively, of the function $C(\rho)$ in nats. Here, E_b denotes the transmission energy per information bit and

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

denotes the minimum E_b/N_0 required for reliable communication. Another motivation for the asymptotic expansions found in this paper is the comparison of the minimum energy and wideband slope achieved by various distributions.

Motivated by the lack of explicit expressions for many channels and inputs of interest, the input-output mutual information has been studied in the low-power regime in several previous works [5, 6, 7, 8, 2, 9, 1]. NonGaussian additive noise is treated in [6, 9], with not only input power restrictions but also peak-power limitations. With the exception of [7, 8, 1], those works limit themselves to the first-order asymptotic expansion of mutual information, and the scope of [8] is a channel not subject to fading or impairments other than additive noise which is slightly nonGaussian (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 2, 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 3 gives the statement of our main results, which are proved in Section 4.

2 Model

2.1 Nonlinear Channels

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

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

where $\mathbf{n}$ 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 $\mathbf{x} = (\mathsf{x}_1, \ldots, \mathsf{x}_n)^T$ also has complex-valued components with finite second moments; and the input-dependent received signal is a proper-complex[2] random m-vector $\mathbf{g}(\mathbf{x})$. We assume throughout that $\mathbf{g}(\mathbf{x})$ and $\mathbf{n}$ are mutually independent.

[1] T denotes transposition

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

2.2 Fading Channels

A fading channel

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

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

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

For the fading channel (4), $\mathbf{x}, \mathbf{H}$, and $\mathbf{n}$ 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 $\mathbf{H}$ at the receiver.

2.3 Additive White Gaussian Noise Channels

If $m = n$ and $\mathbf{H} = \mathbf{I}$ a.s. where $\mathbf{I}$ is the $m \times m$ identity matrix, then (4) reduces to the ordinary additive white Gaussian noise (AWGN) channel

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

2.4 Almost-Gaussian Noise Channels

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

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

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

$$\mathbf{y} = \mathbf{x} + \mathbf{u} + \mathbf{n}, \tag{8}$$

As we let $N_0 \to \infty$, we obtain a channel with almost Gaussian noise, where the contaminating weak noise $\mathbf{u}$ can be considered as a random parameter.

2.5 Notation

In this paper, we are interested in finding the asymptotic behavior of the mutual information $I(\mathbf{x}; \mathbf{g}(\mathbf{x}) + \mathbf{n})$ up to terms of order $o(N_0^{-2})$, or equivalently if N_0 is held fixed, the asymptotics of $I(\varepsilon\mathbf{x}; \varepsilon\mathbf{g}(\mathbf{x}) + \mathbf{n})$ 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 $\mathbf{g}(\mathbf{x})$ by

$$\overline{\mathbf{g}}(\mathbf{x}) = E[\mathbf{g}(\mathbf{x})|\mathbf{x}] \tag{9}$$

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

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

and

$$\mathrm{cov}(\mathbf{g}(\mathbf{x})|\mathbf{x}) = E[(\mathbf{g}(\mathbf{x}) - \overline{\mathbf{g}}(\mathbf{x}))(\mathbf{g}(\mathbf{x}) - \overline{\mathbf{g}}(\mathbf{x}))^{\dagger}|\mathbf{x}]. \tag{11}$$

We will also use the covariance matrix $\mathrm{cov}(\overline{\mathbf{g}}(\mathbf{x}))$ of the conditional mean vector $\overline{\mathbf{g}}(\mathbf{x})$ which, according to (10), is defined as

$$\begin{aligned} \mathrm{cov}(\overline{\mathbf{g}}(\mathbf{x})) &= E[(\overline{\mathbf{g}}(\mathbf{x}) - E[\overline{\mathbf{g}}(\mathbf{x})])(\overline{\mathbf{g}}(\mathbf{x}) - E[\overline{\mathbf{g}}(\mathbf{x})])^{\dagger}] \\ &= E[(\overline{\mathbf{g}}(\mathbf{x}) - E[\mathbf{g}(\mathbf{x})])(\overline{\mathbf{g}}(\mathbf{x}) - E[\mathbf{g}(\mathbf{x})])^{\dagger}]. \end{aligned} \tag{12}$$

Note also that

$$E[\mathrm{cov}(\mathbf{g}(\mathbf{x})|\mathbf{x})] = \mathrm{cov}(\mathbf{g}(\mathbf{x})) \tag{13}$$

Throughout the paper, $\mathrm{trace}(\cdot)$ denotes the trace of the corresponding matrix and the base of $\log(\cdot)$ is the unit in which we measure the mutual information. Let $||\mathbf{x}||^2 = \sum\limits_j |\mathsf{x}_j|^2$ and $||\mathbf{H}||^2 = \sum\limits_{i,j} |\mathsf{h}_{ij}|^2$ be the Euclidean squared norms of vectors and matrices, respectively. In this paper, we make frequent use of the Euclidean (or Frobenius) squared norm of a covariance matrix, which is equal to the sum of the squares of all its coefficients, or equivalently, the trace of the square of the covariance matrix.

3 Main Results

3.1 Statements

The following theorem is the main result of the paper.

Theorem 1 *Assume that the probability distribution of the input signal* $\mathbf{x}$ *satisfies the condition*

$$P\{||\mathbf{x}|| > \delta\} \le \exp\{-\delta^{\nu}\} \tag{14}$$

for all sufficiently large $\delta > 0$*, where* $\nu > 0$ *is a positive constant. Assume also that* $\mathbf{g}(\mathbf{x})$ *and* $\mathbf{g}(\mathbf{x}_0)$ *for any fixed* $\mathbf{x}_0$ *are proper complex random vectors and there exist some finite constants* $\alpha > 0$*,* $c > 0$*,* $k > 0$*, and* $l > 0$ *such that*

$$E[||\mathbf{g}(\mathbf{x})||^{4+\alpha}] < c, \tag{15}$$

[3] $\dagger$ denotes conjugate transposition

$$E[||\mathbf{g}(\mathbf{x})||^{4+\alpha}|\mathbf{x}] \leq c||\mathbf{x}||^k \quad a.s., \tag{16}$$

and

$$\text{trace}\left\{\text{cov}^2(\mathbf{g}(\mathbf{x})|\mathbf{x})\right\} \leq c||\mathbf{x}||^l \quad a.s. \tag{17}$$

Then, as $N_0 \to \infty$*, uniformly in all* $\mathbf{x}$ *satisfying conditions of the theorem the following asymptotic expression holds:*

$$\begin{aligned} I(\mathbf{x};\, \mathbf{g}(\mathbf{x}) + \mathbf{n}) &= \frac{\log e}{N_0}\text{trace}\left\{\text{cov}(\overline{\mathbf{g}}(\mathbf{x}))\right\} \\ &+ \frac{\log e}{2N_0^2}\left[\text{trace}\left\{E[\text{cov}^2(\mathbf{g}(\mathbf{x})|\mathbf{x})]\right\} - \text{trace}\left\{\text{cov}^2(\mathbf{g}(\mathbf{x}))\right\}\right] + o(N_0^{-2}). \end{aligned} \tag{18}$$

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* $\mathbf{x}$ *satisfies condition* (14) *of Theorem 1, and that* $\bar{\mathbf{H}}\mathbf{x}$ *is proper complex where* $\bar{\mathbf{H}} = E[\mathbf{H}]$*. Assume that the channel matrix is such that there exists a constant* $\alpha > 0$ *such that*

$$E[||\mathbf{H}||^{4+\alpha}] < \infty \tag{19}$$

Furthermore, the real and imaginary parts [4] *of the coefficients in* $\mathbf{H}$ *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* $\mathbf{x}$ *satisfying conditions of the theorem we have*

$$\begin{aligned} I(\mathbf{x};\, \mathbf{H}\mathbf{x} + \mathbf{n}) &= \frac{\log e}{N_0}E[||\bar{\mathbf{H}}(\mathbf{x} - E[\mathbf{x}])||^2] \\ &+ \frac{\log e}{2N_0^2}\left[\text{trace}\left\{E[\text{cov}^2(\mathbf{H}\mathbf{x}\,|\,\mathbf{x})]\right\} - \text{trace}\left\{\text{cov}^2(\mathbf{H}\mathbf{x})\right\}\right] + o(N_0^{-2}). \end{aligned} \tag{20}$$

Corollary 1. In particular, if $\bar{\mathbf{H}} = 0$, then

$$I(\mathbf{x};\, \mathbf{H}\mathbf{x} + \mathbf{n}) = \frac{\log e}{2N_0^2}\text{trace}\left\{E\left[\left(E\left[\mathbf{H}\mathbf{x}\mathbf{x}^\dagger\mathbf{H}^\dagger|\mathbf{x}\right]\right)^2\right] - \left(E\left[\mathbf{H}E[\mathbf{x}\mathbf{x}^\dagger]\mathbf{H}^\dagger\right]\right)^2\right\} + o(N_0^{-2}). \tag{21}$$

[4] The real and imaginary parts of x are denoted by $\Re\mathsf{x}$ and $\Im\mathsf{x}$ respectively

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

$$I(\mathbf{x};\, \bar{\mathbf{H}}\mathbf{x} + \mathbf{n}) = \frac{\log e}{N_0}\mathrm{trace}\left\{\bar{\mathbf{H}}\mathrm{cov}(\mathbf{x})\bar{\mathbf{H}}^\dagger\right\} - \frac{\log e}{2N_0^2}\,\mathrm{trace}\left\{(\bar{\mathbf{H}}\mathrm{cov}(\mathbf{x})\bar{\mathbf{H}}^\dagger)^2\right\} + o(N_0^{-2}). \quad (22)$$

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

$$I(\mathsf{x};\, \mathsf{hx} + \mathsf{n}) = \frac{\log e}{N_0}\mathrm{var}[\mathsf{x}]|\bar{\mathsf{h}}|^2 + \frac{\log e}{2N_0^2}\left[E[|\mathsf{x}|^4]\mathrm{var}^2[\mathsf{h}] - \mathrm{var}^2[\mathsf{hx}]\right] + o(N_0^{-2}). \quad (23)$$

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

$$I(\mathsf{x};\, \mathsf{hx} + \mathsf{n}) = \frac{\log e}{2N_0^2}\left(E[|\mathsf{h}|^2]\right)^2 \mathrm{var}[|\mathsf{x}|^2] \; + o(N_0^{-2}), \quad (24)$$

and for a deterministic channel, we obtain

$$I(\mathsf{x};\, \bar{\mathsf{h}}\mathsf{x} + \mathsf{n}) = \frac{\log e}{N_0}\mathrm{var}[\mathsf{x}]|\bar{\mathsf{h}}|^2 - \frac{\log e}{2N_0^2}\mathrm{var}^2[\mathsf{x}]|\bar{\mathsf{h}}|^4 + o(N_0^{-2}). \quad (25)$$

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

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

$$E[||\mathbf{x}||^{4+\alpha}] < c. \quad (26)$$

Assume also that the matrix $\mathbf{H}$ *satisfies a condition similar to* (14)*, i.e.,*

$$P\left\{||\mathbf{H}|| > \delta\right\} \le \exp\left\{-\delta^\nu\right\} \quad (27)$$

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

$$I(\mathbf{x};\, \mathbf{H}\mathbf{x} + \mathbf{n}\;|\mathbf{H}) = \frac{\log e}{N_0}\mathrm{trace}\left\{E[\mathbf{H}\mathrm{cov}(\mathbf{x})\mathbf{H}^\dagger]\right\} - \frac{\log e}{2N_0^2}\,\mathrm{trace}\left\{E[(\mathbf{H}\mathrm{cov}(\mathbf{x})\mathbf{H}^\dagger)^2]\right\} + o(N_0^{-2}). \quad (28)$$

Corollary 4. Under the conditions of Theorem 3, and assuming that the receiver knows the channel but the transmitter 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 identically distributed Gaussian complex random variables. The corollary follows immediately upon noting from (28) 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 (28) we can obtain the derivatives obtained in [1] by differentiating the log-determinant capacity formula:

$$\dot{C}(0) = \frac{\log e}{2n} \text{trace} \left\{ E[\mathbf{H}^{\dagger}\mathbf{H}] \right\} \tag{29}$$

and

$$\ddot{C}(0) = -\frac{m \log e}{2n^2} \text{trace} \left\{ E[(\mathbf{H}^{\dagger}\mathbf{H})^2] \right\} \tag{30}$$

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).

3.2 Remarks

Remark 1. In the case where conditions of proper complexity of $\mathbf{g}(\mathbf{x})$ and/or $\mathbf{g}(\mathbf{x}_0)$ given $\mathbf{x}_0$ are not fulfilled but other conditions of Theorem 1 are satisfied, the asymptotic expression (18) remains true if we replace $\mathbf{x}$ and $\mathbf{g}(\mathbf{x})$ on the right-hand side of (18) and in condition (17) by real-valued random vectors of lengths $2n$ and $2m$ whose components are real and imaginary parts of the complex-valued components of vectors $\mathbf{x}$ and $\mathbf{g}(\mathbf{x})$, respectively. This assertion follows from the proof of Theorem 1 (see formula 78). Similar remarks apply to Theorem 2 and Theorem 3.

Remark 2. For the fading channel (4) the mutual information between input and output signal (and therefore its asymptotic expression (20)) depends on the mean of the input signal $\mathbf{x}$ when the channel is not known at the receiver. Moreover, the second term of (20) depends, in general, not only on covariance matrix of $\mathbf{x}$ but also on mixed fourth moments of the components of $\mathbf{x}$. In contrast, when the channel is known at the receiver, the second-order expansion (28) 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 3. 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 $I(\mathbf{x}; \mathbf{x} + \mathbf{n})$ up to terms of order $o(N_0^{-s})$, $N_0 \to \infty$, was obtained under the condition that $E[||\mathbf{x}||^{s+\alpha}] < \infty$ for some $\alpha > 0$.

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

$$I(\mathbf{x};\, \mathbf{g}(\mathbf{x}) + \mathbf{n}) \le \frac{\log e}{N_0} \text{trace} \left\{ \text{cov}(\overline{\mathbf{g}}(\mathbf{x})) \right\} + \frac{\log e}{2N_0^2} \text{trace} \left\{ E[\text{cov}^2(\mathbf{g}(\mathbf{x}) \,|\, \mathbf{x})] \right\}. \tag{31}$$

Note that this bound holds for all N_0. But, comparing (31) and (18), we can observe that, in general, the upper bound (31) is not tight asymptotically, although the leading terms on the right-hand sides of (31) and (18) coincide if $N_0 \to \infty$.

Remark 5. Our main result suggests the definition of the following measure of the dependence between arbitrary random vectors $\mathbf{x}$ and $\mathbf{y}$:

$$J(\mathbf{x};\mathbf{y}) = \text{trace}\left\{E\left[E^2[\mathbf{y}\mathbf{y}^\dagger|\mathbf{x}]\right] - E^2[\mathbf{y}\mathbf{y}^\dagger]\right\} \tag{32}$$

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

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

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

$$E[\mathbf{M}^2] - E^2[\mathbf{M}] \geq 0. \tag{33}$$

Furthermore only when $\mathbf{M}$ is deterministic, does the trace of the matrix in the left side of (33) become zero. Letting $\mathbf{M} = E[\mathbf{y}\mathbf{y}^\dagger|\mathbf{x}]$ we see that $E\left[E^2[\mathbf{y}\mathbf{y}^\dagger|\mathbf{x}]\right] - E^2[\mathbf{y}\mathbf{y}^\dagger]$ is nonnegative definite, with zero trace only if $E[\mathbf{y}\mathbf{y}^\dagger|\mathbf{x}]$ is deterministic.

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

$$J(\mathbf{x};\mathbf{y}) = J(\mathbf{x};\mathbf{y}+\mathbf{n}), \tag{34}$$

and if $E[\mathbf{y}|\mathbf{x}]$ is not a function of $\mathbf{x}$, then

$$J(\mathbf{x};\mathbf{y}) = J(\mathbf{x};\mathbf{y}+\mathbf{m}), \tag{35}$$

for any deterministic $\mathbf{m}$. Thus, $J(\mathbf{x};\mathbf{y})$ is a measure of the "multiplicative dependence" between $\mathbf{x}$ and $\mathbf{y}$.

Note that in the special case in which $\bar{\mathbf{g}}(\mathbf{x}) = 0$,

$$J(\mathbf{x};\mathbf{g}(\mathbf{x})) = \text{trace}\left\{E[\text{cov}^2(\mathbf{g}(\mathbf{x})|\mathbf{x})]\right\} - \text{trace}\left\{\text{cov}^2(\mathbf{g}(\mathbf{x}))\right\} \tag{36}$$

the nonnegativity of which is required (under $\bar{\mathbf{g}}(\mathbf{x}) = 0$) by the asymptotic expansion of Theorem 1. Under the technical conditions in Theorem 1, if $E[\mathbf{z}|\mathbf{x} = 0]$, then

$$\lim_{N_0 \to \infty} N_0^2 I(\mathbf{x};\mathbf{z} + \sqrt{N_0}\mathbf{n}) = J(\mathbf{x};\mathbf{z}) \tag{37}$$

where the vector $\mathbf{n}$) is proper-complex Gaussian with independent unit-variance components.

Remark 6. According to Theorem 1 in [1], the derivative with respect to signal-to-noise ratio of channel capacity at zero signal-to-noise ratio is given by the maximum channel gain

$$\mathsf{G} = \sup_{P_{\mathbf{x}}} \frac{E[||\mathbf{H}\mathbf{x}||^2]}{E[||\mathbf{x}||^2]} \tag{38}$$

Note that this does not contradict (20) as the asymptotically optimal input distribution does not satisfy condition (14).

3.3 Examples

Example 1: Ricean channel.

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

$$\mathbf{H} = (\bar{h} + g)\mathbf{I}, \tag{39}$$

where $\bar{h}$ is deterministic and g is zero-mean proper complex Gaussian with variance γ^2 independent of $\mathbf{x}$ and $\mathbf{n}$. 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 general formula (20) reduces to

$$\begin{aligned} I(\mathbf{x}; \mathbf{Hx} + \mathbf{n}) &= \frac{\log e}{N_0} |\bar{h}|^2 \mathrm{var}[\mathbf{x}] \\ + \quad \frac{\log e}{2N_0^2} &\left[\gamma^4 E[||\mathbf{x}||^4] - \mathrm{trace}\left\{ \left[\gamma^2 E[\mathbf{x}\mathbf{x}^\dagger] + |\bar{h}|^2 \mathrm{cov}(\mathbf{x}) \right]^2 \right\} \right] + o(N_0^{-2}). \end{aligned} \tag{40}$$

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

$$\begin{aligned} I(\mathbf{x}; \mathbf{Hx} + \mathbf{n}) &= \frac{\log e}{N_0} |\bar{h}|^2 \mathrm{var}[\mathbf{x}] \\ + \quad \frac{\log e}{2N_0^2} &\left[\gamma^4 E[||\mathbf{x}||^4] - (|\bar{h}|^2 + \gamma^2)^2 \sum_{i,j} |\mathrm{cov}(\mathrm{x}_i, \mathrm{x}_j)|^2 \right] + o(N_0^{-2}), \end{aligned} \tag{41}$$

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

$$I(\mathbf{x}; \mathbf{Hx} + \mathbf{n}) = \frac{\log e}{2N_0^2} \gamma^4 \left[\sum_{i,j} \mathrm{var}(\mathrm{x}_i \mathrm{x}_j^*) \right] + o(N_0^2). \tag{42}$$

Example 2: A channel with almost Gaussian noise.

Consider the channel (8) with almost Gaussian noise. In this case, $\mathbf{g}(\mathbf{x}) = \mathbf{x} + \mathbf{u}$ and it is easy to see that the general formula (18) has the form

$$I(\mathbf{x}; \mathbf{x} + \mathbf{u} + \mathbf{n}) = \frac{\log e}{N_0} \mathrm{var}[\mathbf{x}] - \frac{\log e}{2N_0^2} \left[\mathrm{trace}\left\{ \mathrm{cov}^2(\mathbf{x}) + 2\mathrm{cov}(\mathbf{x})\mathrm{cov}(\mathbf{u}) \right\} \right] + o(N_0^{-2}). \tag{43}$$

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

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

for a constant $\alpha > 0$. The proof of this assertion easily follows from Theorem 5 (see Section 4).

Example 3: Multiantenna Block-Fading Channel.

In the Marzetta-Hochwald channel model [12] 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 2.2. The resulting channel matrix $\mathbf{H}$ is $NT \times MT$, block diagonal with each block equal to the same $N \times M$ matrix $\mathbf{G}$ of propagation coefficients. For an arbitrary signaling scheme characterized by the joint distribution of the columns (transmitted vectors across the antennas in consecutive times)

$$[\mathbf{x}_1 \cdots \mathbf{x}_T] .$$

Denoting by $\mathbf{x}$ 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

$$\left(\mathrm{cov}(\mathbf{Hx}\,|\,\mathbf{x})\right)_{jk} = \zeta^2 \mathbf{x}_j^\dagger \mathbf{x}_k \tag{44}$$

Accordingly, the mutual information for a receiver which has no side information of the channel becomes

$$I(\mathbf{x}; \mathbf{Hx} + \mathbf{n}) = \frac{\zeta^4 \log e}{2N_0^2} \sum_{j,k} E[(\mathbf{x}_j^\dagger \mathbf{x}_k)^2] - E^2[(\mathbf{x}_j^\dagger \mathbf{x}_k)] + o(N_0^{-2}). \tag{45}$$

3.4 NonGaussian Noise

Consider the additive noise channel

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{46}$$

where now we assume that the noise components of $\mathbf{n}$ 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 nonGaussian 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

$$p(\mathsf{x}) = \frac{1}{N_0} \exp\left(-\frac{2}{\sqrt{N_0}} (|\Re \mathsf{x}| + |\Im \mathsf{x}|) \right). \tag{47}$$

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 $D(P_{Y|X=\mathbf{x}} \parallel P_{Y|X=0})$ which can be explicitly calculated in that case. Here, $P_{Y|X=\mathbf{x}}(\cdot)$ stands for the distribution of $\mathbf{y} = \mathbf{x} + \mathbf{n}$ conditioned on $\mathbf{x}$, and, in particular, $P_{Y|X=0}(\cdot)$ stands for the distribution of $\mathbf{n}$. 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*

$$\lim_{\mathbf{x}_0 \to 0} \frac{D(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0})}{||\mathbf{x}_0||^2} = a, \quad o < a < \infty, \tag{48}$$

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

$$D(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0}) \leq \left(a||\mathbf{x}_0||^2 - b||\mathbf{x}_0||^{2+\beta}\right) 1\left\{||\mathbf{x}_0|| \leq M\right\} + h||\mathbf{x}_0||^2 1\left\{||\mathbf{x}_0|| > M\right\} \tag{49}$$

Then

$$\mathcal{S}_0 = 0 \tag{50}$$

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

$$\sup_{\mathbf{x}_0} \frac{D(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0})}{||\mathbf{x}_0||^2} = \lim_{\mathbf{x}_0 \to 0} \frac{D(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0})}{||\mathbf{x}_0||^2} = a, \tag{51}$$

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

$$\dot{C}(0) = N_0 \sup_{\mathbf{x}_0} \frac{D(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0})}{||\mathbf{x}_0||^2}. \tag{52}$$

4 Proofs

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

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

$$E[||\mathbf{z}_\varepsilon||^{4+\alpha}] \le \left(\log \frac{1}{\varepsilon}\right)^\mu, \tag{53}$$

then

$$\begin{aligned} I(\varepsilon \mathbf{z}_\varepsilon; \varepsilon \mathbf{z}_\varepsilon + \mathbf{w}) &= I(\varepsilon \widehat{\mathbf{z}}_\varepsilon;\, \varepsilon \widehat{\mathbf{z}}_\varepsilon + \mathbf{w}) + o(\varepsilon^4) \\ &= \frac{\log e}{2\sigma^2} \mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}_\varepsilon)\right\} \varepsilon^2 \\ &- \frac{\log e}{4\sigma^4} \mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}_\varepsilon)\right\} \varepsilon^4 + o(\varepsilon^4), \quad \varepsilon \to 0, \end{aligned} \tag{54}$$

where $\widehat{\mathbf{z}}_\varepsilon$ *is a Gaussian random vector with the same covariance matrix as* $\mathbf{z}_\varepsilon$ *and* $\dfrac{o(\varepsilon^4)}{\varepsilon^4} \to 0$ *as* $\varepsilon \to 0$ *uniformly with respect to all random vectors* $\mathbf{z}_\varepsilon$ *satisfying* (53)*.*

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.

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

$$I(\mathbf{x}\,;\, \mathbf{g}(\mathbf{x}) + \mathbf{n}) = I(\mathbf{x}^{(r)}\,;\, \mathbf{g}^{(r)}(\mathbf{x}) + \mathbf{n}^{(r)}), \tag{55}$$

where $\mathbf{x}^{(r)} = \begin{bmatrix} \Re \mathbf{x} \\ \Im \mathbf{x} \end{bmatrix}$ denotes the real-valued vector of length $2n$ generated by real and imaginary parts of $\mathbf{x}$. The real-valued vectors $\mathbf{g}^{(r)}(\mathbf{x}) = \begin{bmatrix} \Re \mathbf{g}(\mathbf{x}) \\ \Im \mathbf{g}(\mathbf{x}) \end{bmatrix}$ and $\mathbf{n}^{(r)} = \begin{bmatrix} \Re \mathbf{n} \\ \Im \mathbf{n} \end{bmatrix}$ of length $2m$ are similarly defined. Therefore, it is clear that

$$I(\mathbf{x}\,;\, \mathbf{g}(\mathbf{x}) + \mathbf{n}) = I(\mathbf{g}^{(r)}(\mathbf{x})\,;\, \mathbf{g}^{(r)}(\mathbf{x}) + \mathbf{n}^{(r)}) - I(\mathbf{g}^{(r)}(\mathbf{x})\,;\, \mathbf{g}^{(r)}(\mathbf{x}) + \mathbf{n}^{(r)} |\, \mathbf{x}^{(r)}). \tag{56}$$

Let

$$\mathbf{z}(\mathbf{x}^{(r)}) \stackrel{\mathrm{def}}{=} \mathbf{g}^{(r)}(\mathbf{x}), \qquad \mathbf{w} \stackrel{\mathrm{def}}{=} \varepsilon \mathbf{n}^{(r)} \tag{57}$$

where

$$\varepsilon \stackrel{\mathrm{def}}{=} \sqrt{\frac{2}{N_0}}. \tag{58}$$

Then we can rewrite (56) as

$$I(\mathbf{x}\,;\, \mathbf{g}(\mathbf{x}) + \mathbf{n}) = I(\varepsilon \mathbf{z}(\mathbf{x}^{(r)})\,;\, \varepsilon \mathbf{z}(\mathbf{x}^{(r)}) + \mathbf{w}) - I(\varepsilon \mathbf{z}(\mathbf{x}^{(r)})\,;\, \varepsilon \mathbf{z}(\mathbf{x}^{(r)}) + \mathbf{w}\,|\, \mathbf{x}^{(r)}), \tag{59}$$

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

Condition (15) implies that

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

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

$$\begin{aligned} I(\varepsilon\mathbf{z}(\mathbf{x}^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}^{(r)})+\mathbf{w}) &= \frac{\log e}{2}\text{trace}\left\{\text{cov}(\mathbf{z}(\mathbf{x}^{(r)}))\right\}\varepsilon^2 \\ &- \frac{\log e}{4}\text{trace}\left\{\text{cov}^2(\mathbf{z}(\mathbf{x}^{(r)}))\right\}\varepsilon^4 + o(\varepsilon^4), \quad \varepsilon \to 0. \end{aligned} \tag{60}$$

Now note that we cannot directly apply Theorem 5 to write out the asymptotics for the second term of the right-hand side of (56) because, in general, condition (53) is not uniformly satisfied for all values of $\mathbf{x}^{(r)}$. Therefore, to obtain the asymptotic expression for the second term of (56), let us represent $I(\mathbf{g}(\mathbf{x})\,;\,\varepsilon\mathbf{g}(\mathbf{x})+\mathbf{n}\,|\,\mathbf{x})$ as

$$I(\varepsilon\mathbf{z}(\mathbf{x}^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}^{(r)})+\mathbf{w}\,|\,\mathbf{x}^{(r)}) = I_1 + I_2, \tag{61}$$

where

$$I_1 = \int\limits_{A_\varepsilon} I(\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})+\mathbf{w})P_X(d\mathbf{x}_0^{(r)}) \tag{62}$$

and

$$I_2 = \int\limits_{\overline{A}_\varepsilon} I(\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})+\mathbf{w})P_X(d\mathbf{x}_0^{(r)}). \tag{63}$$

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

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

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

It follows from condition (16) that for almost all $\mathbf{x}_0^{(r)} \in A_\varepsilon$ we have

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

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

$$I_1 = \int\limits_{A_\varepsilon}\left[\frac{\log e}{2}\text{trace}\left\{\text{cov}(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\}\varepsilon^2 - \frac{\log e}{4}\text{trace}\left\{\text{cov}^2(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\}\varepsilon^4 + o(\varepsilon^4)\right]P_X(d\mathbf{x}_0^{(r)}), \tag{64}$$

where $o(\varepsilon^4)/\varepsilon^4 \to 0$ as $\varepsilon \to 0$ uniformly in all $\mathbf{x}_0^{(r)} \in A_\varepsilon$. Using (64) we intend to show that

$$\begin{aligned} I_1 &= \frac{\log e}{2}\text{trace}\left\{E[\text{cov}(\mathbf{z}(\mathbf{x}^{(r)})|\mathbf{x}^{(r)})]\right\}\varepsilon^2 \\ &- \frac{\log e}{4}\text{trace}\left\{E[\text{cov}^2(\mathbf{z}(\mathbf{x}^{(r)})|\mathbf{x}^{(r)})]\right\}\varepsilon^4 + o(\varepsilon^4), \quad \varepsilon \to 0. \end{aligned} \tag{65}$$

To this end it suffices to show that

$$\overline{I}_{11} \stackrel{\text{def}}{=} \int\limits_{\overline{A}_\varepsilon} \text{trace}\left\{\text{cov}(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\} P_X(d\mathbf{x}_0^{(r)}) = o(\varepsilon^2), \quad \varepsilon \to 0, \tag{66}$$

and

$$\overline{I}_{12} \stackrel{\text{def}}{=} \int\limits_{\overline{A}_\varepsilon} \text{trace}\left\{\text{cov}^2(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\} P_X(d\mathbf{x}_0^{(r)}) = o(1), \quad \varepsilon \to 0. \tag{67}$$

First observe that

$$\text{trace}\left\{\text{cov}(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\} = \text{trace}\left\{\text{cov}(\mathbf{g}(\mathbf{x}_0))\right\} = E[||\mathbf{g}(\mathbf{x}_0) - E[\mathbf{g}(\mathbf{x}_0)]||^2] \tag{68}$$

and therefore condition (16) implies that [5]

$$\text{trace}\left\{\text{cov}(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\} \le E[||\mathbf{g}(\mathbf{x}_0)||^2] \le \left(E[||\mathbf{g}(\mathbf{x}_0)||^{4+\alpha}]\right)^{\frac{2}{4+\alpha}} \le c_1||\mathbf{x}_0||^{k_1} = c_1||\mathbf{x}_0^{(r)}||^{k_1} \tag{69}$$

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

$$\overline{I}_{11} \le c_1 \int\limits_{\overline{A}_\varepsilon} ||\mathbf{x}_0^{(r)}||^{k_1} P_X(d\,\mathbf{x}_0^{(r)}). \tag{70}$$

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

$$\text{trace}\left\{\text{cov}^2(\mathbf{z}(\mathbf{x}_0^{(r)}))\right\} = \frac{1}{2}\text{trace}\left\{\text{cov}^2(\mathbf{g}(\mathbf{x}_0))\right\} \tag{71}$$

if $\mathbf{g}(\mathbf{x}_0)$ is a proper complex random vector. Therefore, it follows from (71), condition (17), and equality (67) that

$$\overline{I}_{12} \le c_2 \int\limits_{\overline{A}_\varepsilon} ||\mathbf{x}_0^{(r)}||^{l} P_X(d\,\mathbf{x}_0^{(r)}). \tag{72}$$

It is well known that condition (14) implies

$$\int\limits_{||\mathbf{x}_0^{(r)}||>\delta} ||\mathbf{x}_0^{(r)}||^{t} P_X(d\,\mathbf{x}_0^{(r)}) \le \delta^M \exp\{-\delta^\nu\} \tag{73}$$

for any $t \ge 0$ (where $M > 0$ is a constant depending on t) and for all sufficiently large δ. Using (73) we can easily derive from (70) and (72) 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 $\int ||\mathbf{x}_0^{(r)}||^l P_X(d\,\mathbf{x}_0^{(r)})$ converges). Thus, equalities (66) and (67) and therefore (65) are proved.

Now, let us show that

$$I_2 = o(\varepsilon^4), \quad \varepsilon \to 0, \tag{74}$$

[5]Here and throughout below $c, c_1, ...$ denote finite constants, perhaps, different in different inequalities

where I_2 is defined in (63). Note that for $\mathbf{x}_0^{(r)} \in \overline{A}_\varepsilon$, we cannot apply the statement of Theorem 5 to write out the asymptotic expression for $I(\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)}) + \mathbf{w})$ in form of (60). But we can easily estimate I_2 taking into account that

$$I(\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)})\,;\,\varepsilon\mathbf{z}(\mathbf{x}_0^{(r)}) + \mathbf{w}) \le I(\varepsilon\widehat{\mathbf{z}}(\mathbf{x}_0^{(r)}); \varepsilon\widehat{\mathbf{z}}(\mathbf{x}_0^{(r)}) + \mathbf{w}), \tag{75}$$

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

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

(where λ_i^0, $i = 1, \dots, 2m$, are eigenvalues of $\mathrm{cov}(\mathbf{z}(\mathbf{x}_0^{(r)}))$), the required equality (74) follows from (75) and the already proved equality (66).

Thus, from (56), (60), (61), (65), and (74) we obtain

$$\begin{aligned} I(\mathbf{x}\,;\,\mathbf{g}(\mathbf{x}) + \mathbf{n}) \;&=\; \frac{\log e}{2}\left[\mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}(\mathbf{x}^{(r)}))\right\} - \mathrm{trace}\left\{E[\mathrm{cov}(\mathbf{z}(\mathbf{x}^{(r)})\,|\,\mathbf{x}^{(r)})]\right\}\right]\varepsilon^2 \\ &+\; \frac{\log e}{4}\left[\mathrm{trace}\left\{E[\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)})\,|\,\mathbf{x}^{(r)})]\right\} - \mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)}))\right\}\right]\varepsilon^4 \\ &+\; o(\varepsilon^4), \quad \varepsilon \to 0. \end{aligned} \tag{76}$$

It can easily be seen that

$$\mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}(\mathbf{x}^{(r)}))\right\} - \mathrm{trace}\left\{E[\mathrm{cov}(\mathbf{z}(\mathbf{x}^{(r)})|\mathbf{x}^{(r)})]\right\} = \mathrm{trace}\left\{\mathrm{cov}(\overline{\mathbf{z}}(\mathbf{x}^{(r)}))\right\}, \tag{77}$$

where

$$\overline{\mathbf{z}}(\mathbf{x}^{(r)}) \stackrel{\mathrm{def}}{=} E[\mathbf{z}(\mathbf{x}^{(r)})\,|\,\mathbf{x}^{(r)}].$$

Hence,

$$\begin{aligned} I(\mathbf{x}\,;\,\mathbf{g}(\mathbf{x}) + \mathbf{n}) \;&=\; \frac{\log e}{2}\left[\mathrm{trace}\left\{\mathrm{cov}(\overline{\mathbf{z}}(\mathbf{x}^{(r)}))\right\}\right]\varepsilon^2 \\ &+\; \frac{\log e}{4}\left[\mathrm{trace}\left\{E[\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)})\,|\,\mathbf{x}^{(r)})]\right\} - \mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)}))\right\}\right]\varepsilon^4 \\ &+\; o(\varepsilon^4), \quad \varepsilon \to 0. \end{aligned} \tag{78}$$

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

$$\mathrm{trace}\left\{\mathrm{cov}(\overline{\mathbf{z}}(\mathbf{x}^{(r)}))\right\} = \mathrm{trace}\left\{\mathrm{cov}(\overline{\mathbf{g}}(\mathbf{x}))\right\} \tag{79}$$

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

$$\mathrm{trace}\left\{E[\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)})\,|\,\mathbf{x}^{(r)})]\right\} = \frac{1}{2}\mathrm{trace}\left\{E[\mathrm{cov}^2(\mathbf{g}(\mathbf{x})\,|\,\mathbf{x})]\right\} \tag{80}$$

and

$$\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}(\mathbf{x}^{(r)}))\right\} = \frac{1}{2}\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{g}(\mathbf{x}))\right\} \tag{81}$$

are valid if $\mathbf{g}(\mathbf{x}_0)$ is proper for any fixed $\mathbf{x}_0$ and $\mathbf{g}(\mathbf{x})$ is proper, respectively (inequality (80) was already mentioned above (see (71)). Substituting $\varepsilon^2 = 2/N_0$ into (78) and taking into account equalities (79)–(81), we obtain the required equality (18). □

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

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

To see that $\mathbf{Hx}$ is proper complex note that

$$E[(\mathbf{Hx}-\bar{\mathbf{H}}\bar{\mathbf{x}})(\mathbf{Hx}-\bar{\mathbf{H}}\bar{\mathbf{x}})^T] = \bar{\mathbf{H}}E[(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T]\bar{\mathbf{H}}^T + E[(\mathbf{H}-\bar{\mathbf{H}})\mathbf{x}\mathbf{x}^T(\mathbf{H}-\bar{\mathbf{H}})^T]$$

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

In addition we need to show that for fading channels (4) conditions (19) and (14) imply conditions (15)–(17).

Indeed, first note that, as follows from condition (14), all the moments of $\mathbf{x}$ exist and are finite. Therefore, by independence of $\mathbf{H}$ and $\mathbf{x}$ and by condition (19) we easily obtain

$$E[||\mathbf{Hx}||^{4+\alpha}] \le cE[||\mathbf{H}||^{4+\alpha}]E[||\mathbf{x}||^{4+\alpha}] < c_1$$

and hence condition (15) holds. Moreover, it is also follows from the inequality above that

$$E[||\mathbf{Hx}||^{4+\alpha}|\mathbf{x}] \le c||\mathbf{x}||^{4+\alpha} \quad a.s.,$$

i.e., condition (16) is also satisfied.

Finally, one can easily verify that

$$\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{Hx}|\mathbf{x})\right\} \le c||\mathbf{x}||^4 \quad a.s.$$

and, therefore, condition (17) is fulfilled. To complete the proof of the theorem we only need to note that the equality

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

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

Proof of Theorem 3. First of all, we have

$$I(\mathbf{x}\,;\,\mathbf{Hx}+\mathbf{n}) = I(\mathbf{Hx}\,;\,\mathbf{Hx}+\mathbf{n}) = I(\varepsilon(\mathbf{Hx})^{(r)}\,;\,\varepsilon(\mathbf{Hx})^{(r)}+\mathbf{w}\,|\,\mathbf{H}^{(r)}) \tag{82}$$

where the quantities

$$(\mathbf{Hx})^{(r)} = \begin{bmatrix} \Re(\mathbf{Hx}) \\ \Im(\mathbf{Hx}) \end{bmatrix}, \quad \mathbf{H}^{(r)} = \begin{bmatrix} \Re\mathbf{H} \\ \Im\mathbf{H} \end{bmatrix}, \quad \mathbf{w} = \varepsilon\mathbf{n}^{(r)} = \varepsilon \begin{bmatrix} \Re\mathbf{n} \\ \Im\mathbf{n} \end{bmatrix}, \quad \varepsilon = \sqrt{\frac{2}{N_0}} \tag{83}$$

are defined similar to $\mathbf{x}^{(r)}$, $\mathbf{g}^{(r)}(\mathbf{x})$, etc. in the proof of Theorem 1 (cf. (57) and (58)). Note now that the right-hand side of (82) has the same form as the second term of the right-hand side of (59) (but now we have $(\mathbf{Hx})^{(r)}$ and $\mathbf{H}^{(r)}$ instead of $\mathbf{z}(\mathbf{x}^{(r)})$ and $\mathbf{x}^{(r)}$, respectively). Therefore the same reasoning as was used in the proof of Theorem 1 shows that, as $\varepsilon \to 0$, we have

$$\begin{aligned} I(\varepsilon(\mathbf{Hx})^{(r)}\,;\,\varepsilon(\mathbf{Hx})^{(r)} + \mathbf{w}\,|\,\mathbf{H}^{(r)}) &= \frac{\log e}{2}\mathrm{trace}\left\{E[\mathrm{cov}((\mathbf{Hx})^{(r)}\,|\,\mathbf{H}^{(r)})]\right\}\varepsilon^2 \\ &- \frac{\log e}{4}\mathrm{trace}\left\{E[\mathrm{cov}^2((\mathbf{Hx})^{(r)}\,|\,\mathbf{H}^{(r)})]\right\}\varepsilon^4 + o(\varepsilon^4) \end{aligned} \tag{84}$$

if $\mathbf{x}$ is proper, the matrix $\mathbf{H}$ satisfies condition (27), and

$$E[||\mathbf{Hx}||^{4+\alpha}] < c, \tag{85}$$

$$E[||\mathbf{Hx}||^{4+\alpha}|\mathbf{H}] \le c||\mathbf{H}||^k \quad a.s., \tag{86}$$

$$\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{Hx}|\mathbf{H})\right\} \le c||\mathbf{H}||^l \quad a.s. \tag{87}$$

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 (85) – (87) are fulfilled if $\mathbf{x}$ and $\mathbf{H}$ satisfy conditions (26) and (27).

To complete the proof of Theorem 3 we only need to verify that (84) implies (28). Indeed, one can check by direct calculations that

$$\mathrm{trace}\left\{E[\mathrm{cov}((\mathbf{Hx})^{(r)}\,|\,\mathbf{H}^{(r)})]\right\} = \mathrm{trace}\left\{E[\mathrm{cov}(\mathbf{Hx}\,|\,\mathbf{H})]\right\} = \mathrm{trace}\left\{E[\mathbf{H}\mathrm{cov}(\mathbf{x})\mathbf{H}^\dagger]\right\} \tag{88}$$

and

$$\begin{aligned} \mathrm{trace}\left\{E[\mathrm{cov}^2((\mathbf{Hx})^{(r)}\,|\,\mathbf{H}^{(r)})]\right\} &= \frac{1}{2}\mathrm{trace}\left\{E[\mathrm{cov}^2(\mathbf{Hx}\,|\,\mathbf{H})]\right\} \\ &= \frac{1}{2}\mathrm{trace}\left\{E[(\mathbf{H}\mathrm{cov}(\mathbf{x})\mathbf{H}^\dagger)^2]\right\} \end{aligned} \tag{89}$$

since $\mathbf{x}$ is proper (cf. (79) and (81)). Substituting $\varepsilon^2 = 2/N_0$ and the right-hand side expressions from (88) and (89) instead of the left-hand side ones into (84), we get the required formula (28). $\square$

Proof of Theorem 4. Taking into account expression (1) for $\mathcal{S}_0$, we see that we need to show that $\ddot{C} = -\infty$.

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

$$\begin{aligned}
I(\mathbf{x};\mathbf{y}) - m\dot{C}(0)\rho &= D\left(P_{Y|X=\mathbf{x}} \parallel P_{Y|X=0}|P_{\mathbf{x}}\right) - D\left(P_Y \parallel P_{Y|X=0}\right) - amN_0\rho \\
&\le D\left(P_{Y|X=\mathbf{x}} \parallel P_{Y|X=0}|P_{\mathbf{x}}\right) - aE[||\mathbf{x}||^2] \\
&= \int\limits_{\mathcal{R}^n} \left[D\left(P_{Y|X=\mathbf{x}_0} \parallel P_{Y|X=0}\right) - a||\mathbf{x}_0||^2\right] P_X(d\,\mathbf{x}_0) \\
&= I' + I'', \qquad (90)
\end{aligned}$$

where

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

and

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

Using conditions (i) and (ii) we easily obtain

$$I' \le -b \int\limits_{||\mathbf{x}_0|| \le M} ||\mathbf{x}_0||^{2+\beta} P_X(d\,\mathbf{x}_0) \le -b \left[\int\limits_{||\mathbf{x}_0|| \le M} ||\mathbf{x}_0||^2 P_X(d\,\mathbf{x}_0)\right]^{\frac{2+\beta}{2}} \qquad (91)$$

and

$$I'' \le -(a-h) \int\limits_{||\mathbf{x}_0|| > M} ||\mathbf{x}_0||^2 P_X(d\,\mathbf{x}_0). \qquad (92)$$

Set

$$q_1^2 \stackrel{\text{def}}{=} \int\limits_{||\mathbf{x}_0|| \le M} ||\mathbf{x}_0||^2 P_X(d\,\mathbf{x}_0)$$

and

$$q_2^2 \stackrel{\text{def}}{=} \int\limits_{||\mathbf{x}_0|| > M} ||\mathbf{x}_0||^2 P_X(d\,\mathbf{x}_0).$$

It follows from (90) – (92) that

$$\sup_{\mathbf{x}:\, E[||\mathbf{x}||^2] \le mN_0\rho} I(\mathbf{x};\mathbf{y}) - m\dot{C}(0)\rho \le -\inf_{q_1, q_2} \left[b q_1^{2+\beta} + (a-h) q_2^2\right], \qquad (93)$$

where the infimum on the right-hand side of (93) 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

$$\sup_{\mathbf{x}:\, E[||\mathbf{x}||^2] \le mN_0\rho} I(\mathbf{x};\mathbf{y}) - m\dot{C}(0)\rho \le -b(mN_0\rho)^{(2+\beta)/2}. \qquad (94)$$

From (94) 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.

□

5 Appendix

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 $\mathbf{Q} = \mathbf{Q}_\varepsilon$ be the matrix of an orthogonal transformation in $\mathcal{R}^l$ such that

$$\mathbf{Q}\operatorname{cov}(\mathbf{z}_\varepsilon)\mathbf{Q}^\dagger = \lambda\mathbf{I},$$

where $\lambda = \lambda_\varepsilon = (\lambda_{1\varepsilon}, \ldots, \lambda_{l\varepsilon})$ is a vector whose components are eigenvalues of the covariance matrix $\operatorname{cov}(\mathbf{z}_\varepsilon)$. Note that

$$\mathbf{Q}\operatorname{cov}(\mathbf{w})\mathbf{Q}^\dagger = \operatorname{cov}(\mathbf{w})$$

since $\operatorname{cov}(\mathbf{w})$ is a diagonal matrix. In other words, the random vector $\mathbf{Q}\mathbf{z}_\varepsilon$ has uncorrelated components and, moreover,

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

as obviously follows from (53). It is well known that any orthogonal transformation of random vectors does not change the mutual information, i.e.,

$$I(\varepsilon\mathbf{Q}\mathbf{z}_\varepsilon; \varepsilon\mathbf{Q}\mathbf{z}_\varepsilon + \mathbf{Q}\mathbf{w}) = I(\varepsilon\mathbf{z}_\varepsilon; \varepsilon\mathbf{z}_\varepsilon + \mathbf{w}).$$

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

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

Lemma 2. *Under conditions of Theorem* 2 *the probability density function $q_\varepsilon(\mathbf{t})$ of the random vector $\varepsilon\mathbf{z}_\varepsilon + \mathbf{w}$ can be represented as*

$$q_\varepsilon(\mathbf{t}) = p(\mathbf{t}) + \sum_{j=2}^{4}\frac{(-1)^j\varepsilon^j}{j!}E\left[\left(\frac{\partial}{\partial t_1}\mathsf{z}_{1\varepsilon} + \ldots + \frac{\partial}{\partial t_l}\mathsf{z}_{l\varepsilon}\right)^j p(\mathbf{t})\right] + r_\varepsilon(\mathbf{t}), \tag{95}$$

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

$$|r_\varepsilon(\mathbf{t})| \le c\varepsilon^{4+\gamma} \quad \textit{for some} \quad \gamma > 0, \tag{96}$$

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

Proof of Lemma 2. Formula (95) follows from the representation

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

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

$$r_\varepsilon(\mathbf{t}) = \frac{\varepsilon^4}{24}\int_{\mathcal{R}^l}\left[\left(\frac{\partial}{\partial t_1}u_1 + \ldots + \frac{\partial}{\partial t_l}u_l\right)^4 p(\mathbf{t} - \theta\varepsilon\mathbf{u}) - \left(\frac{\partial}{\partial t_1}u_1 + \ldots + \frac{\partial}{\partial t_l}u_l\right)^4 p(\mathbf{t})\right] dF_\varepsilon(\mathbf{u}). \tag{97}$$

Here $F_\varepsilon(\cdot)$ denotes the distribution function of $\mathbf{z}_\varepsilon$. Therefore, to prove the lemma we only need to show that $r_\varepsilon(\mathbf{t})$ satisfies inequality (96).

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

$$\Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u})) \triangleq \left| \frac{\partial^4 p}{\partial t_{i_1} \ldots \partial t_{i_4}}(\mathbf{t} - \theta\varepsilon\mathbf{u}) - \frac{\partial^4 p}{\partial t_{i_1} \ldots \partial t_{i_4}}(\mathbf{t}) \right| \tag{98}$$

and upper bound the integrals

$$I_{\mathbf{i}}(k) = \int\limits_{\mathcal{R}^l} u_k^4 \Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u}))dF_\varepsilon(\mathbf{u}). \tag{99}$$

It is clear that

$$I_{\mathbf{i}}(k) \leq \sum_{j=0}^{l} I_{\mathbf{i}}(j,k), \tag{100}$$

where

$$I_{\mathbf{i}}(0,k) = \int\limits_{\left\{|u_j| \leq \varepsilon^{\beta-1},\ j=1,\ldots,l\right\}} u_k^4 \Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u}))dF_\varepsilon(\mathbf{u}),$$

$$I_{\mathbf{i}}(j,k) = \int\limits_{\left\{|u_j| \geq \varepsilon^{\beta-1},\ |u_k| \leq \varepsilon^{\beta-1}\right\}} u_k^4 \Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u}))dF_\varepsilon(\mathbf{u}), \quad \text{if} \quad j \neq k,$$

and

$$I_{\mathbf{i}}(k,k) = \int\limits_{\left\{|u_k| \geq \varepsilon^{\beta-1}\right\}} u_k^4 \Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u}))dF_\varepsilon(\mathbf{u}),$$

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

Now we have

$$\begin{aligned} I_{\mathbf{i}}(0,k) &\leq c\varepsilon^\beta \int\limits_{\mathcal{R}^l} u_k^4 dF_\varepsilon(\mathbf{u}) \\ &\leq c\varepsilon^\beta \left(E[||\mathbf{z}_\varepsilon||^{4+\alpha}] \right)^{\frac{4}{4+\alpha}} \\ &\leq c\varepsilon^\beta \left(\log \frac{1}{\varepsilon} \right)^{\frac{4\mu}{4+\alpha}} \leq c\varepsilon^{\beta'}, \end{aligned} \tag{101}$$

where $\beta' > 0$ is a constant such that $\beta' < \beta$. Here we have used the fact that $\Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u})) \leq c\varepsilon||\mathbf{u}||$, the well-known inequalities between moments of random variables, and condition (53) of Theorem 2.

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

$$\begin{aligned} I_{\mathbf{i}}(j,k) &\leq c\varepsilon^{4(\beta-1)} P\left\{ ||\mathbf{z}_\varepsilon|| > \varepsilon^{\beta-1} \right\} \\ &\leq c\varepsilon^{4(\beta-1)} E[||\mathbf{z}_\varepsilon||^{4+\alpha}] \varepsilon^{-(\beta-1)(4+\alpha)} \\ &\leq c\varepsilon^{\alpha-\alpha\beta} \left(\frac{1}{\varepsilon} \right)^\mu \leq c\varepsilon^{\alpha'}, \end{aligned} \tag{102}$$

where α' is a constant such that $\alpha' < \alpha - \alpha\beta$ (note that $< \alpha - \alpha\beta > 0$ if $\beta < 1$). To get (102) we have used inequality $\Delta_{\mathbf{i}}(\mathbf{t}, \theta\varepsilon\mathbf{u})) <$ const $< \infty$ for all $\mathbf{t}$ and $\mathbf{u}$, the Chebyshev's inequality, and condition (53).

Finally, we have

$$\begin{aligned} I_{\mathbf{i}}(k,k) &\le c\varepsilon^{\beta} \int\limits_{\{||u_k|>\varepsilon^{\beta-1}|\}} u_k^4 dF_\varepsilon(\mathbf{u}) \\ &= c\varepsilon^{4(\beta-1)} \int\limits_{\{||u_k|>\varepsilon^{\beta-1}|\}} \left(\frac{u_k}{\varepsilon^{\beta-1}}\right)^4 dF_\varepsilon(\mathbf{u}) \\ &\le c\varepsilon^{\alpha-\alpha\beta} \int\limits_{\mathcal{R}^l} ||\mathbf{u}||^{4+\alpha} dF_\varepsilon(\mathbf{u}) \\ &\le c\varepsilon^{\alpha-\alpha\beta} \left(\log\frac{1}{\varepsilon}\right)^{\mu} \le c\varepsilon^{\alpha'}. \end{aligned} \tag{103}$$

Now, it follows from (100)–(103) that

$$0 \le I_{\mathbf{i}}(k) \le c\varepsilon^{\gamma}, \tag{104}$$

for all $\mathbf{i}$ and k where $\gamma > 0$ is a constant. Taking into account (104), (97)–(99), and the fact that $|u_{i_1}\cdots u_{i_4}| \le \sum_{k=1}^{l} u_k^4$ for any $i_1, ..., i_4 \in \{1, ..., l\}$, we arrive at the conclusion that $r_\varepsilon(\mathbf{t})$ satisfies (96). This completes the proof of Lemma 2. □

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

$$G_{\varepsilon,\kappa} \stackrel{\text{def}}{=} \left\{\mathbf{t} \in \mathcal{R}^l : \exp\left[-\frac{||\mathbf{t}||^2}{2\sigma^2}\right] \ge \varepsilon^{4+\kappa}\right\}, \quad \overline{G}_{\varepsilon,\kappa} \stackrel{\text{def}}{=} \left\{\mathbf{t} \in \mathcal{R}^l : \exp\left[-\frac{||\mathbf{t}||^2}{2\sigma^2}\right] < \varepsilon^{4+\kappa}\right\}, \tag{105}$$

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

$$a_\varepsilon(\mathbf{t}) \triangleq \sum_{j=2}^{4} \frac{(-1)^j\varepsilon^j}{j!} E\left[\left(\frac{\partial}{\partial t_1}\mathsf{z}_{1\varepsilon} + \ldots + \frac{\partial}{\partial t_n}\mathsf{z}_{l\varepsilon}\right)^j p(\mathbf{t})\right]. \tag{106}$$

It follows from Lemma 2, definition of $\overline{G}_{\varepsilon,\kappa}$, and condition (54), that

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

for all sufficiently small ε. Therefore, using formulas (95), (106), we easily get (throughout below we will omit arguments $\mathbf{t}$ and ε in the notations of $a_\varepsilon(\mathbf{t})$, $r_\varepsilon(\mathbf{t})$, and $p(\mathbf{t})$)

$$I(\varepsilon\mathbf{z}_\varepsilon; \varepsilon\mathbf{z}_\varepsilon + \mathbf{w}) \;\ge\; -\int\limits_{G_{\varepsilon,\kappa}} (a+r)\log p\, d\mathbf{t} + \int\limits_{\overline{G}_{\varepsilon,\kappa}} p\log p\, d\mathbf{t}$$

$$
\begin{aligned}
& - \quad \log e \left[\int\limits_{G_{\varepsilon,\kappa}} (p+a+r) \ln \left(1 + \frac{a+r}{p} \right) d\mathbf{t} \right] \\
& = \quad - \int\limits_{\mathcal{R}^l} a \log p \, d\mathbf{t} + \int\limits_{\overline{G}_{\varepsilon,\kappa}} p \log p \, d\mathbf{t} + \int\limits_{\overline{G}_{\varepsilon,\kappa}} a \log p \, d\mathbf{t} - \int\limits_{G_{\varepsilon,\kappa}} r \log p \, d\mathbf{t} \\
& - \quad \log e \left[\int\limits_{G_{\varepsilon,\kappa}} (a+r) \, d\mathbf{t} + \sum_{j=2}^{\infty} \frac{(-1)^j}{(j-1)j} \int\limits_{G_{\varepsilon,\kappa}} \frac{(a+r)^j}{p^{j-1}} \, d\mathbf{t} \right]. \qquad (107)
\end{aligned}
$$

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

$$
\int\limits_{\overline{G}_{\varepsilon,\kappa}} p \log p \, d\mathbf{t} = o(\varepsilon^4), \quad \int\limits_{\overline{G}_{\varepsilon,\kappa}} a \log p \, d\mathbf{t} = o(\varepsilon^4), \quad \varepsilon \to 0, \qquad (108)
$$

$$
\int\limits_{G_{\varepsilon,\kappa}} r \log p \, d\mathbf{t} = o(\varepsilon^4), \quad \int\limits_{G_{\varepsilon,\kappa}} (a+r) \, d\mathbf{t} = o(\varepsilon^4), \quad \varepsilon \to 0, \qquad (109)
$$

and

$$
\sum_{j=3}^{\infty} \frac{(-1)^j}{(j-1)j} \int\limits_{G_{\varepsilon,\kappa}} \frac{(a+r)^j}{p^{j-1}} \, d\mathbf{t} = o(\varepsilon^4), \quad \varepsilon \to 0. \qquad (110)
$$

Proof of Lemma 3. The proof equalities (108)–(110) is based on Lemma 2, properties of the Gaussian density function $p(\mathbf{t})$), and the definitions of the sets $G_{\varepsilon,\kappa}$ and $\overline{G}_{\varepsilon,\kappa}$.

The first equality in (108) follows from rather obvious relations

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

where δ is defined by the equality

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

The proof of the second equality in (108) is similar to the first one if we note that

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

Equality (109) can be proved by the same method as (108) taking into account that

$$
\int\limits_{G_{\varepsilon,\kappa}} a \, d\mathbf{t} = \int\limits_{\overline{G}_{\varepsilon,\kappa}} a \, d\mathbf{t}.
$$

The proof of (110) is a little more complicated. First of all note that

$$\left|\frac{a}{p^{(j-1)/j}}\right| \le c\varepsilon^2 \log^2\frac{1}{\varepsilon}, \quad \mathbf{t}\in G_{\varepsilon,\kappa}$$

and

$$\left|\frac{r}{p^{(j-1)/j}}\right| \le c\frac{\varepsilon^{4+\gamma}}{\varepsilon^{(4+\kappa)(j-1)/j}} = c\varepsilon^{\frac{4+\kappa}{j}}\varepsilon^{\gamma-\kappa}, \quad \mathbf{t}\in G_{\varepsilon,\kappa}$$

since it is clear that

$$|a| \le c\varepsilon^2\left(\log\frac{1}{\varepsilon}\right)p\,, \quad \mathbf{t}\in G_{\varepsilon,\kappa}.$$

Therefore, we get

$$\left|\frac{(a+r)^j}{p^{j-1}}\right| \le \left[c\varepsilon^{\frac{4+\kappa}{j}}\left(\varepsilon^{2-\frac{4+\kappa}{j}}\log^2\frac{1}{\varepsilon}+\varepsilon^{\gamma-\kappa}\right)\right]^j \le \varepsilon^{4+\kappa}\cdot o(1), \quad \mathbf{t}\in G_{\varepsilon,\kappa} \tag{111}$$

if $j\ge 3$ and κ is sufficiently small. Now equality (111) implies (110). Lemma 3 is proved. □

Lemma 3 and inequality (107) imply

$$I(\varepsilon\mathbf{z}_\varepsilon;\varepsilon\mathbf{z}_\varepsilon+\mathbf{w}) \ge -\int\limits_{\mathcal{R}^l} a\log p\,d\mathbf{t} - \frac{\log e}{2}\int\limits_{G_{\varepsilon,\kappa}}\frac{(a+r)^2}{p}\,d\mathbf{t}+o(\varepsilon^4), \quad \varepsilon\to 0. \tag{112}$$

By direct calculations one can verify that

$$-\int\limits_{\mathcal{R}^l} a\log p\,d\mathbf{t} = \frac{\log e}{2\sigma^2}\mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}_\varepsilon)\right\}\varepsilon^2. \tag{113}$$

Furthermore, we have

$$\int\limits_{G_{\varepsilon,\kappa}}\frac{(a+r)^2}{p}\,d\mathbf{t} = \int\limits_{\mathcal{R}^l}\frac{a^2}{p}\,d\mathbf{t} - \int\limits_{\overline{G}_{\varepsilon,\kappa}}\frac{a^2}{p}\,d\mathbf{t} + 2\int\limits_{G_{\varepsilon,\kappa}}\frac{ar}{p}\,d\mathbf{t} + \int\limits_{G_{\varepsilon,\kappa}}\frac{r^2}{p}\,d\mathbf{t}. \tag{114}$$

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

$$\int\limits_{\overline{G}_{\varepsilon,\kappa}}\frac{a^2}{p}\,d\mathbf{t} = o(\varepsilon^4), \quad \int\limits_{G_{\varepsilon,\kappa}}\frac{ar}{p}\,d\mathbf{t} = o(\varepsilon^4), \quad \int\limits_{G_{\varepsilon,\kappa}}\frac{r^2}{p}\,d\mathbf{t} = o(\varepsilon^4), \quad \varepsilon\to 0. \tag{115}$$

It is not difficult to show that

$$\int\limits_{\mathcal{R}^l}\frac{a^2}{p}\,d\mathbf{t} = \frac{1}{2\sigma^4}\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}_\varepsilon)\right\}\varepsilon^4+o(\varepsilon^4), \quad \varepsilon\to 0. \tag{116}$$

Thus, relations (112)–(116) yield

$$\begin{aligned} I(\varepsilon\mathbf{z}_\varepsilon;\varepsilon\mathbf{z}_\varepsilon+\mathbf{w}) &\ge \frac{\log e}{2\sigma^2}\mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}_\varepsilon)\right\}\varepsilon^2 \\ &- \frac{\log e}{4\sigma^4}\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}_\varepsilon)\right\}\varepsilon^4+o(\varepsilon^4), \quad \varepsilon\to 0. \end{aligned} \tag{117}$$

On the other hand, we have

$$\begin{aligned} I(\varepsilon\mathbf{z}_\varepsilon;\varepsilon\mathbf{z}_\varepsilon+\mathbf{w}) &\le I(\varepsilon\widehat{\mathbf{z}}_\varepsilon;\varepsilon\widehat{\mathbf{z}}_\varepsilon+\mathbf{w}) = \frac{\log e}{2\sigma^2}\mathrm{trace}\left\{\mathrm{cov}(\mathbf{z}_\varepsilon)\right\}\varepsilon^2 \\ &- \frac{\log e}{4\sigma^4}\mathrm{trace}\left\{\mathrm{cov}^2(\mathbf{z}_\varepsilon)\right\}\varepsilon^4+o(\varepsilon^4), \quad \varepsilon\to 0, \end{aligned} \tag{118}$$

where $\widehat{\mathbf{z}}_\varepsilon$ is a Gaussian random vector with the same covariance matrix as $\mathbf{z}_\varepsilon$. Theorem 5 follows from (117) and (118). □

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 (47), we can take $m=1$ without loss of generality. In this case we have

$$D\left(P_{Y|X=\mathsf{x}_0} \parallel P_{Y|X=0}\right) = -2+\frac{2}{\sqrt{N_0}}(|\Re\mathsf{x}_0|+|\Im\mathsf{x}_0|)+\exp\left[-\frac{2|\Re\mathsf{x}_0|}{\sqrt{N_0}}\right]+\exp\left[-\frac{2|\Im\mathsf{x}_0|}{\sqrt{N_0}}\right]. \tag{119}$$

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

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

Denote for simplicity

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

and let

$$D\left(P_{Y|X=\mathsf{x}_0} \parallel P_{Y|X=0}\right) = D_1(x)+D_2(y), \tag{120}$$

where

$$D_1(x) \stackrel{\text{def}}{=} 1+\frac{2}{\sqrt{N_0}}|x|+\exp\left[-\frac{2|x|}{\sqrt{N_0}}\right] \tag{121}$$

and

$$D_2(y) \stackrel{\text{def}}{=} 1+\frac{2}{\sqrt{N_0}}|y|+\exp\left[-\frac{2|y|}{\sqrt{N_0}}\right]. \tag{122}$$

It is clear that

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

and therefore $a = 2/N_0$.

Now, let us show that

$$D\left(P_{Y|X=\mathsf{x}_0} \parallel P_{Y|X=0}\right) \leq \frac{2}{N_0}||\mathsf{x}_0||^2 - \frac{1}{3\sqrt{2}(N_0)^{3/2}}||\mathsf{x}_0||^3 \quad \text{for} \quad ||\mathsf{x}_0|| \leq \frac{\sqrt{N_0}}{2}. \tag{123}$$

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

$$D_1(x) \leq \frac{2}{N_0}|x|^2 - \frac{4}{3(N_0)^{3/2}}|x|^3 + \frac{2}{3(N_0)^2}|x|^4 \leq \frac{2}{N_0}|x|^2 - \frac{2}{3(N_0)^{3/2}}|x|^3. \tag{124}$$

Here, we have used the well-known property of the Leibnitz power series. Similarly,

$$D_2(y) \leq \frac{2}{N_0}|y|^2 - \frac{2}{3(N_0)^{3/2}}|y|^3 \quad \text{for} \quad |y| \leq \sqrt{N_0}/2. \tag{125}$$

Thus,

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

for $||\mathsf{x}_0|| \leq \sqrt{N_0}/2$, i.e., (123) is proved.

It remains to show that

$$D\left(P_{Y|X=\mathsf{x}_0} \parallel P_{Y|X=0}\right) \leq \frac{1.9}{N_0}||\mathsf{x}_0||^2 \quad \text{for} \quad ||\mathsf{x}_0|| > \frac{\sqrt{N_0}}{2}. \tag{126}$$

To this end, note that

$$-1 + 2t + e^{-2t} < 1.7t^2 \quad \text{for} \quad t > \frac{1}{2\sqrt{2}} \tag{127}$$

since, as can easily be seen,

$$f\left(\frac{1}{2\sqrt{2}}\right) > 0, \quad f'\left(\frac{1}{2\sqrt{2}}\right) > 0, \quad \text{and} \quad f''(t) > 0 \quad \text{for} \quad t > \frac{1}{2\sqrt{2}},$$

where $f(t) = 1.7t^2 + 1 - 2t - e^{-2t}$. It follows from (127) that

$$D\left(P_{Y|X=\mathsf{x}_0} \parallel P_{Y|X=0}\right) \leq \frac{1.7}{N_0}||\mathsf{x}_0||^2 \quad \text{if} \quad |x| > \frac{\sqrt{N_0}}{2\sqrt{2}} \quad \text{and} \quad |y| > \frac{\sqrt{N_0}}{2\sqrt{2}}. \tag{128}$$

Assume for definiteness that $|x|^2 = \lambda||\mathsf{x}_0||^2$, $0 \leq \lambda \leq 1/2$. Then, taking into account (128), (124), and (125), we obtain

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

for all $||\mathsf{x}_0|| > \sqrt{N_0}/2$, i.e., inequality (126) is proved. This completes the proof of the lemma. $\square$

References

[1] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Information Theory*, to appear, 2002.

[2] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Information Theory*, vol. 36 (5), pp. 1019–1030, Sep. 1990.

[3] J. S. Richters, "Communication over fading dispersive channels," Tech. Rep., MIT Research Laboratory of Electronics, Nov. 1967.

[4] I. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. on Information 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. Peredach. 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. Transmission*, vol. 8, no. 4, pp. 28–39, Oct. 1972.

[7] V. V. Prelov, "Communication channel capacity with almost Gaussian noise," *Theory of Probability and its Applications*, vol. 33, no. 3, pp. 405–422, July 1988.

[8] V. V. Prelov, "Asymptotic expansions for the mutual information and for the capacity of continuous memoryless channels with weak input signal," *Probl. Control and Inform. Theory*, vol. 18 (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. on Information Theory*, pp. 1728–1735, Sep. 1993.

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

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

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