entropy

MDPI

Review

A View of Information-Estimation Relations in Gaussian Networks

Alex Dytso [1,*], Ronit Bustin [2,*], H. Vincent Poor [1,*] and Shlomo Shamai (Shitz) [2,*]

[1] Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA
[2] Department of Electrical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel
* Correspondence: adytso@princeton.edu (A.D.); bustin@technion.ac.il (R.B.); poor@princeton.edu (H.V.P.); sshlomo@ee.technion.ac.il (S.S.S.); Tel.: +1-609-258-1816 (H.V.P.)

Received: 31 May 2017; Accepted: 1 August 2017; Published: 9 August 2017

Abstract: Relations between estimation and information measures have received considerable attention from the information theory community. One of the most notable such relationships is the I-MMSE identity of Guo, Shamai and Verdú that connects the mutual information and the minimum mean square error (MMSE). This paper reviews several applications of the I-MMSE relationship to information theoretic problems arising in connection with multi-user channel coding. The goal of this paper is to review the different techniques used on such problems, as well as to emphasize the added-value obtained from the information-estimation point of view.

Keywords: network information theory; estimation theory; I-MMSE

1. Introduction

The connections between information theory and estimation theory go back to the late 1950s in the work of Stam in which he uses the de Bruijn identity [1], attributed to his PhD advisor, which connects the differential entropy and the Fisher information of a random variable contaminated by additive white Gaussian noise. In 1968 Esposito [2] and then in 1971 Hatsell and Nolte [3] identified connections between the Laplacian and the gradient of the log-likelihood ratio and the conditional mean estimate. Information theoretic measure can indeed be put in terms of log-likelihood ratios, however, these works did not make this additional connecting step. In the early 1970s continuous-time signals observed in white Gaussian noise received specific attention in the work of Duncan [4] and Kadota et al. [5] who investigated connections between the mutual information and causal filtering. In particular, Duncan and Zakai (Duncan's theorem was independently obtained by Zakai in the general setting of inputs that may depend causally on the noisy output in a 1969 unpublished Bell Labs Memorandum (see [6])) [4,7] showed that the input-output mutual information can be expressed as a time integral of the causal minimum mean square error (MMSE). It was only in 2005 that Guo, Shamai and Verdú revealed the I-MMSE relationship [8], which similarly to the de Bruijn identity, relates information theoretic quantities to estimation theoretic quantities over the additive white Gaussian noise channel. Moreover, the fact that the I-MMSE relationship connects the mutual information with the MMSE has made it considerably more applicable, specifically to information theoretic problems.

The I-MMSE type relationships have received considerable attention from the information theory community and a number of extensions have been found. In [9], in the context of multiple-input multiple-output (MIMO) Gaussian channels, it was shown that the gradient of the mutual information with respect to the channel matrix is equal to the channel matrix times the MMSE matrix. In [10] a version of the I-MMSE identity has been shown for Gaussian channels with feedback. An I-MMSE type relationship has been found for additive non-Gaussian channels in [11] and non-additive channels with a well-defined notion of the signal-to noise ratio (SNR) in [12–15]. A relationship between the MMSE and the relative entropy has been established in [16], and between the score function and

Rényi divergence and f-divergence in [17]. The I-MMSE relationship has been extended to continuous time channels in [8] and generalized in [18] by using Malliavin calculus. For other continuous time generalizations the reader is referred to [19–21]. Finally, Venkat and Weissman [22] dispensed with the expectation and provided a point-wise identity that has given additional insight into this relationship. For a comprehensive summary of results on the interplay between estimation and information measures the interested reader is referred to [23].

In this survey we provide an overview of several applications of the I-MMSE relationship to multi-user information theoretic problems. We consider three types of applications:

1. Capacity questions, including both converse proofs and bounds given additional constraints such as discrete inputs;
2. The MMSE SNR-evolution, meaning the behavior of the MMSE as a function of SNR for asymptotically optimal code sequences (code sequences that approach capacity as $n \to \infty$); and
3. Finite blocklength effects on the SNR-evolution of the MMSE and hence effects on the rate as well.

Our goal in this survey is both to show the strength of the I-MMSE relationship as a tool to tackle network information theory problems, and to overview the set of tools used in conjunction with the I-MMSE relationship such as the "single crossing point" property. As will be seen such tools lead to alternative and, in many cases, simpler proofs of information theoretic converses.

We are also interested in using estimation measures in order to upper or lower bound information measures. Such bounds lead to simple yet powerful techniques that are used to find "good" capacity approximations. At the heart of this technique is a generalization of the Ozarow-Wyner bound [24] based on minimum mean p-th error (MMPE). We hope that this overview will enable future application of these properties in additional multi-user information theoretic problems.

The outline of the paper is as follows:

1. In Section 2 we review information and estimation theoretic tools that are necessary for the presentation of the main results.
2. In Section 3 we go over point-to-point information theory and give the following results:
   - In Section 3.1, using the I-MMSE and a basic MMSE bound, a simple converse is shown for the Gaussian point-to-point channel;
   - In Section 3.2, a lower bound, termed the Ozarow-Wyner bound, on the mutual information achieved by a discrete input on an AWGN channel, is presented. The bound holds for vector discrete inputs and yields the sharpest known version of this bound; and
   - In Section 3.3, it is shown that the MMSE can be used to identify optimal point-to-point codes. In particular, it is shown that an optimal point-to-point code has a unique SNR-evolution of the MMSE.
3. In Section 4 we focus on the wiretap channel and give the following results:
   - In Section 4.1, using estimation theoretic properties a simple converse is shown for the Gaussian wiretap channel that avoids the use of the entropy power inequality (EPI); and
   - In Section 4.2, some results on the SNR-evolution of the code sequences for the Gaussian wiretap channel are provided, showing that for the secrecy capacity achieving sequences of codes the SNR-evolution is unique.
4. In Section 5 we study a communication problem in which the transmitter wishes to maximize its communication rate, while subjected to a constraint on the disturbance it inflicts on the secondary receiver. We refer to such scenarios as communication with a disturbance constraint and give the following results:
   - In Section 5.1 it is argued that an instance of a disturbance constraint problem, when the disturbance is measured by the MMSE, has an important connection to the capacity of a two-user Gaussian interference channel;

- In Section 5.2 the capacity is characterized for the disturbance problem when the disturbance is measured by the MMSE;
- In Section 5.3 the capacity is characterized for the disturbance problem when the disturbance is measured by the mutual information. The MMSE and the mutual information disturbance results are compared. It is argued that the MMSE disturbance constraint is a more natural measure in the case when the disturbance measure is chosen to model the unintended interference;
- In Section 5.4 new bounds on the MMSE are derived and are used to show upper bounds on the disturbance constraint problem with the MMSE constraint when the block length is finite; and
- In Section 5.5 a notion of mixed inputs is defined and is used to show lower bounds on the rates of the disturbance constraint problem when the block length is finite.

5. In Section 6 we focus on the broadcast channel and give the following results:
   - In Section 6.1, the converse for a scalar Gaussian broadcast channel, which is based only on the estimation theoretic bounds and avoids the use of the EPI, is derived; and
   - In Section 6.2, similarly to the Gaussian wiretap channel, we examine the SNR-evolution of asymptotically optimal code sequences for the Gaussian broadcast channel, and show that any such sequence has a unique SNR-evolution of the MMSE.
6. In Section 7 the SNR-evolution of the MMSE is derived for the K-user broadcast channel.
7. In Section 8, building on the MMSE disturbance problem in Section 5.1, it is shown that for the two-user Gaussian interference channel a simple transmission strategy of treating interference as noise is approximately optimal.

Section 9 concludes the survey by pointing out interesting future directions.

1.1. Notation

Throughout the paper we adopt the following notational conventions:

- Random variables and vectors are denoted by upper case and bold upper case letters, respectively, where r.v. is short for either random variable or random vector, which should be clear from the context. The dimension of these random vectors is n throughout the survey. Matrices are denoted by bold upper case letters;
- If A is an r.v. we denote the support of its distribution by $\mathsf{supp}(A)$;
- The symbol $|\cdot|$ may denote different things: $|\mathbf{A}|$ is the determinant of the matrix $\mathbf{A}$, $|\mathcal{A}|$ is the cardinality of the set $\mathcal{A}$, $|X|$ is the cardinality of $\mathsf{supp}(X)$, or $|x|$ is the absolute value of the real-valued x;
- The symbol $\|\cdot\|$ denotes the Euclidian norm;
- $\mathbb{E}[\cdot]$ denotes the expectation;
- $\mathcal{N}(\mathbf{m_X}, \mathbf{K_X})$ denotes the density of a real-valued Gaussian r.v. $\mathbf{X}$ with mean vector $\mathbf{m_X}$ and covariance matrix $\mathbf{K_X}$;
- $X \sim \mathsf{PAM}\left(N, d_{\min(X)}\right)$ denotes the uniform probability mass function over a zero-mean pulse amplitude modulation (PAM) constellation with $|\mathsf{supp}(X)| = N$ points, minimum distance $d_{\min(X)}$, and therefore average energy $\mathbb{E}[X^2] = d^2_{\min(X)} \frac{N^2-1}{12}$;
- The identity matrix is denoted by $\mathbf{I}$;
- The reflection of the matrix $\mathbf{A}$ along its main diagonal, or the transpose operation, is denoted by $\mathbf{A}^\mathsf{T}$;
- The trace operation on the matrix $\mathbf{A}$ is denoted by $\mathrm{Tr}(\mathbf{A})$;
- The order notation $\mathbf{A} \succeq \mathbf{B}$ implies that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix;
- $\log(\cdot)$ denotes the logarithm to the base e;
- $[n_1 : n_2]$ is the set of integers from n_1 to $n_2 \geq n_1$;
- For $x \in \mathbb{R}$ we let $\lfloor x \rfloor$ denote the largest integer not greater than x;
- For $x \in \mathbb{R}$ we let $[x]^+ := \max(x, 0)$ and $\log^+(x) := [\log(x)]^+$;

- Let $f(x), g(x)$ be two real-valued functions. We use the Landau notation $f(x) = O(g(x))$ to mean that for *some* $c > 0$ there exists an x_0 such that $f(x) \leq c\, g(x)$ for all $x \geq x_0$, and $f(x) = o(g(x))$ to mean that for *every* $c > 0$ there exists an x_0 such that $f(x) < cg(x)$ for all $x \geq x_0$; and
- We denote the upper incomplete gamma function and the gamma function by

$$\Gamma(x;a) := \int_a^\infty t^{x-1} e^{-t} dt,\ x \in \mathbb{R}, a \in \mathbb{R}^+, \tag{1a}$$

$$\Gamma(x) := \Gamma(x;0). \tag{1b}$$

2. Estimation and Information Theoretic Tools

In this section, we overview relevant information and estimation theoretic tools. The specific focus is to show how estimation theoretic measures can be used to represent or bound information theoretic measures such as entropy and mutual information.

2.1. Estimation Theoretic Measures

Of central interest to us is the following estimation measure constructed from the L_p norm.

Definition 1. *For the random vector* $\mathbf{V} \in \mathbb{R}^n$ *and* $p > 0$ *let*

$$\|\mathbf{V}\|_p := \left(\frac{1}{n}\mathbb{E}\left[\|\mathbf{V}\|^p\right]\right)^{\frac{1}{p}} = \left(\frac{1}{n}\mathbb{E}\left[\left(\mathrm{Tr}(\mathbf{V}\mathbf{V}^\mathsf{T})\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}. \tag{2a}$$

We define the minimum mean p-th error (MMPE) of estimating $\mathbf{X}$ *from* $\mathbf{Y}$ *as*

$$\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p) = \inf_f \|\mathbf{X} - f(\mathbf{Y})\|_p^p, \tag{2b}$$

where the minimization is over all possible Borel measurable functions $f(\mathbf{Y})$*. Whenever the optimal MMPE estimator exists, we shall denote it by* $f_p(\mathbf{X}|\mathbf{Y})$*.*

In particular, for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ the norm in (2a) is given by

$$n\|\mathbf{Z}\|_p^p = \mathbb{E}\left[\left(\sum_{i=1}^n Z_i^2\right)^{\frac{p}{2}}\right] = 2^{\frac{p}{2}} \frac{\Gamma\left(\frac{n}{2} + \frac{p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},\ \text{for } n \in \mathbb{N}, p \geq 0, \tag{3}$$

and for $\mathbf{V}$ uniform over the n dimensional ball of radius r the norm in (2a) is given by

$$n\|\mathbf{V}\|_p^p = \frac{1}{\mathrm{Vol}(B(r))} \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^r \rho^p \rho^{n-1} d\rho = \frac{n}{2p+2n} r^p,\ \text{for } n \in \mathbb{N}, p \geq 0. \tag{4}$$

We shall denote

$$\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p) = \mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p), \tag{5}$$

if $\mathbf{Y}$ and $\mathbf{X}$ are related as

$$\mathbf{Y} = \sqrt{\mathsf{snr}}\,\mathbf{X} + \mathbf{Z}, \tag{6}$$

where $\mathbf{Z}, \mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{X}$, and $\mathsf{snr} \geq 0$ is the SNR. When it will be necessary to emphasize the SNR at the output $\mathbf{Y}$, we will denote it by $\mathbf{Y}_{\mathsf{snr}}$. Since the distribution of the noise is fixed $\mathrm{mmpe}(\mathbf{X}|\mathbf{Y};p)$ is completely determined by the distribution of $\mathbf{X}$ and snr and there is no ambiguity in using the notation $\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p)$. Applications to the Gaussian noise channel will be the main focus of this paper.

In the special, case when $p = 2$, we refer to the MMPE as the minimum mean square error (MMSE) and use the notation

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, 2) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{7}$$

in which case we also have that $f_2(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$.

Remark 1. *The notation $f_p(\mathbf{X}|\mathbf{Y})$, for the optimal estimator in (2) is inspired by the conditional expectation $\mathbb{E}[\mathbf{X}|\mathbf{Y}]$, and $f_p(\mathbf{X}|\mathbf{Y})$ should be thought of as an operator on $\mathbf{X}$ and a function of $\mathbf{Y}$. Indeed, for $p = 2$, the MMPE reduces to the MMSE; that is, $\mathrm{mmpe}(\mathbf{X}|\mathbf{Y}; 2) = \mathrm{mmse}(\mathbf{X}|\mathbf{Y})$ and $f_2(\mathbf{X}|\mathbf{Y}) = \mathbb{E}[\mathbf{X}|\mathbf{Y}]$.*

Finally, similarly to the conditional expectation, the notation $f_p(\mathbf{X}|\mathbf{Y} = \mathbf{y})$ should be understood as an evaluation for a realization of a random variable $\mathbf{Y}$, while $f_p(\mathbf{X}|\mathbf{Y})$ should be understood as a function of a random variable $\mathbf{Y}$ which itself is a random variable.

Lemma 1. *(Existence of the Optimal Estimator [25]) For any $\mathbf{X}$ and $\mathbf{Y}$ given by (6) an optimal estimator exists and the infimum in (2) can be attained.*

In certain cases the optimal estimator might not be unique and the interested reader is referred to [25] for such examples. In general we do not have a closed form solution for the MMPE optimal estimator in (2). Interestingly, the optimal estimator for Gaussian inputs can be found and is linear for all $p \geq 1$

Proposition 1. *(MMPE of a Gaussian Input [25–27]) For $\mathbf{X}_G \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $p \geq 1$*

$$\mathrm{mmpe}(\mathbf{X}_G, \mathsf{snr}, p) = \frac{\|\mathbf{Z}\|_p^p}{(1 + \mathsf{snr})^{\frac{p}{2}}}, \tag{8a}$$

with the optimal estimator given by

$$f_p(\mathbf{X}_G|\mathbf{Y} = \mathbf{y}) = \frac{\sqrt{\mathsf{snr}}\,\mathbf{y}}{1 + \mathsf{snr}}. \tag{8b}$$

Note that unlike the Gaussian case in general the estimator will be a function of the order p. For $X = \pm 1$ equally likely (i.e., binary phase shift keying—BPSK) the optimal estimator is given by

$$f_p(X|Y = y) = \tanh\left(\frac{y\sqrt{\mathsf{snr}}}{p - 1}\right). \tag{9}$$

Often the MMPE is difficult to compute, even for $p = 2$ (MMSE), and one instead is interested in deriving upper bounds on the MMPE. One of the most useful upper bounds on the MMPE can be obtained by restricting the optimization in (2) to linear functions.

Proposition 2. *(Asymptotically Gaussian is the "hardest" to estimate [25]) For $\mathsf{snr} \geq 0$, $p \geq 1$, and a random variable $\mathbf{X}$ such that $\|\mathbf{X}\|_p^p \leq \sigma^p \|\mathbf{Z}\|_p^p$, we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \leq \kappa_{p,\sigma^2\mathsf{snr}} \cdot \frac{\sigma^p \|\mathbf{Z}\|_p^p}{(1 + \mathsf{snr}\sigma^2)^{\frac{p}{2}}}, \tag{10a}$$

where

$$\text{for } p = 2 : \kappa_{p,\sigma^2\mathsf{snr}}^{\frac{1}{p}} = 1, \tag{10b}$$

$$\text{for } p \neq 2 : 1 \leq \kappa_{p,\sigma^2\mathsf{snr}}^{\frac{1}{p}} = \frac{1+\sqrt{\sigma^2\mathsf{snr}}}{\sqrt{1+\sigma^2\mathsf{snr}}} \leq 1 + \frac{1}{\sqrt{1+\sigma^2\mathsf{snr}}}. \tag{10c}$$

Moreover, a Gaussian $\mathbf{X}$ *with per-dimension variance* σ^2 *(i.e.,* $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$*) asymptotically achieves the bound in* (10a)*, since* $\lim_{\mathsf{snr}\to\infty} \kappa_{p,\sigma^2\mathsf{snr}} = 1$.

For the case of $p = 2$, the bound in (10a) is achieved with a Gaussian input for all SNR's. Moreover, this special case of the bound in (10a), namely

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \frac{\sigma^2}{1+\sigma^2\mathsf{snr}}, \tag{11}$$

for all $\|\mathbf{X}\|_2^2 \leq \sigma^2$, is referred to as the linear minimum mean square error (LMMSE) upper bound.

2.2. Mutual Information and the I-MMSE

For two random variables $(\mathbf{X}, \mathbf{Y})$ distributed according to $P_{\mathbf{XY}}$ the *mutual information* is defined as

$$I(\mathbf{X};\mathbf{Y}) = \mathbb{E}\left[\log \frac{dP_{\mathbf{XY}}}{d(P_{\mathbf{X}} \times P_{\mathbf{Y}})}\right], \tag{12}$$

where $\frac{dP_{\mathbf{XY}}}{d(P_{\mathbf{X}} \times P_{\mathbf{Y}})}$ is the Radon-Nikodym derivative. For the channel in (6) the mutual information between $\mathbf{X}$ and $\mathbf{Y}$ takes the following form:

$$I(\mathbf{X};\mathbf{Y}) = \mathbb{E}\left[\log\left(\frac{f_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{f_{\mathbf{Y}}(\mathbf{Y})}\right)\right], \tag{13}$$

and it will be convenient to use the normalized mutual information

$$I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{n} I(\mathbf{X};\mathbf{Y}). \tag{14}$$

The basis for much of our analysis is the fundamental relationship between information theory and estimation theory, also known as the Guo, Shamai and Verdú I-MMSE relationship [8].

Theorem 1. *(I-MMSE [8]) For any* $\mathbf{X}$ *(independent of* snr*) we have that*

$$\frac{d}{d\mathsf{snr}} I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{15a}$$

$$I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt. \tag{15b}$$

In [28] the I-MMSE relationship has been partially extended to the limit as $n \to \infty$. This result was then extended in [29] under the assumption that the mutual information sequence converges.

Proposition 3. *(I-MMSE limiting expression [29]) Suppose that* $\|\mathbf{X}\|_2^2 \leq \sigma^2 < \infty$ *and*

$$\lim_{n\to\infty} I_n(\mathbf{X}, \mathsf{snr}) = I_\infty(\mathbf{X}, \mathsf{snr}), \tag{16}$$

exists. (The limit here is taken with respect to a sequence of input distributions over $\{\mathbf{X}_n\}_{n\geq 1}$ *which induces a sequence of input-output joint distributions. The second moment constraint* $\|\mathbf{X}_n\|_2^2$ *should be understood in a similar manner, as a constraint for every n in the sequence.) Then,*

$$\lim_{n\to\infty} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}) = \mathrm{mmse}_\infty(\mathbf{X}, \mathsf{snr}), \tag{17}$$

and the I-MMSE relationship holds for the following limiting expression:

$$I_\infty(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}_\infty(\mathbf{X}, t)dt. \tag{18}$$

Proof. The proof is given in Appendix A. □

Properties of the MMSE, with the specific focus on the I-MMSE identity, as a function of the input distribution and the noise distribution have been thoroughly studied and the interested reader is referred to [17,30–32]. For the derivation of the I-MMSE and a comprehensive summary of various extension we refer the reader to [23].

For a continuous random vector $\mathbf{X}$ with the density $f_{\mathbf{X}}$ the differential entropy is defined as

$$h(\mathbf{X}) = -\mathbb{E}\left[\log f_{\mathbf{X}}(\mathbf{X})\right]. \tag{19}$$

Moreover, for a discrete random vector $\mathbf{X}$ the discrete entropy is defined as

$$H(\mathbf{X}) = -\sum_{\mathbf{x}\in\mathsf{supp}(\mathbf{X})} p_{\mathbf{x}} \log p_{\mathbf{x}}, \text{ where } p_{\mathbf{x}} = \mathbb{P}[\mathbf{X} = \mathbf{x}]. \tag{20}$$

2.3. Single Crossing Point Property

Upper bounds on the MMSE are useful, thanks to the I-MMSE relationship, as tools to derive information theoretic converse results, and have been used in [23,30,33,34] to name a few. The key MMSE upper bound that will be used in conjunction with the I-MMSE to derive information theoretic converses is the single crossing point property (SCPP).

Proposition 4. *(SCPP [30,33]) Let* $\|\mathbf{X}\|_2 \leq 1$*. Then for any fixed* snr_0 *there exists a unique* $\alpha \in [0, 1]$ *such that*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\alpha}{1 + \alpha\mathsf{snr}_0}. \tag{21a}$$

Moreover, for every $\mathsf{snr} > \mathsf{snr}_0$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \frac{\alpha}{1 + \alpha\mathsf{snr}}, \tag{21b}$$

and for every $\mathsf{snr} \leq \mathsf{snr}_0$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \geq \frac{\alpha}{1 + \alpha\mathsf{snr}}. \tag{21c}$$

Even though the statement of Proposition 4 seems quite simple it turns out that it is sufficient to show a special case of the EPI [33]:

$$\mathrm{e}^{\frac{2}{n}h(\mathbf{X}+\mathbf{Z})} \geq \mathrm{e}^{\frac{2}{n}h(\mathbf{X})} + \mathrm{e}^{\frac{2}{n}h(\mathbf{Z})}, \tag{22}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\mathbf{Z})$. Interestingly, the I-MMSE appears to be a very powerful tool in deriving EPI type inequalities; the interested reader is referred to [35–38].

In [25] it has been pointed out that the SCPP upper bound can also be shown for the MMPE as follows.

Proposition 5. *(Generalized SCPP upper bound [25]) Let* $\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}_0, p) = \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta \mathsf{snr}_0}$ *for some* $\beta \geq 0$*. Then,*

$$\mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X}, \mathsf{snr}, p) \leq c_p \cdot \frac{\beta \|\mathbf{Z}\|_p^2}{1+\beta\, \mathsf{snr}}, \text{ for } \mathsf{snr} \geq \mathsf{snr}_0, \tag{23a}$$

where

$$c_p = \begin{cases} 2 & p \neq 2 \\ 1 & p = 2 \end{cases}. \tag{23b}$$

Proof. The proof of Propositions 4 and 5 uses a clever choice of a sub-optimal estimator. The interested reader is referred to Appendix B for the proof. □

2.4. Complementary SCPP Bounds

Note that the SCPP allows us to upper bound the MMSE for all values of $\mathsf{snr} \geq \mathsf{snr}_0$, and as will be shown later this is a very powerful tool in showing information theoretic converses. Another interesting question is whether we can produce a complementary upper bound to that of the SCPP. That is, can we show an upper bounds on the MMSE for $\mathsf{snr} \leq \mathsf{snr}_0$? As will be demonstrated in Section 5, such complementary SCPP bounds are useful in deriving information theoretic converses for problems of communication with a disturbance constraint.

The next result shows that this is indeed possible.

Proposition 6. *(Complementary SCPP bound [25]) For* $0 < \mathsf{snr} \leq \mathsf{snr}_0$, $\mathbf{X}$ *and* $p \geq 0$*, we have*

$$\mathrm{mmpe}(\mathbf{X}, \mathsf{snr}, p) \leq \kappa_{n,t}\, \mathrm{mmpe}^{\frac{1-t}{1+t}}\left(\mathbf{X}, \mathsf{snr}_0, \frac{1+t}{1-t} \cdot p\right),$$
$$\text{where } \kappa_{n,t} := \left(\frac{2^n}{n^2}\right)^{\frac{t}{t+1}} \left(\frac{1}{1-t}\right)^{\frac{nt}{t+1}-\frac{1}{2}},\ t = \frac{\mathsf{snr}_0 - \mathsf{snr}}{\mathsf{snr}_0}.$$

An interesting property of the bound in Proposition 6 is that the right hand side of the inequality keeps the channel SNR fixed and only varies the order of the MMPE while the left hand side of the inequality keeps the order fixed and changes the SNR value.

2.5. Bounds on Differential Entropy

Another common application of estimation theoretic measures is to bound information measures. Next, we presented one such bound.

For any random vector $\mathbf{V}$ such that $|\mathrm{Cov}(\mathbf{V})| < \infty$, $h(\mathbf{V}) < \infty$, and any random vector $\mathbf{Y}$, the following inequality is considered to be a continuous analog of Fano's inequality [39]:

$$h(\mathbf{V}|\mathbf{Y}) \leq \frac{n}{2} \log(2\pi \mathrm{e}\, |\mathrm{Cov}(\mathbf{V}|\mathbf{Y})|^{\frac{1}{n}}) \tag{24}$$
$$\leq \frac{n}{2} \log(2\pi \mathrm{e}\, \mathrm{mmse}(\mathbf{V}|\mathbf{Y})), \tag{25}$$

where the inequality in (25) is a consequence of the arithmetic-mean geometric-mean inequality, that is, for any $0 \preceq \mathbf{A}$ we have used $|\mathbf{A}|^{\frac{1}{n}} = (\prod_{i=1}^n \lambda_i)^{\frac{1}{n}} \leq \frac{\sum_{i=1}^n \lambda_i}{n} = \frac{\mathrm{Tr}(\mathbf{A})}{n}$ where λ_i's are the eigenvalues of $\mathbf{A}$.

The inequality in (25) can be generalized in the following way.

Theorem 2. *(Conditional Entropy Bound [25]) Let* $\mathbf{V} \in \mathbb{R}^n$ *be such that* $h(\mathbf{V}) < \infty$ *and* $\|\mathbf{V}\|_p < \infty$*. Then, for any* $p \in (0, \infty)$ *and for any* $\mathbf{Y} \in \mathbb{R}^n$*, we have*

$$h(\mathbf{V}|\mathbf{Y}) \leq \frac{n}{2} \log \left(k_{n,p}^2 \cdot n^{\frac{2}{p}} \cdot \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{V}|\mathbf{Y}; p) \right), \tag{26}$$

where $k_{n,p} = \frac{\sqrt{\pi} \left(\frac{pe}{n}\right)^{\frac{1}{p}} \Gamma^{\frac{1}{n}} \left(\frac{n}{p}+1\right)}{\Gamma^{\frac{1}{n}} \left(\frac{n}{2}+1\right)}$.

While the MMPE is still a relatively new tool it has already found several applications:

- The MMPE can be used to bound the conditional entropy (see Theorem 2 in Section 2.5). These bounds are generally tighter than the MMSE based bound especially for highly non-Gaussian statistics;
- The MMPE can be used to develop bounds on the mutual information of discrete inputs via the generalized Ozarow-Wyner bound (see Theorem 4 in Section 3.2); The MMPE and the Ozarow-Wyner bound can be used to give tighter bounds on the gap to capacity achieved by PAM input constellations (see Figure 2);
- The MMPE can be used as a key tool in finding complementary bounds on the SCPP (see Theorem 10 in Section 5.4). Note that using the MMPE as a tool produces the correct phase transition behavior; and
- While not mentioned, another application is to use the MMPE to bound the derivatives of the MMSE; see [25] for further details.

3. Point-to-Point Channels

In this section, we review Shannon's basic theorem for point-to-point communication and introduce relevant definitions used throughout the paper. The point-to-point channel is also a good starting point for introducing many of the techniques that will be used in this survey.

A classical point-to-point channel is shown in Figure 1. The transmitter wishes to reliably communicate a message W at a rate R bits per transmission to a receiver over a noisy channel. To that end, the transmitter encodes the message W into a signal $\mathbf{X}$ and transmits it over a channel in n time instances. Upon receiving a sequence $\mathbf{Y}$, a corrupted version of $\mathbf{X}$, the receiver decodes it to obtain the estimate $\hat{W}$.

Definition 2. *A* memoryless channel *(MC), assuming no feedback,* $(\mathcal{X}, P_{Y|X}, \mathcal{Y})$ *(in short* $P_{Y|X}$*) consists of an input set* $\mathcal{X}$*, an output set* $\mathcal{Y}$*, and a collection of transition probabilities* $P_{Y|X}$ *on* $\mathcal{Y}$ *for every* $x \in \mathcal{X}$*. The transition of a length-n vector* $\mathbf{X}$ *through such a channel then has the following conditional distribution:*

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} P_{Y|X}(y_i|x_i). \tag{27}$$

Definition 3. *A* code *of length n and rate R, denoted by* $(2^{nR}, n)$*, of the channel* $P_{Y|X}$ *consist of the following:*

- *A* message set $\{1, 2, ..., 2^{nR}\}$*. We assume that the message W is chosen uniformly over the message set.*
- *An* encoding function $\mathbf{X} : \{1, 2, ..., 2^{nR}\} \to \mathcal{X}^n$ *that maps messages W to codewords* $\mathbf{X}(W)$*. The set of all codewords is called the* codebook *and is denoted by* $\mathcal{C}$*; and*
- *A* decoding function $g : \mathcal{Y}^n \to \{1, 2, ..., 2^{nR}\}$ *that assigns an estimate* $\hat{W}$ *to each received sequence.*

The average probability of error for a $(2^{nR}, n)$ *code is defined as*

$$\mathbb{P}[\hat{W} \neq W] = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \mathbb{P}[\hat{W} \neq w \mid W = w].$$

Definition 4. *A rate R is said to be achievable over a point-to-point channel if there exists a sequence of codes, $(2^{nR}, n)$, such that*

$$\lim_{n\to\infty} \mathbb{P}[\hat{W} \neq W] = 0. \tag{28}$$

The capacity C of a point-to-point channel is the supremum over all achievable rates.

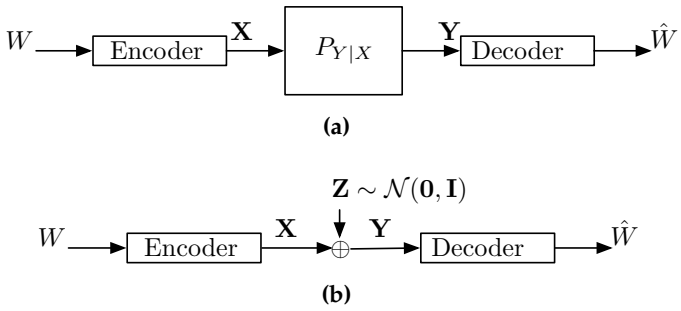


Figure 1. A point-to-point communication system. (**a**) A memoryless point-to-point channel with a transition probability $P_{Y|X}$; (**b**) A Gaussian point-to-point channel.

A crowning achievement of Shannon's 1948 paper [40] is a simple characterization of the capacity of a point-to-point channel.

Theorem 3. *(Channel Coding Theorem [40]) The capacity of the channel $P_{Y|X}$ is given by*

$$C = \max_{P_X} I(X;Y). \tag{29}$$

For a formal derivation of the capacity expression in (29) the reader is referred to classical texts such as [39,41,42].

3.1. A Gaussian Point-to-Point Channel

In this section we consider the practically relevant Gaussian point-to-point channel shown in Figure 1b and given by

$$Y = \sqrt{\mathsf{snr}}X + Z, \tag{30}$$

where Z is standard Gaussian noise and there is an additional input power constraint $\mathbb{E}[X^2] \leq 1$. The capacity in this setting was solved in the original paper by Shannon and is given by

$$C = \frac{1}{2}\log(1+\mathsf{snr}). \tag{31}$$

To show the converse proof of the capacity (the upper bound on the capacity) in (31), Shannon used the maximum entropy principle. In contrast to Shannon's proof, we show the converse can be derived by using the I-MMSE and the LMMSE upper bound in (11)

$$\begin{aligned} I(X;Y) &= \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(X,\gamma)d\gamma \\ &\leq \frac{1}{2}\int_0^{\mathsf{snr}} \frac{1}{1+\gamma}d\gamma \\ &= \frac{1}{2}\log(1+\mathsf{snr}). \end{aligned} \tag{32}$$

It is well know that the upper bound in (32) is achievable if and only if the input is $X \sim \mathcal{N}(0,1)$.

The main idea behind the upper bounding technique in (32) is to find an upper bound on the MMSE that holds for all SNR's and integrate it to get an upper bound on the mutual information. This simple, yet powerful, idea will be used many times throughout this paper to show information theoretic converses for multi-user channels.

3.2. Generalized Ozarow-Wyner Bound

In practice, Gaussian inputs are seldom used and it is important to assess the performance of more practical discrete constellations (or inputs). Another reason is that discrete inputs often outperform Gaussian inputs in competitive multi-user scenarios, such as the interference channel, as will be demonstrated in Section 7. For other examples of discrete inputs being useful in multi-user settings, the interested readers is referred to [43–46].

However, computing an exact expression for the mutual information between the channel input and output when the inputs are discrete is often impractical or impossible. Therefore, the goal is to derive a good computable lower bound on the mutual information that is not too far away from the true value of the mutual information. As we will see shortly, estimation measures such as the MMSE and the MMPE will play an important role in establishing good lower bounds on the mutual information.

The idea of finding good capacity approximations can be traced back to Shannon. Shannon showed, in his unpublished work in 1948 [47], the asymptotic optimality of a PAM input for the point-to-point power-constrained Gaussian noise channel. Another such observation about approximate optimality of a PAM input was made by Ungerboeck in [48] who, through numerical methods, observed that the rate of a properly chosen PAM input is always a constant away from the AWGN capacity.

Shannon's and Ungerboeck's arguments were solidified by Ozarow and Wyner in [24] where firm lower bounds on the achievable rate with a PAM input were derived and used to show optimality of PAM to within 0.41 bits [24].

In [24] the following "Ozarow-Wyner lower bound" on the mutual information achieved by a discrete input X_D transmitted over an AWGN channel was shown:

$$[H(X_D) - \mathsf{gap}]^+ \leq I(X_D;Y) \leq H(X_D), \tag{33a}$$

$$\mathsf{gap} \leq \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) + \frac{1}{2}\log\left(1 + \frac{\mathrm{lmmse}(X,\mathsf{snr})}{d_{\min}(X_D)^2}\right), \tag{33b}$$

where $\mathrm{lmmse}(X|Y)$ is the LMMSE. The advantage of the bound in (33) compared to the existing bounds is its computational simplicity. The bound depends only on the entropy, the LMMSE, and the minimum distance, which are usually easy to compute.

The bound in (33) has also been proven to be useful for other problems such as two-user Gaussian interference channels [45,49], communication with a disturbance constraint [50], energy harvesting problems [51,52], and information-theoretic security [53].

The bound on the gap in (33) has been sharpened in [45] to

$$\mathsf{gap} \leq \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) + \frac{1}{2}\log\left(1 + \frac{\mathrm{mmse}(X,\mathsf{snr})}{d_{\min}(X_D)^2}\right), \tag{34}$$

since $\mathrm{lmmse}(X,\mathsf{snr}) \geq \mathrm{mmse}(X,\mathsf{snr})$.

Finally, the following generalization of the bound in (34) to discrete vector input, which is the sharpest known bound on the gap term, was derived in [25].

Theorem 4. *(Generalized Ozarow-Wyner Bound [25]) Let $\mathbf{X}_D$ be a discrete random vector with finite entropy, and let $\mathcal{K}_p$ be a set of continuous random vectors, independent of $\mathbf{X}_D$, such that for every $\mathbf{V} \in \mathcal{K}_p$, $h(\mathbf{V})$, $\|\mathbf{V}\|_p < \infty$, and*

$$H(\mathbf{X}_D|\mathbf{X}_D + \mathbf{V}) = 0, \forall\, \mathbf{V} \in \mathcal{K}_p. \tag{35a}$$

Then for any $p > 0$

$$[H(\mathbf{X}_D) - \mathsf{gap}_p]^+ \leq I(\mathbf{X}_D; \mathbf{Y}) \leq H(\mathbf{X}_D), \tag{35b}$$

where

$$n^{-1}\mathsf{gap}_p \leq \inf_{\mathbf{V}\in\mathcal{K}_p} \left(G_{1,p}(\mathbf{V}, \mathbf{X}_D) + G_{2,p}(\mathbf{V})\right),$$

$$G_{1,p}(\mathbf{V}, \mathbf{X}_D) = \log\left(\frac{\|\mathbf{V} + \mathbf{X}_D - f_p(\mathbf{X}_D|\mathbf{Y})\|_p}{\|\mathbf{V}\|_p}\right) \leq \begin{cases} \log\left(1 + \frac{\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}_D, \mathsf{snr}, p)}{\|\mathbf{V}\|_p}\right), p \neq 2 \\ \frac{1}{2}\log\left(1 + \frac{\mathrm{mmse}(\mathbf{X}_D, \mathsf{snr})}{\|\mathbf{V}\|_2^2}\right), p = 2 \end{cases}, \tag{35c}$$

$$G_{2,p}(\mathbf{V}) = \log\left(\frac{k_{n,p} \cdot n^{\frac{1}{p}} \cdot \|\mathbf{V}\|_p}{\mathrm{e}^{\frac{1}{n}h_{\mathrm{e}}(\mathbf{V})}}\right). \tag{35d}$$

Remark 2. *The condition in (35a) can be enforced by, for example, selecting the support of $\mathbf{V}$ to satisfy a non-overlap condition given by*

$$\mathsf{supp}(\mathbf{V} + \mathbf{x}_i) \cap \mathsf{supp}(\mathbf{V} + \mathbf{x}_j) = \emptyset, \ \forall\, \mathbf{x}_i, \mathbf{x}_j \in \mathsf{supp}(\mathbf{X}_D), i \neq j, \tag{36}$$

as was done in [54].

It is interesting to note that the lower bound in (35b) resembles the bound for lattice codes in [55], where $\mathbf{V}$ can be thought of as a dither, $G_{2,p}$ corresponds to the log of the normalized p-moment of a compact region in $\mathbb{R}^n$, $G_{1,p}$ corresponds to the log of the normalized MMSE term, and $H(\mathbf{X}_D)$ corresponds to the capacity C.

In order to show the advantage of Theorem 4 over the original Ozarow-Wyner bound (case of $n = 1$ and with LMMSE instead of MMPE), we consider X_D uniformly distributed with the number of points equal to $N = \lfloor\sqrt{1 + \mathsf{snr}}\rfloor$, that is, we choose the number of points such that $H(X_D) \approx \frac{1}{2}\log(1 + \mathsf{snr})$. Figure 2 shows:

- The solid cyan line is the "shaping loss" $\frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right)$ for a one-dimensional infinite lattice and is the limiting gap if the number of points N grows faster than $\sqrt{\mathsf{snr}}$;
- The solid magenta line is the gap in the original Ozarow-Wyner bound in (33); and
- The dashed purple, dashed-dotted blue and dotted green lines are the new gap given by Theorem 4 for values of $p = 2, 4, 6$, respectively, and where we choose $V \sim \mathcal{U}\left[-\frac{d_{\min}(X_D)}{2}, \frac{d_{\min}(X_D)}{2}\right]$.

We note that the version of the Ozarow-Wyner bound in Theorem 4 provides the sharpest bound for the gap term. An open question, for $n = 1$, is what value of p provide the smallest gap and whether it coincides with the ultimate "shaping loss".

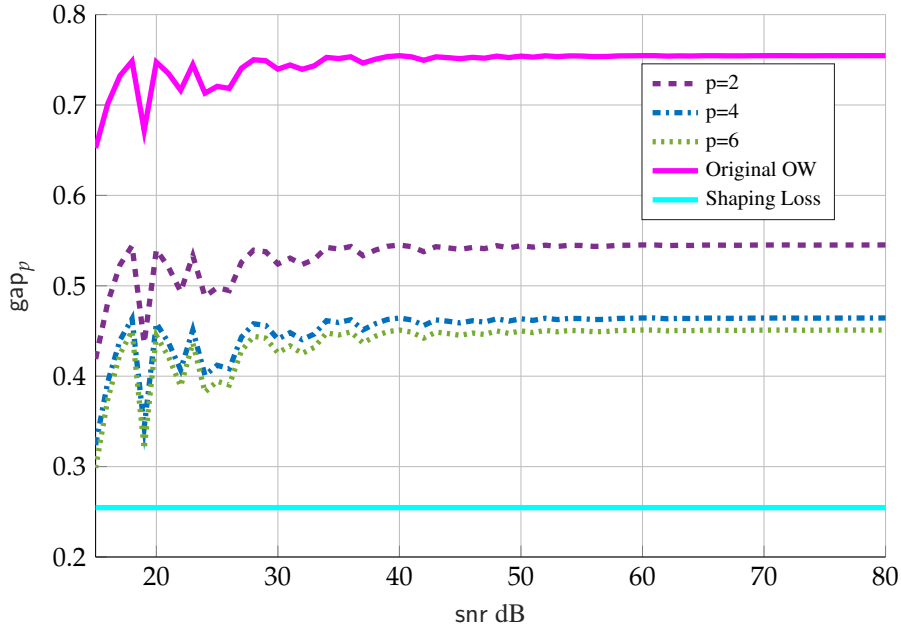


Figure 2. Gaps in Equations (33a) and (35) vs. snr.

For the AWGN channel there exist a number of other bounds that use discrete inputs as well (see [46,56–58] and references therein). The advantage of using Ozarow-Wyner type bounds, however, lies in their simplicity as they only depend on the number of signal constellation points and the minimum distance of the constellation.

The Ozarow-Wyner bound will play a key role in Sections 5 and 8 where we examine achievable schemes for a point-to-point channel with a disturbance constraint and for a two-user Gaussian interference channel.

For recent applications of the bound in Theorem 4 to non-Gaussian and MIMO channels the reader is referred to [59–61].

3.3. SNR Evolution of Optimal Codes

The I-MMSE can also be used in the analysis of the MMSE SNR-evolution of asymptotically optimal code sequences (code sequences that approach capacity in the limit of blocklength). In particular, using the I-MMSE relationship one can exactly identify the MMSE SNR-evolution of asymptotically optimal code sequences for the Gaussian point-to-point channel.

Theorem 5. *(SNR evolution of the MMSE [62,63]) Any code sequence for the Gaussian point-to-point channel attains capacity if and only if*

$$\mathrm{mmse}_\infty(\mathbf{X},\gamma) = \begin{cases} \frac{1}{1+\gamma} & \gamma \leq \mathsf{snr}, \\ 0 & \gamma \geq \mathsf{snr}. \end{cases} \tag{37}$$

Figure 3 depicts the SNR evolution of the MMSE as described in Theorem 5. The discontinuity of the MMSE at snr is ofter referred to as the *phase transition*. From Theorem 5 it is clear that the optimal point-to-point code must have the same MMSE profile as the Gaussian distribution for all SNR's before snr and experience a phase transition at snr. Intuitively, the phase transition happens because an optimal point-to-point code designed to operate at snr can be reliably decoded at snr and SNR's larger than snr, and both the decoding and estimation errors can be driven to zero. It is also important to point out that the area under (37) is twice the capacity.

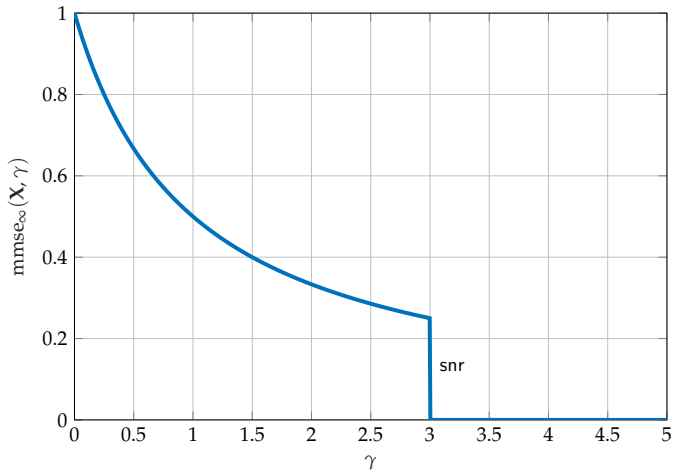
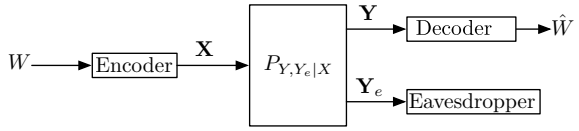


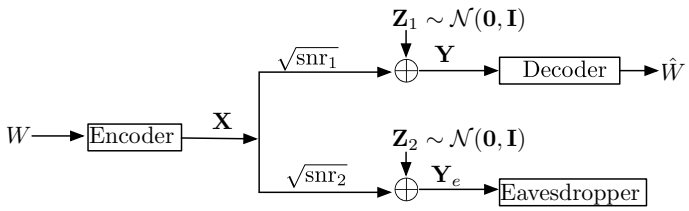
Figure 3. SNR evolution of the MMSE for $\mathsf{snr} = 3$.

4. Applications to the Wiretap Channel

In this section, by focusing on the wiretap channel, it is shown how estimation theoretic techniques can be applied to multi-user information theory. The wiretap channel, introduced by Wyner in [64], is a point-to-point channel with an additional eavesdropper (see Figure 4a). The input is denoted by $\mathbf{X}$, the output of the legitimate user is denoted by $\mathbf{Y}$, and the output of an eavesdropper is denoted by $\mathbf{Y}_\mathrm{e}$. The transmitter of $\mathbf{X}$, commonly referred to as Alice, wants to reliably communicate a message W to the legitimate receiver $\mathbf{Y}$, commonly referred to as Bob, while keeping the message W secure to some extent from the eavesdropper $\mathbf{Y}_\mathrm{e}$, commonly referred to as Eve.



(a)



(b)

Figure 4. Wiretap Channels. (**a**) The Wiretap Channel; (**b**) The Gaussian Wiretap Channel.

Definition 5. *A rate-equivocation pair (R,d) is said to be achievable over a wiretap channel if there exists a sequence of $(2^{nR},n)$ codes such that*

$$\lim_{n\to\infty} \mathbb{P}[\hat{W} \neq W] = 0, \qquad \textit{reliability constraint,} \tag{38a}$$

$$\lim_{n\to\infty} \frac{1}{n} I(W;\mathbf{Y}_e) \leq R - d, \qquad \textit{information leakage or secrecy constraint.} \tag{38b}$$

The rate-equivocation region $\mathcal{R}_s$ *is defined as the closure of all achievable rate-equivocation pairs, and the secrecy capacity is defined as*

$$C_s = \sup_R \{R : (R,R) \in \mathcal{R}_s\}. \tag{39}$$

The secrecy capacity of a general wiretap channel was shown by Csiszár and Körner [65] and is given by

$$C_s = \max_{P_{UX}} \{I(U;Y) - I(U;Y_e)\} \tag{40}$$

where U is an auxiliary random variable that satisfies the Markov relationship $U \leftrightarrow X \leftrightarrow (Y, Y_e)$.

In the case of a *degraded* wiretap channel (i.e., a wiretap channel obeying the Markov relationship $X \leftrightarrow Y \leftrightarrow Y_e$) the expression in (40) reduces to

$$C_s = \max_{P_X} \{I(X;Y) - I(X;Y_e)\}. \tag{41}$$

In fact, the expression in (41) for the degraded channel predates the expression in (40) and was shown in the original work of Wyner [64].

4.1. Converse of the Gaussian Wiretap Channel

In this section, we consider the practically relevant scalar Gaussian wiretap channel given by

$$Y = \sqrt{\mathsf{snr}} X_1 + Z_1, \tag{42a}$$
$$Y_e = \sqrt{\mathsf{snr}_0} X_2 + Z_2, \tag{42b}$$

where $\mathsf{snr} \geq \mathsf{snr}_0$, with an additional input power constraint $\mathbb{E}[X^2] \leq 1$. This setting was considered in [66], and the secrecy capacity was shown to be

$$C_s = \frac{1}{2} \log \left(\frac{1+\mathsf{snr}}{1+\mathsf{snr}_0} \right). \tag{43}$$

In contrast to the proof in [66], where the key technical tool used to maximize the expression in (41) was the EPI, by using the I-MMSE relationship the capacity in (43) can be shown via the following simple three line argument [30]:

$$I(X;Y) - I(X;Y_e) = \frac{1}{2} \int_{\mathsf{snr}_0}^{\mathsf{snr}} \mathrm{mmse}(X,t) dt \tag{44a}$$
$$\leq \frac{1}{2} \int_{\mathsf{snr}_0}^{\mathsf{snr}} \frac{1}{1+t} dt \tag{44b}$$
$$= \frac{1}{2} \log \left(\frac{1+\mathsf{snr}}{1+\mathsf{snr}_0} \right), \tag{44c}$$

where the inequality follows by using the LMMSE upper bound in (11). It is also interesting to point out that the technique in (44) can be easily mimicked to derive the entire rate-equivocation region; for details see [23].

4.2. SNR Evolution of Optimal Wiretap Codes

In the previous section, we saw that the I-MMSE relationship is a very powerful mathematical tool and can be used to provide a simple derivation of the secrecy capacity of the scalar Gaussian wiretap channel. In fact, as shown in [28,34], the I-MMSE relationship can also be used to obtain practical insights. Specifically, it was shown to be useful in identifying key properties of optimal wiretap codes.

Theorem 6. *(SNR evolution of the MMSE [28]) Any code sequence for the Gaussian wiretap channel attains a rate equivocation pair (R, C_s), meaning it attains the maximum level of equivocation, if and only if*

$$\mathrm{mmse}_\infty(\mathbf{X}; \gamma \mid W) = 0, \ \gamma \geq \mathsf{snr}_0, \tag{45}$$

and

$$\mathrm{mmse}_{\infty}(\mathbf{X};\gamma) = \begin{cases} \frac{1}{1+\gamma}, & \gamma \leq \mathsf{snr}, \\ 0, & \gamma > \mathsf{snr}, \end{cases} \tag{46}$$

regardless of the rate of the code, meaning for any $R \geq C_s$.

Note that, as shown in Theorem 5, (46) is the SNR-evolution of any point-to-point capacity achieving code sequence, C, to $\mathbf{Y}$ as shown in [62,63]; however, only a one-to-one mapping over this codebook sequence leads to the maximum point-to-point rate. The idea is that the maximum level of equivocation determines the SNR-evolution of $\mathrm{mmse}(\mathbf{X};\gamma)$ regardless of the rate.

The additional condition given in (45) is required in order to fully define the sub-group of code sequences that are (R, C_s) codes for the Gaussian wiretap channel. Still, these conditions do not fully specify the rate of the code sequence, as the group contains codes of different rates R as long as $R \geq d_{\max}$. Note that the rate of the code is determined solely by the SNR-evolution of mmse $(\mathbf{X};\gamma|W)$ in the region of $\gamma \in [0, \mathsf{snr}_0)$ and is given by

$$R = \frac{1}{2}\int_0^{\mathsf{snr}} \left[\mathrm{mmse}_{\infty}(\mathbf{X};\gamma) - \mathrm{mmse}_{\infty}(\mathbf{X};\gamma|W)\right] d\gamma. \tag{47}$$

The immediate question that arises is: Can we find MMSE properties that will distinguish code sequences of different rates? The answer is affirmative in the two extreme cases: (i) When $R = C_s$, meaning a completely secure code; (ii) When $R = C$, meaning maximum point-to-point capacity. In the latter case, one-to-one mapping is required, and the conditional MMSE is simply zero for all SNR. Figure 5 considers the former case of perfect secrecy as well as an arbitrary intermediate case in which the rate is between the secrecy capacity and the point-to-point capacity.

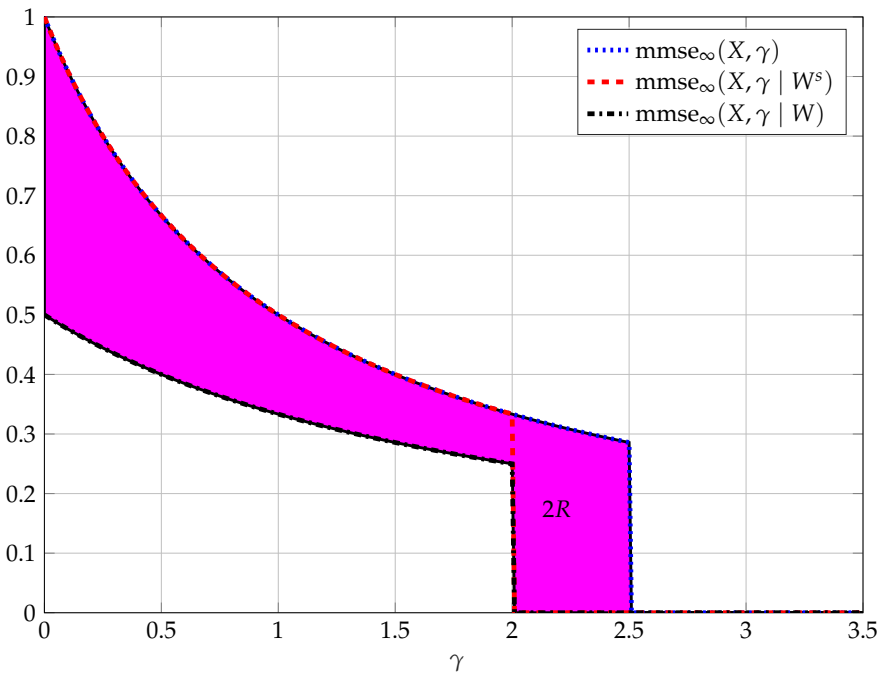


Figure 5. The above figure depicts the behavior of $\mathrm{mmse}_{\infty}(\mathbf{X};\gamma)$ as a function of γ assuming $d_{\max}$ (in dotted blue), the behavior $\mathrm{mmse}_{\infty}(\mathbf{X};\gamma|W^s)$ assuming complete secrecy (in dashed red) and the behavior of $\mathrm{mmse}_{\infty}(\mathbf{X};\gamma|W)$ for some arbitrary code of rate above secrecy capacity and below point-to-point capacity (in dash-dot black). We mark twice the rate as the area between $\mathrm{mmse}_{\infty}(\mathbf{X};\gamma)$ and $\mathrm{mmse}_{\infty}(\mathbf{X};\gamma|W)$ (in magenta). Parameters are $\mathsf{snr}_0 = 2$ and $\mathsf{snr} = 2.5$.

According to the above result, constructing a completely secure code sequence requires splitting the possible codewords into sub-codes that are asymptotically optimal for the eavesdropper. This approach is exactly the one in Wyner's original work [64], and also emphasized by Massey in [67], wherein the achievability proof the construction of the code sequence is such that the bins of each secure message are asymptotically optimal code sequences to the eavesdropper (saturating the eavesdropper). The above claim extends this observation by claiming that any mapping of messages to codewords (alternatively, any binning of the codewords) that attains complete secrecy must saturate the eavesdropper, thus supporting the known achievability scheme of Wyner. Moreover, it is important to emphasize that the maximum level of equivocation can be attained with no loss in rate, meaning the reliable receiver can continue communicating at capacity.

Another important point to note is that these results supports the necessity of a stochastic encoder for any code sequence for the Gaussian wiretap channel, achieving the maximum level of equivocation and with $R < C$ (as shown in [68] for a completely secure code for the discrete memoryless wiretap channel), since one can show that the conditions guarantee $H(\mathbf{X}|W) > 0$ for any such code sequence.

5. Communication with a Disturbance Constraint

Consider a scenario in which a message, encoded as $\mathbf{X}$, must be decoded at the primary receiver $\mathbf{Y}$ while it is also seen at the unintended/secondary receiver $\mathbf{Y}_0$ for which it is interference, as shown in Figure 6a. The transmitter wishes to maximize its communication rate, while subject to a constraint on the disturbance it inflicts on the secondary receiver, and where the disturbance is measured by some function $F(\mathbf{X}, \mathbf{Y}_0)$. It is common to refer to such a scenario as communication with a disturbance constraint. The choice of $F(\mathbf{X}, \mathbf{Y}_0)$ depends on the application one has in mind. For example, a common application is to limit the interference that the primary user inflicts on the secondary. In this case, two possible choices of $F(\mathbf{X}, \mathbf{Y}_0)$ are the mutual information $I(\mathbf{X}; \mathbf{Y}_0)$ and the MMSE $\mathrm{mmse}(\mathbf{X}|\mathbf{Y}_0)$, considered in [69,70], respectively. In what follows we review these two possible measures of disturbance, so as to explain the advantages of the MMSE as a measure of disturbance that best models the interference.

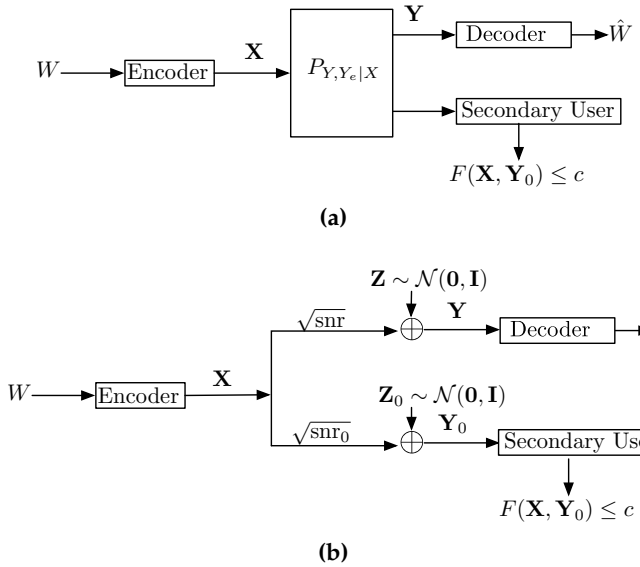


Figure 6. Channels with disturbance constraints. (**a**) A point-to-point channel with a disturbance constraint; (**b**) A Gaussian point-to-point channel with the disturbance constraint.

5.1. Max-I Problem

Consider a Gaussian noise channel and take the disturbance to be measured in terms of the MMSE (i.e., $F(\mathbf{X}, \mathbf{Y}_0) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)$), as shown on Figure 6b. Intuitively, the MMSE disturbance constraint

quantifies the remaining interference after partial interference cancellation or soft-decoding have been performed [47,70]. Formally, the following problem was considered in [50]:

Definition 6. (Max-I problem.) *For some* $\beta \in [0,1]$

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} I_n(\mathbf{X}, \mathsf{snr}), \tag{48a}$$

$$s.t.\ \|\mathbf{X}\|_2^2 \leq 1,\ \textit{power constraint}, \tag{48b}$$

$$\textit{and}\ \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0},\ \textit{MMSE constraint}. \tag{48c}$$

The subscript n in $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ emphasizes that we consider length n inputs $\mathbf{X} \in \mathbb{R}^n$. Clearly $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ is a non-decreasing function of n. The scenario depicted in Figure 6b is captured when $n \to \infty$ in the Max-I problem, in which case the objective function has a meaning of reliable achievable rate.

The scenario modeled by the Max-I problem is motivated by the two-user *Gaussian interference channel* (G-IC), whose capacity is known only for some special cases. The following strategies are commonly used to manage interference in the G-IC:

1. *Interference is treated as Gaussian noise*: in this approach the interference is not explicitly decoded. Treating interference as noise with Gaussian codebooks has been shown to be sum-capacity optimal in the so called very-weak interference regime [71–73].
2. *Partial interference cancellation*: by using the Han-Kobayashi (HK) achievable scheme [74], part of the interfering message is jointly decoded with part of the desired signal. Then the decoded part of the interference is subtracted from the received signal, and the remaining part of the desired signal is decoded while the remaining part of the interference is treated as Gaussian noise. With Gaussian codebooks, this approach has been shown to be capacity achieving in the strong interference regime [75] and optimal within 1/2 bit per channel per user otherwise [76].
3. *Soft-decoding/estimation*: the unintended receiver employs soft-decoding of part of the interference. This is enabled by using non-Gaussian inputs and designing the decoders that treat interference as noise by taking into account the correct (non-Gaussian) distribution of the interference. Such scenarios were considered in [44,46,49], and shown to be optimal to within either a constant or a $O(\log\log(\mathsf{snr}))$ gap for all regimes in [45].

Even though the Max-I problem is somewhat simplified, compared to that of determining the capacity of the G-IC, as it ignores the existence of the second transmission, it can serve as an important building block towards characterizing the capacity of the G-IC [47,70], especially in light of the known (but currently uncomputable) limiting expression for the capacity region [77]:

$$\mathcal{C}_\infty^{\mathrm{IC}} = \lim_{n\to\infty} \mathsf{co} \bigcup_{P_{\mathbf{X}_1\mathbf{X}_2} = P_{\mathbf{X}_1} P_{\mathbf{X}_2}} \left\{ \begin{array}{l} 0 \leq R_1 \leq I_n(\mathbf{X}_1; \mathbf{Y}_1) \\ 0 \leq R_2 \leq I_n(\mathbf{X}_2; \mathbf{Y}_2) \end{array} \right\}, \tag{49}$$

where co denotes the convex closure operation. Moreover, observe that for any finite n we have that the capacity region can be inner bounded by

$$\mathcal{C}_n^{\mathrm{IC}} \subset \mathcal{C}_\infty^{\mathrm{IC}}, \tag{50}$$

where

$$\mathcal{C}_n^{\mathrm{IC}} = \mathsf{co} \bigcup_{P_{\mathbf{X}_1\mathbf{X}_2} = P_{\mathbf{X}_1} P_{\mathbf{X}_2}} \left\{ \begin{array}{l} 0 \leq R_1 \leq I_n(\mathbf{X}_1; \mathbf{Y}_1) \\ 0 \leq R_2 \leq I_n(\mathbf{X}_2; \mathbf{Y}_2) \end{array} \right\}. \tag{51}$$

The inner bound $\mathcal{C}_n^{\mathrm{IC}}$ will be referred to as the *treating interference as noise (TIN)* inner bound. Finding the input distributions $P_{\mathbf{X}_1}P_{\mathbf{X}_2}$ that exhaust the achievable region in $\mathcal{C}_n^{\mathrm{IC}}$ is an important open problem. In Section 8, for a special case of $n = 1$, we will demonstrate that $\mathcal{C}_1^{\mathrm{IC}}$ is within a constant or $O(\log\log(\mathsf{snr}))$ from the capacity $\mathcal{C}_\infty^{\mathrm{IC}}$. Therefore, the Max-I problem, denoted by $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (48), can serve as an important step in characterizing the structure of optimal input distributions for $\mathcal{C}_n^{\mathrm{IC}}$. We also note that in [47,70] it was conjectured that the optimal input for $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ is discrete. For other recent works on optimizing the TIN region in (51), we refer the reader to [43,46,49,78,79] and the references therein.

The importance of studying models of communication systems with disturbance constraints has been recognized previously. For example, in [69] Bandemer et al. studied the following problem related to the Max-I problem in (48).

Definition 7. *(Bandemer et al. problem [69]) For some $R \geq 0$*

$$\mathcal{I}_n(\mathsf{snr}_0, \mathsf{snr}, R) := \max_{\mathbf{X}} I_n(\mathbf{X}, \mathsf{snr}_0), \tag{52a}$$

$$\text{s.t. } \|\mathbf{X}\|_2^2 \leq 1, \text{ power constraint,} \tag{52b}$$

$$\text{and } I_n(\mathbf{X}, \mathsf{snr}_0) \leq R, \text{ disturbance constraint.} \tag{52c}$$

In [69] it was shown that the optimal solution for $\mathcal{I}_n(\mathsf{snr}, \mathsf{snr}_0, R)$, for any n, is attained by $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \alpha\mathbf{I})$ where $\alpha = \min\left(1, \frac{\mathrm{e}^{2R}-1}{\mathsf{snr}_0}\right)$; here α is such that the most stringent constraint between (52b) and (52c) is satisfied with equality. In other words, the optimal input is independent and identically distributed (i.i.d.) Gaussian with power reduced such that the disturbance constraint in (52c) is not violated.

Theorem 7 ([69])**.** *The rate-disturbance region of the problem in (52) is given by*

$$\mathcal{I}_n(\mathsf{snr}_0, \mathsf{snr}, R) \leq \frac{1}{2}\log\left(1 + \alpha\mathsf{snr}\right), \tag{53}$$

with equality if and only if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \alpha\mathbf{I})$ where $\alpha = \min\left(1, \frac{\mathrm{e}^{2R}-1}{\mathsf{snr}_0}\right)$.

Measuring the disturbance with the mutual information as in (52), in contrast to the MMSE as in (48), suggests that it is always optimal to use Gaussian codebooks with reduced power without any rate splitting. Moreover, while the mutual information constraint in (52) limits the amount of information transmitted to the unintended receiver, it may not be the best choice for measuring the interference, since any information that can be reliably decoded by the unintended receiver is not really interference. For this reason, it has been argued in [47,70] that the Max-I problem in (48) with the MMSE disturbance constraint is a more suitable building block to study the G-IC, since the MMSE constraint accounts for the interference, and captures the key role of rate splitting.

We also refer the reader to [80] where, in the context of discrete memoryless channels, the disturbance constraint was modeled by controlling the type (i.e., empirical distribution) of the interference at the secondary user. Moreover, the authors of [80] were able to characterize the tradeoff between the rate and the type of the induced interference by exactly characterizing the capacity region of the problem at hand.

We first consider a case of the Max-I problem when $n \to \infty$.

5.2. Characterization of $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ as $n \to \infty$

For the practically relevant case of $n \to \infty$, which has an operational meaning, $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ has been characterized in [70] and is given by the following theorem.

Theorem 8 ([70]). *For any* $\mathsf{snr}, \mathsf{snr}_0 \geq 0$ *and* $\beta \in [0,1]$

$$\begin{aligned}
\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) &= \lim_{n\to\infty} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta), \\
&= \begin{cases} \frac{1}{2}\log(1+\mathsf{snr}), & \mathsf{snr} \leq \mathsf{snr}_0, \\ \frac{1}{2}\log(1+\beta\mathsf{snr}) + \frac{1}{2}\log\left(1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}\right), & \mathsf{snr} \geq \mathsf{snr}_0, \end{cases} \\
&= \frac{1}{2}\log^+\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(1+\min(\mathsf{snr}, \mathsf{snr}_0)\right), \qquad (54)
\end{aligned}$$

which is achieved by using superposition coding with Gaussian codebooks.

The proof of the achievability part of Theorem 8 is by using superposition coding and is outside of the scope of this work. The interested reader is referred to [63,70,81] for a detailed treatment of MMSE properties of superposition codes.

Next, we show a converse proof of Theorem 8. In addition, to the already familiar use of the LMMSE bound technique, as in the wiretap channel in Section 4.1, we also show an application of the SCPP bound. The proof for the case of $\mathsf{snr} \leq \mathsf{snr}_0$ follows by ignoring the MMSE constraint at snr_0 and using the LMMSE upper bound

$$\begin{aligned}
I_n(\mathbf{X}, \mathsf{snr}) &= \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \\
&\leq \frac{1}{2}\int_0^{\mathsf{snr}} \frac{1}{1+t}dt \\
&= \frac{1}{2}\log(1+\mathsf{snr})dt.
\end{aligned}$$

Next, we focus on the case of $\mathsf{snr} \geq \mathsf{snr}_0$

$$\begin{aligned}
I_n(\mathbf{X}, \mathsf{snr}) &= \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \\
&= \frac{1}{2}\int_0^{\mathsf{snr}_0} \mathrm{mmse}(\mathbf{X}, t)dt + \frac{1}{2}\int_{\mathsf{snr}_0}^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \\
&\leq \frac{1}{2}\int_0^{\mathsf{snr}_0} \frac{1}{1+t}dt + \frac{1}{2}\int_{\mathsf{snr}_0}^{\mathsf{snr}} \frac{\beta}{1+\beta t}dt \\
&= \frac{1}{2}\log(1+\beta\mathsf{snr}) + \frac{1}{2}\log\left(1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}\right),
\end{aligned}$$

where the last inequality follows by upper bounding the integral over $[0, \mathsf{snr}_0]$ by the LMMSE bound in (11) and by upper bounding the integral over $[\mathsf{snr}_0, \mathsf{snr}]$ using the SCPP bound in (21).

Figure 7 shows a plot of $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (54) normalized by the capacity of the point-to-point channel $\frac{1}{2}\log(1+\mathsf{snr})$. The region $\mathsf{snr} \leq \mathsf{snr}_0$ (flat part of the curve) is where the MMSE constraint is inactive since the channel with snr_0 can decode the interference and guarantee zero MMSE. The regime $\mathsf{snr} \geq \mathsf{snr}_0$ (curvy part of the curve) is where the receiver with snr_0 can no-longer decode the interference and the MMSE constraint becomes active, which in practice is the more interesting regime because the secondary receiver experiences "weak interference" that cannot be fully decoded (recall that in this regime superposition coding appears to be the best achievable strategy for the two-user Gaussian interference channel, but it is unknown whether it achieves capacity [76]).

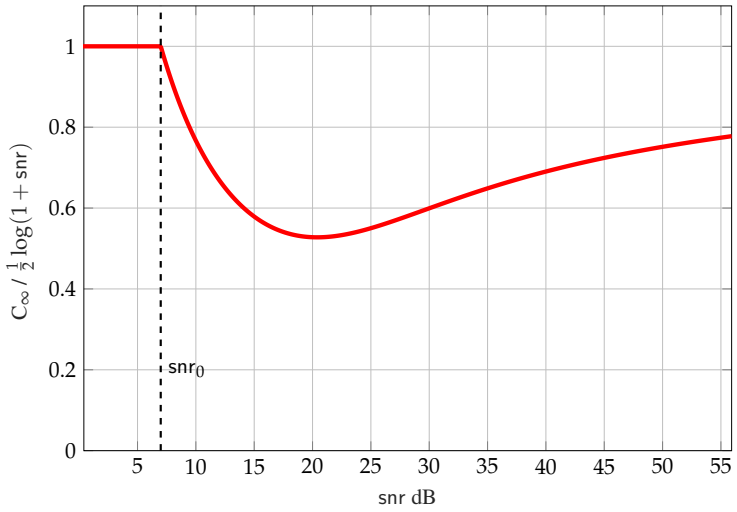


Figure 7. Plot of $\frac{\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta)}{\frac{1}{2}\log(1+\mathsf{snr})}$ vs. snr in dB, for $\beta = 0.01$, $\mathsf{snr}_0 = 5 = 6.989$ dB.

5.3. Proof of the Disturbance Constraint Problem with a Mutual Information Constraint

In this section we show that the mutual information disturbance constraint problem in (52) can also be solved via an estimation theoretic approach.

An Alternative Proof of the Converse Part of Theorem 7. Observe that, similarly to the Max-I problem, the interesting case of the $\mathcal{I}_n(\mathsf{snr}_0, \mathsf{snr}, R)$ is the "weak interference" regime (i.e., $\mathsf{snr} \geq \mathsf{snr}_0$). This, follows since for the "strong interference" regime (i.e., $\mathsf{snr} \leq \mathsf{snr}_0$) the result follows trivially by the data processing inequality

$$I_n(\mathbf{X}, \mathsf{snr}) \leq I_n(\mathbf{X}, \mathsf{snr}_0) \leq R, \tag{55}$$

and maximizing (55) under the power constraint. To show Theorem 7, for the case of $\mathsf{snr} \geq \mathsf{snr}_0$, observe that

$$0 \leq I_n(\mathbf{X}, \mathsf{snr}_0) \leq \frac{1}{2}\log\left(1 + \mathsf{snr}_0\right), \tag{56}$$

where the inequality on the right is due to the power constraint on $\mathbf{X}$. Therefore, there exists some $\alpha \in [0, 1]$ such that

$$I_n(\mathbf{X}, \mathsf{snr}_0) = \frac{1}{2}\log\left(1 + \alpha\mathsf{snr}_0\right). \tag{57}$$

Using the I-MMSE, (57) can be written as

$$\frac{1}{2}\int_0^{\mathsf{snr}_0} \mathrm{mmse}(\mathbf{X}, t)dt = \frac{1}{2}\int_0^{\mathsf{snr}_0} \frac{\alpha}{1+\alpha t}dt. \tag{58}$$

From (58) and SCPP property we conclude that $\mathrm{mmse}(\mathbf{X}, t)$ and $\frac{\alpha}{1+\alpha t}$ are either equal for all t, or cross each other once in the region $[0, \mathsf{snr}_0)$. In both cases, by the SCPP, we have

$$\mathrm{mmse}(\mathbf{X}, t) \leq \frac{\alpha}{1+\alpha t}, \forall t \in [\mathsf{snr}_0, \infty). \tag{59}$$

We are now in the position to bound the main term of the disturbance constrained problem. By using the I-MMSE relationship the mutual information can be bounded as follows:

$$
\begin{aligned}
I_n(\mathbf{X}, \mathsf{snr}) &= \frac{1}{2}\int_0^{\mathsf{snr}_0} \mathrm{mmse}(\mathbf{X}, t)dt + \frac{1}{2}\int_{\mathsf{snr}_0}^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \\
&= \frac{1}{2}\log\left(1+\alpha\mathsf{snr}_0\right) + \frac{1}{2}\int_{\mathsf{snr}_0}^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \\
&\leq \frac{1}{2}\log\left(1+\alpha\mathsf{snr}_0\right) + \frac{1}{2}\int_{\mathsf{snr}_0}^{\mathsf{snr}} \frac{\alpha}{1+\alpha t}dt \qquad (60) \\
&= \frac{1}{2}\log\left(1+\alpha\mathsf{snr}\right), \qquad (61)
\end{aligned}
$$

where the bound in (60) follows by the inequality in (59). The proof of the converse is concluded by establishing that the maximum value of α in (61) is given by $\alpha = \min\left(1, \frac{\mathrm{e}^{2R}-1}{\mathsf{snr}_0}\right)$ which is a consequence of the bound $I_n(\mathbf{X}, \mathsf{snr}_0) \leq R$.

This concludes the proof of the converse. □

The achievability proof of Theorem 7 follows by using an i.i.d. Gaussian input with power α. This concludes the proof of Theorem 7.

In contrast to the proof in [69] which appeals to the EPI, the proof outlined here only uses the SCPP and the I-MMSE. Note, that unlike the proof of the converse of the Max-I problem, which also requires the LMMSE bound, the only ingredient in the proof of the converse for $\mathcal{I}_n(\mathsf{snr}_0, \mathsf{snr}, R)$ is a clever use of the SCPP bound. In Section 6, we will make use of this technique and show a converse proof for the scalar Gaussian broadcast channel.

Another observation is that the achievability proof of the $\mathcal{I}_n(\mathsf{snr}_0, \mathsf{snr}, R)$ holds for an arbitrary finite n while the achievability proof of the Max-I problem holds only as $n \to \infty$. In the next section, we demonstrate techniques for how to extend the achievability of the Max-I problem to the case of finite n. These techniques will ultimately be used to show an approximate optimality of the TIN inner bound for the two-user G-IC in Section 8.

5.4. Max-MMSE Problem

The Max-I problem in (48) is closely related to the following optimization problem.

Definition 8. *(Max-MMSE problem [50,82]) For some* $\beta \in [0, 1]$

$$
\begin{aligned}
&\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}), && (62a) \\
&\textit{s.t. } \|\mathbf{X}\|_2^2 \leq 1, \textit{ power constraint}, && (62b) \\
&\textit{and } \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}, \textit{ MMSE constraint}. && (62c)
\end{aligned}
$$

The authors of [63,70] proved that

$$
\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) = \lim_{n\to\infty} \mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} < \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr} \geq \mathsf{snr}_0, \end{cases} \qquad (63)
$$

achieved by superposition coding with Gaussian codebooks. Clearly there is a discontinuity in (63) at $\mathsf{snr} = \mathsf{snr}_0$ for $\beta < 1$. This fact is a well known property of the MMSE, and it is referred to as a *phase transition* [63].

The LMMSE bound provides the converse solution for $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (63) in the regime $\mathsf{snr} \leq \mathsf{snr}_0$. An interesting observation is that in this regime the knowledge of the MMSE at snr_0 is not used. The SCPP bound provides the converse in the regime $\mathsf{snr} \leq \mathsf{snr}_0$ and, unlike the LMMSE bound, does use the knowledge of the value of MMSE at snr_0.

The solution of the Max-MMSE problem provides an upper bound on the Max-I problem (for every n including in the limit as $n \to \infty$), through the I-MMSE relationship

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt \leq \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{M}_n(t, \mathsf{snr}_0, \beta)dt. \tag{64}$$

The reason is that in the Max-MMSE problem one maximizes the integrand in the I-MMSE relationship for every γ, and the maximizing input may have a different distribution for each γ. The surprising result is that in the limit as $n \to \infty$ we have equality, meaning that in the limit there exists an input that attains the maximum Max-MMSE solution for every γ. In other words, the integration of $\mathrm{M}_\infty(\gamma, \mathsf{snr}_0, \beta)$ over $\gamma \in [0, \mathsf{snr}]$ results in $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$. In view of the relationship in (64) we focus on the $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ problem.

Note that SCPP gives a solution to the Max-MMSE problem in (62) for $\mathsf{snr} \geq \mathsf{snr}_0$ and any $n \geq 1$ as follows:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta\mathsf{snr}}, \text{ for } \mathsf{snr} \geq \mathsf{snr}_0, \tag{65}$$

achieved by $\mathbf{X} \sim \mathcal{N}(0, \beta\mathbf{I})$.

However, for $\mathsf{snr} \leq \mathsf{snr}_0$, where the LMMSE bound (11) is used without taking the constraint into account, it is no longer tight for every $n \geq 1$. Therefore, the emphasis in the treatment of the Max-MMSE problem is on the regime $\mathsf{snr} \leq \mathsf{snr}_0$. In other words, the phase transition phenomenon can only be observed as $n \to \infty$, and for any finite n the LMMSE bound on the MMSE at $\mathsf{snr} \leq \mathsf{snr}_0$ must be sharpened, as the MMSE constraint at snr_0 must restrict the input in such a way that would effect the MMSE performance at $\mathsf{snr} \leq \mathsf{snr}_0$. We refer to the upper bounds in the regime $\mathsf{snr} \leq \mathsf{snr}_0$ as complementary SCPP bounds. Also, for any finite n, $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ is a continuous function of snr [30]. Putting these two facts together we have that, for any finite n, the objective function $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ must be continuous in snr and converge to a function with a jump-discontinuity at snr_0 as $n \to \infty$. Therefore, $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ must be of the following form:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} \leq \mathsf{snr}_L, \\ T_n(\mathsf{snr}, \mathsf{snr}_0, \beta), & \mathsf{snr}_L \leq \mathsf{snr} \leq \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \leq \mathsf{snr}, \end{cases} \tag{66}$$

for some snr_L. The goal is to characterize snr_L in (66) and the continuous function $T_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ such that

$$T_n(\mathsf{snr}_L, \mathsf{snr}_0, \beta) = \frac{1}{1+\mathsf{snr}_L}, \tag{67a}$$

$$T_n(\mathsf{snr}_0, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta\mathsf{snr}_0}, \tag{67b}$$

and give scaling bounds on the width of the phase transition region defined as

$$W_n := \mathsf{snr}_0 - \mathsf{snr}_L. \tag{68}$$

In other words, the objective is to understand the behavior of the MMSE phase transitions for arbitrary finite n by obtaining complementary upper bounds on the SCPP. We first focus on upper bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$.

Theorem 9. *(D-Bound [50]) For any* $\mathbf{X}$ *and* $0 < \mathsf{snr} \leq \mathsf{snr}_0$*, we have*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) + k_n\left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}\right), \tag{69a}$$

$$k_n \leq n+2, \tag{69b}$$

The proof of Theorem 9 can be found in [50] and relies on developing bounds on the derivative of the MMSE with respect to the SNR.

Theorem 10. *(M-Bound [25]) For* $0 < \mathsf{snr} \leq \mathsf{snr}_0$,

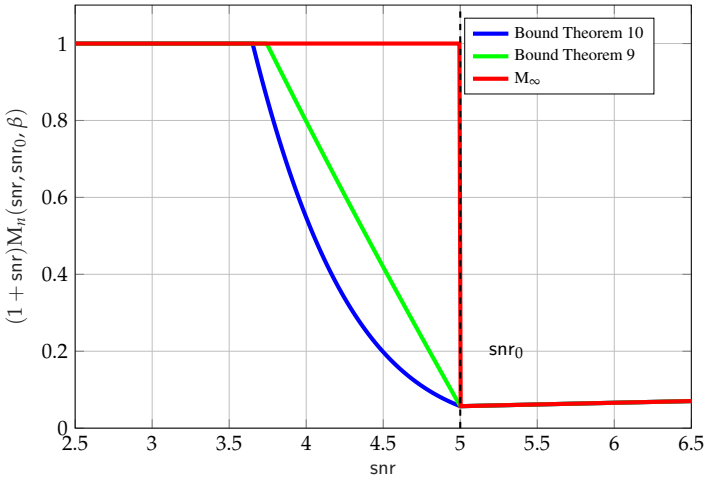
$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \min_{r > \frac{2}{\gamma}} \kappa(r, \gamma, n) \left(\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)\right)^{\frac{\gamma r - 2}{r-2}}, \tag{70a}$$

where $\gamma := \frac{\mathsf{snr}}{2\mathsf{snr}_0 - \mathsf{snr}} \in (0, 1]$, *and*

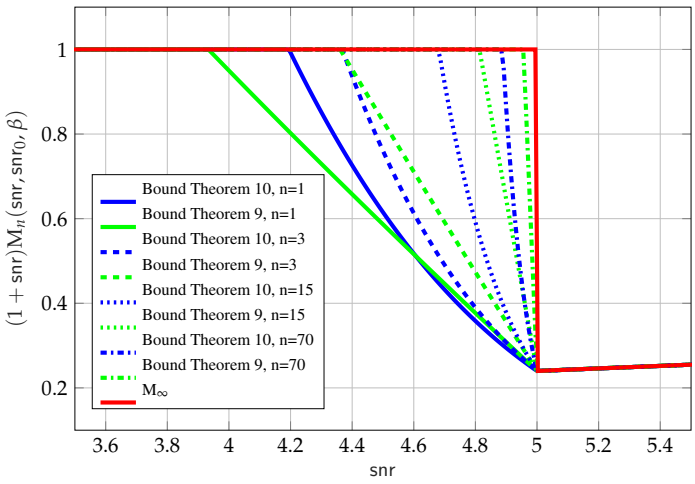
$$\kappa(r, \gamma, n) := \frac{\sqrt{2}}{n^{1-\gamma}} \left(\frac{1+\gamma}{\gamma}\right)^{\frac{n(1-\gamma)-1}{2}} M_r^{\frac{2(1-\gamma)}{r-2}}, \tag{70b}$$

$$M_r := \|\mathbf{X} - \mathbb{E}\left[\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}\right]\|_r^r \leq 2^r \min\left(\frac{\|\mathbf{Z}\|_r^r}{\mathsf{snr}_0^{\frac{r}{2}}}, \|\mathbf{X}\|_r^r\right). \tag{70c}$$

The bounds in (69a) and in (70a) are shown in Figure 8. The key observation is that the bounds in (69a) and in (70a) are sharper versions of the LMMSE bound that take into account the value of the MMSE at snr_0. It is interesting to observe how the bounds converge with n going to ∞.



(a)



(b)

Figure 8. Upper bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr. (**a**) For $\mathsf{snr}_0 = 5$ and $\beta = 0.01$. Here $n = 1$; (**b**) For $\mathsf{snr}_0 = 5$ and $\beta = 0.05$. Several values of n.

The bound in (70a) is asymptotically tighter than the one in (69a). It can be shown that the phase transition region shrinks as $O\left(\frac{1}{\sqrt{n}}\right)$ for (70a), and as $O\left(\frac{1}{n}\right)$ for the bound in (69a). It is not possible in general to assert that (70a) is tighter than (69a). In fact, for small values of n, the bound in (69a) can offer advantages, as seen for the case $n = 1$ shown in Figure 8b. Another advantage of the bound in (69a) is its analytical simplicity.

With the bounds in (69a) and in (70a) at our disposal we can repeat the converse proof outline in (61).

5.5. Mixed Inputs

Another question that arises, in the context of finite n, is how to mimic the achievability of superposition codes? Specifically, how to select an input that will maximize $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ when $\mathsf{snr} \leq \mathsf{snr}_0$.

We propose to use the following input, which in [45] was termed a *mixed input*:

$$\mathbf{X}_{\text{mix}} := \sqrt{1-\delta}\mathbf{X}_D + \sqrt{\delta}\mathbf{X}_G,\ \delta \in [0,1]: \tag{71}$$

$$\mathbf{X}_G \sim \mathcal{N}(0, \mathbf{I}),\ \|\mathbf{X}_D\|_2^2 \leq 1,\ \frac{1}{n}H(\mathbf{X}_D) < \infty, \tag{72}$$

where $\mathbf{X}_G$ and $\mathbf{X}_D$ are independent. The parameter δ and the distribution of $\mathbf{X}_D$ are to be optimized over.

The behavior of the input in (71) exibits many properties of superposition codes and we will see that the discrete part $\mathbf{X}_D$ will behave as the common message and the Gaussian part $\mathbf{X}_G$ will behave as the private message.

The input $\mathbf{X}_{\text{mix}}$ exhibits a decomposition property via which the MMSE and the mutual information can be written as the sum of the MMSE and the mutual information of the $\mathbf{X}_D$ and $\mathbf{X}_G$ components, albeit at different SNR values.

Proposition 7 ([50]). *For* $\mathbf{X}_{\text{mix}}$ *defined in* (71) *we have that*

$$I_n(\mathbf{X}_{\text{mix}}, \mathsf{snr}) = I_n\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + I_n(\mathbf{X}_G, \mathsf{snr}\,\delta), \tag{73a}$$

$$\mathrm{mmse}(\mathbf{X}_{\text{mix}}, \mathsf{snr}) = \frac{1-\delta}{(1+\mathsf{snr}\delta)^2}\mathrm{mmse}\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \delta\,\mathrm{mmse}(\mathbf{X}_G, \mathsf{snr}\,\delta). \tag{73b}$$

Observe that Proposition 7 implies that, in order for mixed inputs (with $\delta < 1$) to comply with the MMSE constraint in (48c) and (62c), the MMSE of $\mathbf{X}_D$ must satisfy

$$\mathrm{mmse}\left(\mathbf{X}_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) \leq \frac{(\beta-\delta)(1+\delta\mathsf{snr}_0)}{(1-\delta)(1+\beta\mathsf{snr}_0)}. \tag{74}$$

Proposition 7 is particularly useful because it allows us to design the Gaussian and discrete components of the mixed input independently.

Next, we evaluate the performance of $\mathbf{X}_{\text{mix}}$ in $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ for the important special case of $n = 1$. Figure 9 shows upper and lower bounds on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ where we show the following:

- The $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ upper bound in (63) (solid red line);
- The upper D-bound (69a) (dashed cyan line) and upper M-bound (dashed red line) (70a);
- The Gaussian-only input (solid green line), with $X \sim \mathcal{N}(0, \beta)$, where the power has been reduced to meet the MMSE constraint;
- The mixed input (blue dashed line), with the input in (71). We used Proposition 7 where we optimized over X_D for $\delta = \beta\frac{\mathsf{snr}_0}{1+\mathsf{snr}_0}$. The choice of δ is motivated by the scaling property of the MMSE, that is, $\delta\mathrm{mmse}(X_G, \mathsf{snr}\delta) = \mathrm{mmse}(\sqrt{\delta}X_G, \mathsf{snr})$, and the constraint on the discrete component in (74). That is, we chose δ such that the power of X_G is approximately β while the MMSE constraint on X_D in (74) is not equal to zero. The input X_D used in Figure 9 was found by a local search algorithm on the space of distributions with $N = 3$, and resulted in

$X_D = [-1.8412, -1.7386, 0.5594]$ with $P_X = [0.1111, 0.1274, 0.7615]$, which we do not claim to be optimal;

- The discrete-only input (Discrete 1 brown dashed-dotted line), with $X_D = [-1.8412, -1.7386, 0.5594]$ with $P_X = [0.1111, 0.1274, 0.7615]$, that is, the same discrete part of the above mentioned mixed input. This is done for completeness, and to compare the performance of the MMSE of the discrete component of the mixed input with and without the Gaussian component; and
- The discrete-only input (Discrete 2 dotted magenta line), with $X_D = [-1.4689, -1.1634, 0.7838]$ with $P_X = [0.1282, 0.2542, 0.6176]$, which was found by using a local search algorithm on the space of discrete-only distributions with $N = 3$ points.

The choice of $N = 3$ is motivated by the fact that it requires roughly $N = \lfloor \sqrt{1 + \mathsf{snr}_0} \rfloor$ points for the PAM input to approximately achieve capacity of the point-to-point channel with SNR value snr_0.

On the one hand, Figure 9 shows that, for $\mathsf{snr} \geq \mathsf{snr}_0$, a Gaussian-only input with power reduced to β maximizes $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in agreement with the SCPP bound (green line). On the other hand, for $\mathsf{snr} \leq \mathsf{snr}_0$, we see that discrete-only inputs (brown dashed-dotted line and magenta dotted line) achieve higher MMSE than a Gaussian-only input with reduced power. Interestingly, unlike Gaussian-only inputs, discrete-only inputs do not have to reduce power in order to meet the MMSE constraint. The reason discrete-only inputs can use full power, as per the power constraint only, is because their MMSE decreases fast enough (exponentially in SNR) to comply with the MMSE constraint. However, for $\mathsf{snr} \geq \mathsf{snr}_0$, the behavior of the MMSE of discrete-only inputs, as opposed to mixed inputs, prevents it from being optimal; this is due to their exponential tail behavior. The mixed input (blue dashed line) gets the best of both (Gaussian-only and discrete-only) worlds: it has the behavior of Gaussian-only inputs for $\mathsf{snr} \geq \mathsf{snr}_0$ (without any reduction in power) and the behavior of discrete-only inputs for $\mathsf{snr} \leq \mathsf{snr}_0$. This behavior of mixed inputs turns out to be important for the Max-I problem, where we need to choose an input that has the largest area under the MMSE curve.

Finally, Figure 9 shows the achievable MMSE with another discrete-only input (Discrete 2, dotted magenta line) that achieves higher MMSE than the mixed input for $\mathsf{snr} \leq \mathsf{snr}_0$ but lower than the mixed input for $\mathsf{snr} \geq \mathsf{snr}_0$. This is again due to the tail behavior of the MMSE of discrete inputs. The reason this second discrete input is not used as a component of the mixed inputs is because this choice would violate the MMSE constraint on X_D in (74). Note that the difference between Discrete 1 and 2 is that, Discrete 1 was found as an optimal discrete component of a mixed input (i.e., $\delta = \beta \frac{\mathsf{snr}_0}{1+\mathsf{snr}_0}$), while Discrete 2 was found as an optimal discrete input without a Gaussian component (i.e., $\delta = 0$).

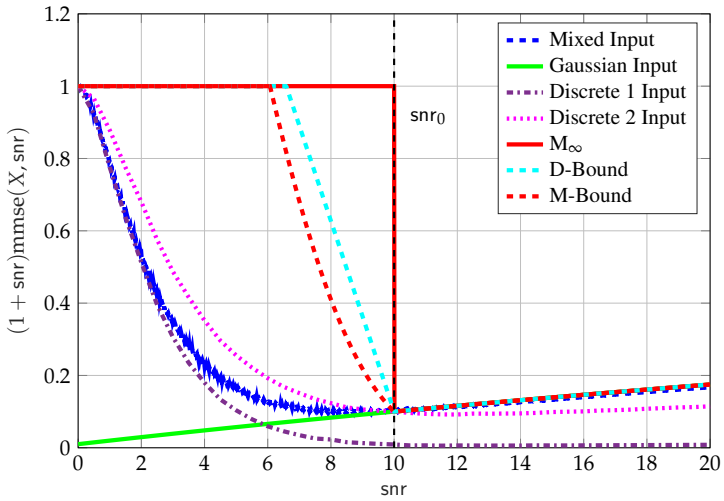


Figure 9. Upper and lower bounds on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.01$, $\mathsf{snr}_0 = 10$.

We conclude this section by demonstrating that an inner bound on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ with the mixed input in (71) is to within an additive gap of the outer bound.

Theorem 11 ([50]). *A lower bound on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ with the mixed input in (71), with $X_D \sim \mathsf{PAM}$ and with input parameters as specified in Table 1, is to within $O\left(\log\log\left(\frac{1}{\mathrm{mmse}(X,\mathsf{snr}_0)}\right)\right)$.*

Table 1. Parameters of the mixed input in (71) used in the proof of Proposition 11.

Regime	Input Parameters
Weak Interference ($\mathsf{snr} \geq \mathsf{snr}_0$)	$N = \left\lfloor \sqrt{1 + c_1 \frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}} \right\rfloor$, $c_1 = \frac{3}{2\log\left(\frac{12(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)}$, $\delta = \beta\frac{\mathsf{snr}_0}{1+\mathsf{snr}_0}$
Strong Interference ($\mathsf{snr} \leq \mathsf{snr}_0$)	$N = \lfloor \sqrt{1 + c_2\mathsf{snr}} \rfloor$, $c_2 = \frac{3}{2\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)}$, $\delta = 0$

We refer the reader to [50] for the details of the proof and extension of Theorem 11 to arbitrary n.

Please note that the gap result in Proposition 11 is constant in snr (i.e., independent of snr) but not in snr_0. Figure 10 compares the inner bounds on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$, normalized by the point-to-point capacity $\frac{1}{2}\log(1+\mathsf{snr})$, with mixed inputs (dashed magenta line) in Proposition 11 to:

- The $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ upper bound in (54) (solid red line);
- The upper bound from integration of the bound in (69a) (dashed blue line);
- The upper bound from integration of the bound in (70a) (dashed red line); and
- The inner bound with $X \sim \mathcal{N}(0, \beta)$, where the reduction in power is necessary to satisfy the MMSE constraint $\mathrm{mmse}(X, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}$ (dotted green line).

Figure 10 shows that Gaussian inputs are sub-optimal and that mixed inputs achieve large degrees of freedom compared to Gaussian inputs. Interestingly, in the regime $\mathsf{snr} \leq \mathsf{snr}_0$, it is approximately optimal to set $\delta = 0$, that is, only the discrete part of the mixed input is used. This in particular supports the conjecture in [70] that discrete inputs may be optimal for $n = 1$ and $\mathsf{snr} \leq \mathsf{snr}_0$. For the case $\mathsf{snr} \geq \mathsf{snr}_0$ our results partially refute the conjecture by excluding the possibility of discrete inputs with finitely many points from being optimal.

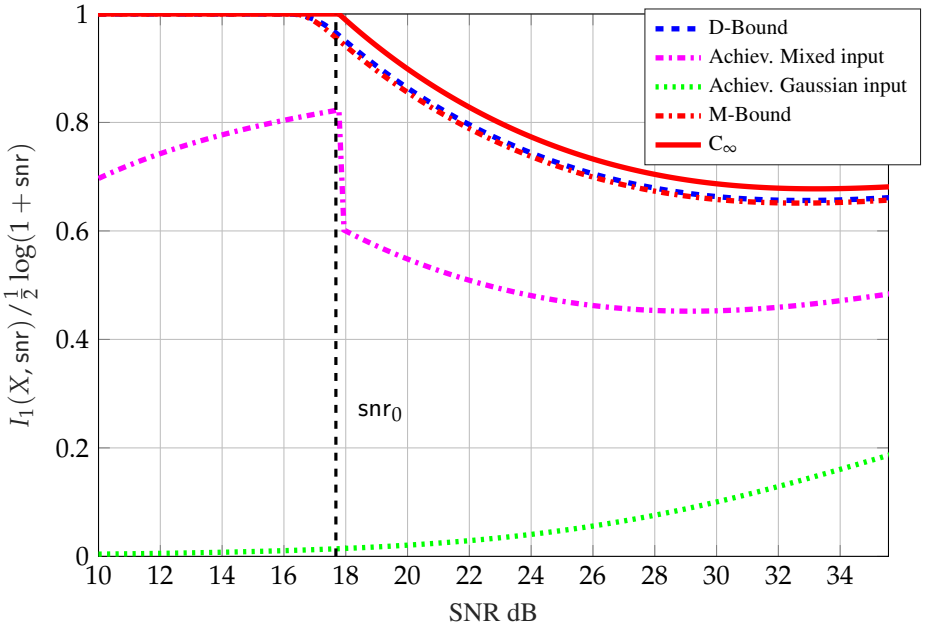


Figure 10. Upper and lower bounds on $\mathcal{C}_{n=1}(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.001$ and $\mathsf{snr}_0 = 60 = 17.6815$ dB.

The key intuition developed in this section about the mixed input and its close resemblance to superposition coding will be used in Section 8 to show approximate optimality of TIN for the two-user G-IC.

6. Applications to the Broadcast Channel

The broadcast channel (BC), introduced by Cover in [83], is depicted in Figure 11a. In the BC the goal of the transmitter is to reliably transmit the message W_1 to receiver 1 and the message W_2 to receiver 2. The transmitter encodes the pair of messages (W_1, W_2) into a transmitted codeword $\mathbf{X}$ of length n. Receiver 1 receives the sequence $\mathbf{Y}_1$ of length n and receiver 2 receives the sequence $\mathbf{Y}_2$ of length n. They both try to decode their respective messages from their received sequence. An achievable rate pair is defined as follows:

Definition 9. *A rate pair (R_1, R_2) is said to be achievable for each n, for a message W_1 of cardinality 2^{nR_1} and a message W_2 of cardinality 2^{nR_2}, if there exists an encoding function*

$$f_n(W_1, W_2) = \mathbf{X},$$

and decoding functions

$$\hat{W}_1 = g_{1,n}(\mathbf{Y}_1),$$
$$\hat{W}_2 = g_{2,n}(\mathbf{Y}_2),$$

such that

$$\lim_{n\to\infty} \mathbb{P}[(W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)] = 0,$$

assuming that W_1 and W_2 are uniformly distributed over the respective message sets.

The capacity is defined as the closure over all achievable rate pairs. Note that one can easily add to the above definition a common message.

The capacity of a general broadcast channel is still an open problem. However, the capacity is known for some important special cases [42] such as the *degraded broadcast channel* which is of interest in this work.

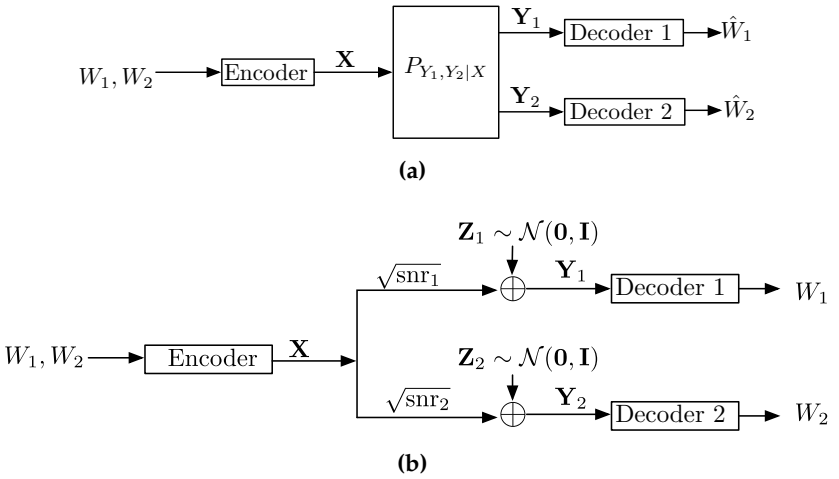


Figure 11. Two-receiver broadcast channel. (**a**) A general BC; (**b**) A Gaussian BC.

As told by Cover in [84] 1973–1974 was a year of "intense activity" where Bergmans, Gallager and others tried to provide a converse proof showing that the natural achievable region (shown in 1973 by Bergmans) is indeed the capacity region. Correspondences were exchanged between Gallager, Bergmans and Wyner until finally one day both Gallager and Bergmans sent a converse proof to Wyner. Gallager's proof tackled the degraded (i.e., $X \leftrightarrow Y_1 \leftrightarrow Y_2$) discrete memoryless BC yielding the following [85]:

$$\mathcal{C}_{BC} = \bigcup_{P_{UX}} \begin{cases} R_1 \leq I(X;Y_1|U) \\ R_2 \leq I(U;Y_2) \end{cases}, \tag{75}$$

where U is an auxiliary random variable with $U \leftrightarrow X \leftrightarrow (Y_1, Y_2)$. It did not consider a constraint on the input.

Bergman's proof directly examined the scalar Gaussian channel under a power constraint $\mathbb{E}[X^2] \leq 1$ and input-output relationship given by

$$Y_1 = \sqrt{\mathsf{snr}_1} X + Z_1, \tag{76}$$

$$Y_2 = \sqrt{\mathsf{snr}_2} X + Z_2, \tag{77}$$

where $\mathsf{snr}_1 \geq \mathsf{snr}_2$ (i.e., the degraded case) and applied the EPI (its first use since Shannon's paper in 1948) [86]:

$$\mathcal{C}_{BC} = \bigcup_{\alpha \in [0,1]} \begin{cases} R_1 \leq \frac{1}{2} \log\left(1 + \alpha \mathsf{snr}_1\right) \\ R_2 \leq \frac{1}{2} \log\left(\frac{1+\mathsf{snr}_2}{1+\alpha \mathsf{snr}_2}\right) \end{cases}. \tag{78}$$

6.1. Converse for the Gaussian Broadcast Channel

In [30] Guo et al. have shown that a converse proof of the scalar (degraded) Gaussian channel can also be derived using the SCPP bound instead of the EPI, when applied on the extension of Gallager's single-letter expression which takes into account also a power constraint.

The power constraint $\mathbb{E}[X^2] \leq 1$, implies that there exists some $\alpha \in [0,1]$ such that

$$I(X;Y_2|U) = \frac{1}{2}\log(1+\alpha \mathsf{snr}_2) = \frac{1}{2}\int_0^{\mathsf{snr}_2} \frac{\alpha}{1+\alpha t} dt. \tag{79}$$

By the chain rule of mutual information

$$\begin{aligned} I(U;Y_2) &= I(U,X;Y_2) - I(X;Y_2|U) \\ &= I(X;Y_2) - I(X;Y_2|U), \end{aligned} \tag{80}$$

where in the last step we have used the Markov chain relationship $U \leftrightarrow X \leftrightarrow (Y_1, Y_2)$. Using (79) and (80) the bound on R_2 is given by

$$\begin{aligned} R_2 &\leq I(U;Y_2) \\ &= \frac{1}{2}\int_0^{\mathsf{snr}_2} \mathrm{mmse}(X,t) dt + \int_0^{\mathsf{snr}_2} \frac{\alpha}{1+\alpha t} dt \\ &\leq \frac{1}{2}\int_0^{\mathsf{snr}_2} \frac{1}{1+t} dt + \int_0^{\mathsf{snr}_2} \frac{\alpha}{1+\alpha t} dt \qquad (81) \\ &= \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_2}{1+\alpha \mathsf{snr}_2}\right), \qquad (82) \end{aligned}$$

where in (81) we have used the LMMSE bound. The inequality in (82) establishes the desired bound on R_2. To bound the R_1 term observe that by using I-MMSE and (79)

$$I(X;Y_2|U) = \frac{1}{2}\int_0^{\mathsf{snr}_2} \mathrm{mmse}(X,t|U)dt = \frac{1}{2}\int_0^{\mathsf{snr}_2} \frac{\alpha}{1+\alpha t}dt, \tag{83}$$

the expression in (83) implies that there exists some $0 \leq \mathsf{snr}_0 \leq \mathsf{snr}_2$ such that

$$\mathrm{mmse}(X,\mathsf{snr}_0|U) = \frac{\alpha}{1+\alpha\mathsf{snr}_0}. \tag{84}$$

The equality in (84) together with the SCPP bound implies the following inequality:

$$\mathrm{mmse}(X,t|U) \leq \frac{\alpha}{1+\alpha t}, \tag{85}$$

for all $t \geq \mathsf{snr}_2 \geq \mathsf{snr}_0$. Therefore,

$$\begin{aligned} R_1 &\leq I(X;Y_1|U) \\ &= \frac{1}{2}\int_0^{\mathsf{snr}_2} \mathrm{mmse}(X,t|U)dt + \frac{1}{2}\int_{\mathsf{snr}_2}^{\mathsf{snr}_1} \mathrm{mmse}(X,t|U)dt \\ &= \frac{1}{2}\log(1+\alpha\mathsf{snr}_2) + \frac{1}{2}\int_{\mathsf{snr}_2}^{\mathsf{snr}_1} \mathrm{mmse}(X,t|U)dt \qquad (86) \\ &\leq \frac{1}{2}\log(1+\alpha\mathsf{snr}_2) + \frac{1}{2}\int_{\mathsf{snr}_2}^{\mathsf{snr}_1} \frac{\alpha}{1+\alpha t}dt \qquad (87) \\ &= \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_2}{1+\alpha\mathsf{snr}_2}\right), \end{aligned}$$

where the expression in (86) follows from (79) and the bound in (87) follows by using the bound in (85). This concludes the proof.

6.2. SNR Evolution of Optimal BC Codes

Similarly to the analysis presented in Section 4.2 the I-MMSE relationship can be used also to obtain practical insights and key properties of optimal code sequences for the scalar Gaussian BC. These were shown in [28,87].

The first result we present explains the implications of reliable decoding in terms of the MMSE behavior.

Theorem 12 ([28]). *Consider a code sequence, transmitting a message pair (W_1, W_2), at rates (R_1, R_2) (not necessarily on the boundary of the capacity region), over the Gaussian BC. W_2 can be reliably decoded from $\mathbf{Y}_2$ if and only if*

$$\mathrm{mmse}_\infty(\mathbf{X};\gamma|W_2) = \mathrm{mmse}_\infty(\mathbf{X};\gamma), \quad \forall \gamma \geq \mathsf{snr}_2. \tag{88}$$

The above theorem formally states a very obvious observation which is that once W_2 can be decoded, it provides no improvement to the estimation of the transmitted codeword, beyond the estimation from the output. This insight is strengthened as it is also a sufficient condition for reliable decoding of the message W_2.

The main observation is an extension of the result given in [63], where it was shown that a typical code from the hierarchical code ensemble (which achieves capacity) designed for a given Gaussian BC has a specific SNR-evolution of the MMSE function. This result was extended and shown to hold for any code sequence on the boundary of the capacity region.

Theorem 13 ([28]). *An achievable code sequence for the Gaussian BC has rates on the boundary of the capacity region, meaning*

$$(R_1, R_2) = \left(\frac{1}{2} \log \left(1 + \alpha \mathsf{snr}_1\right), \frac{1}{2} \log \left(\frac{1 + \mathsf{snr}_2}{1 + \alpha \mathsf{snr}_2} \right) \right), \tag{89}$$

for some $\alpha \in [0,1]$*, if and only if it has a deterministic mapping from* (W_1, W_2) *to the transmitted codeword and*

$$\mathrm{mmse}_\infty(\mathbf{X}; \gamma) = \begin{cases} \frac{1}{1+\gamma}, & \mathsf{snr} \in [0, \mathsf{snr}_2) \\ \frac{\alpha}{1+\alpha\gamma}, & \gamma \in [\mathsf{snr}_2, \mathsf{snr}_1) \\ 0, & \gamma \geq \mathsf{snr}_1 \end{cases}, \tag{90}$$

$$\mathrm{mmse}_\infty(\mathbf{X}; \gamma | W_2) = \begin{cases} \frac{\alpha}{1+\alpha\gamma}, & \mathsf{snr} \in [0, \mathsf{snr}_1) \\ 0, & \mathsf{snr} \geq \mathsf{snr}_1 \end{cases}. \tag{91}$$

Note that the above SNR-evolution holds for any capacity achieving code sequence for the Gaussian BC. This includes also codes designed for decoding schemes such as "dirty paper coding", in which case the decoding at $\mathbf{Y}_1$ does not require the reliable decoding of the known "interference" (the part of the codeword that carries the information of W_2), but simply encodes the desired messages against that "interference". In that sense the result is surprising since one does not expect such a scheme to have the same SNR-evolution as a superposition coding scheme, where the decoding is in layers: first the "interference" and only after its removal, the reliable decoding of the desired message.

Figure 12 depicts the result of Theorem 13 for capacity achieving code sequences.

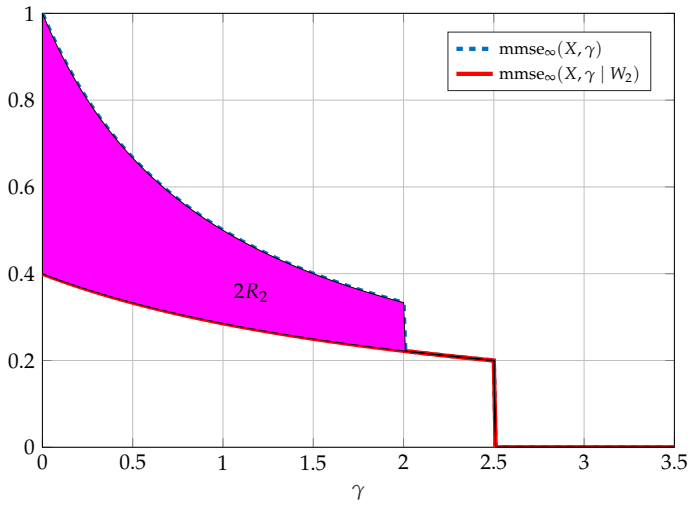


Figure 12. In the above figure we consider the SNR-evolution of $\mathrm{mmse}_\infty(\mathbf{X}; \gamma)$ (in dashed blue) and $\mathrm{mmse}_\infty(\mathbf{X}; \gamma | W_2)$ (in solid red) required from an asymptotically capacity achieving code sequence for the Gaussian BC (rate on the boundary of the capacity region). Twice R_2 is marked as the area between these two functions (in magenta). The parameters are $\mathsf{snr}_1 = 2.5$, $\mathsf{snr}_2 = 2$, and $\alpha = 0.4$.

7. Multi-Receiver SNR-Evolution

In this section we extend the results regarding the SNR-evolution of the Gaussian wiretap channel and the SNR-evolution of the Gaussian broadcast channel, given in Sections 4.2 and 6.2, respectively, to the multi-receiver setting. Moreover, we enhance the graphical interpretation of the SNR-evolution to relate to the basic relevant quantities of rate and equivocation.

More specifically, we now consider a multi-receiver additive Gaussian noise setting in which

$$\mathbf{Y}_i = \sqrt{\mathsf{snr}_i}\mathbf{X} + \mathbf{Z}_i, \tag{92}$$

where we assume that $\mathsf{snr}_1 \leq \mathsf{snr}_2 \leq \cdots \leq \mathsf{snr}_K$ for some $K \geq 2$. Since both rate and equivocation are measured according to the conditional densities at the receivers we may further assume that $\boldsymbol{Z} = \boldsymbol{Z}_i$ for all i. Moreover, $\mathbf{X}$ is the transmitted message encoded at the transmitter, assuming a set of some L messages $(W_1, W_2, \ldots, W_L)$. Each receiver may have a different set of requirements regarding these messages. Such requirements can include:

- Reliably decoding some subset of these messages;
- Begin ignorant to some extent regarding some subset of these messages, meaning having at least some level of equivocation regarding the messages within this subset;
- A receiver may be an "unintended" receiver with respect to some subset of messages, in which case we might wish also to limit the "disturbance" these message have at this specific receiver. We may do so by limiting the MMSE of these messages; and
- Some combination of the above requirements.

There might be, of course, additional requirements, but so far the application of the I-MMSE approach as done in [34,70,87,88], was able to analyze these types of requirements. We will now give the main results from which one can consider other specific cases as discussed at the end of this section.

We first consider only reliable communication, meaning a set of messages intended for receivers at different SNRs, in other words, a K-user Gaussian BC.

Theorem 14 ([88]). *Given a set of messages $(W_1, W_2, \ldots, W_K)$, such that W_i is reliably decoded at snr_i and $\mathsf{snr}_1 \leq \mathsf{snr}_2 \leq \cdots \leq \mathsf{snr}_K$, we have that*

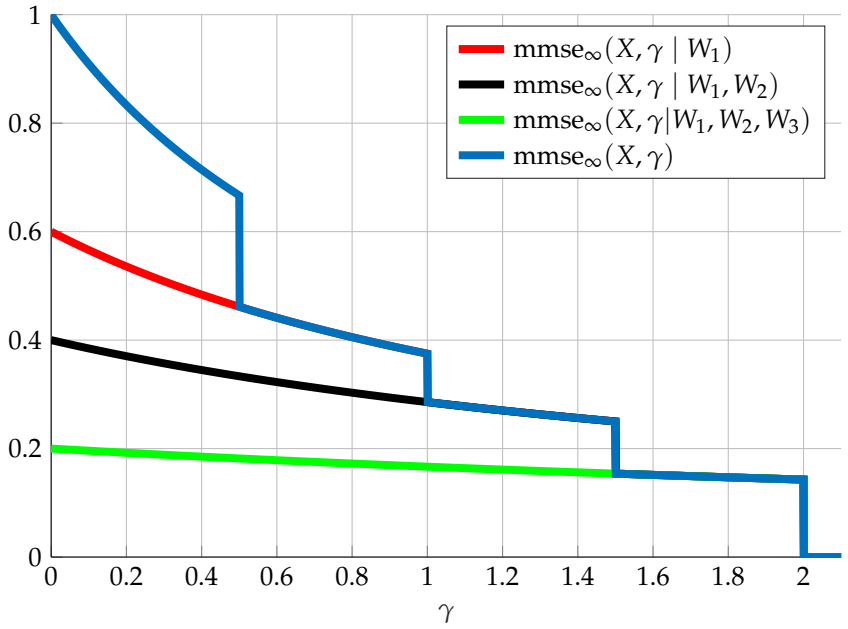
$$R_i = \frac{1}{2}\int_0^{\mathsf{snr}_i} \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_{i-1}) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_i)d\gamma. \tag{93}$$

In the case of R_1 the first MMSE is simply $\mathrm{mmse}_\infty(\mathbf{X};\gamma)$ (meaning $W_0 = \emptyset$).

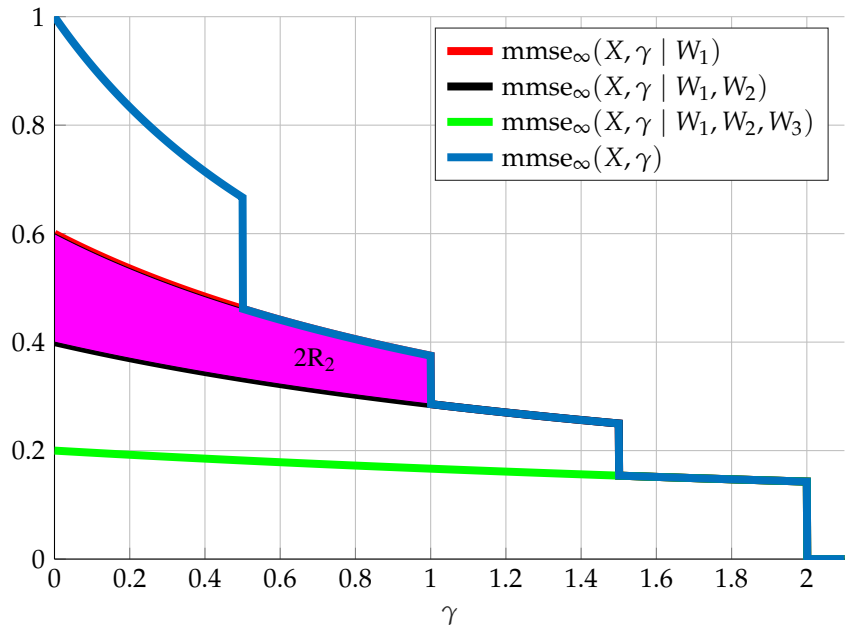
Note that due to the basic ordering of the MMSE quantity, meaning that for all $\gamma \geq 0$

$$\mathrm{mmse}_\infty(\mathbf{X};\gamma) \geq \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1) \geq \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2) \geq \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,W_3) \geq \ldots,$$

we have that the integrand is always non-negative. Thus, the above result slices the region defined by $\mathrm{mmse}_\infty(\mathbf{X};\gamma)$ into distinctive stripes defined by the conditional MMSE functions. Each such stripe corresponds to twice the respective rate. The order of the stripes from top to bottom is by the message first decoded to the one last decoded (see Figure 13); further, taking into account Theorem 12, which gives a necessary and sufficient condition for reliable communication in terms of MMSE functions, we know that for $\mathsf{snr} \geq \mathsf{snr}_i$ the MMSE conditioned on any message reliably decoded at snr_i equals $\mathrm{mmse}_\infty(\mathbf{X};\gamma)$; thus, we may extend the integration in the above result to any $\mathsf{snr} \geq \mathsf{snr}_i$ (or even integrate to infinity).



(a)



(b)

Figure 13. The above figure depicts a general transmission of (W_1, W_2, W_3) independent messages, each required to be reliably decoded at the respective SNR $(\mathsf{snr}_1, \mathsf{snr}_2, \mathsf{snr}_3) = (1/2, 1, 3/2)$. The rates are defined by the areas. (**a**) We observe that due to reliable decoding, the respective conditional MMSE converges to the MMSE; (**b**) we examine the same transmission as in (**a**), however here we observe the respective rates. The rates are defined by the areas. As an example we mark $2R_2$ - twice the rate of message W_2. Similarly one can mark the other rates $2R_1$ and $2R_3$.

We now consider in addition to reliable communication also the equivocation measure.

Theorem 15 ([88]). *Assume a set of independent messages* $(W_1, W_2, \ldots, W_i)$ *such that* $(W_1, W_2, \ldots, W_{i-1})$ *are reliably decoded at* $\mathbf{Y}(\mathsf{snr}_{i-1})$*, however* W_i *is reliably decoded only at some* $\mathsf{snr}_i > \mathsf{snr}_{i-1}$*. The equivocation of* W_i *at* $\mathbf{Y}(\mathsf{snr}_{i-1})$ *equals*

$$H(W_i|\mathbf{Y}(\mathsf{snr}_{i-1})) = \frac{1}{2}\int_{\mathsf{snr}_{i-1}}^{\mathsf{snr}_i} \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_{i-1}) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,\ldots,W_i)d\gamma, \tag{94}$$

which can also be written as

$$H(W_i|\mathbf{Y}(\mathsf{snr}_{i-1})) = \frac{1}{2}\int_{\mathsf{snr}_{i-1}}^{\mathsf{snr}_i} \mathrm{mmse}_\infty(\mathbf{X};\gamma) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_i)d\gamma. \tag{95}$$

The above result together with Theorem 14 provides a novel graphical interpretation. Theorem 14 divides the area below $\mathrm{mmse}_\infty(\mathbf{X};\gamma)$ into stripes, each corresponding to a rate. Theorem 15 further divides these stripes horizontally. The stripe corresponding to the rate of message W_i is an area between $\mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,\ldots,W_{i-1})$ and $\mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,\ldots,W_i)$ from $[0,\mathsf{snr}_i]$. For any point $\mathsf{snr} > \mathsf{snr}_i$ this area is then split into the region between $[0,\mathsf{snr}]$ which corresponds to the information that can be obtained regarding the message by $\mathbf{Y}(\mathsf{snr})$ and the region $[\mathsf{snr},\mathsf{snr}_i]$ which corresponds to the equivocation (see Figure 14 for an example).

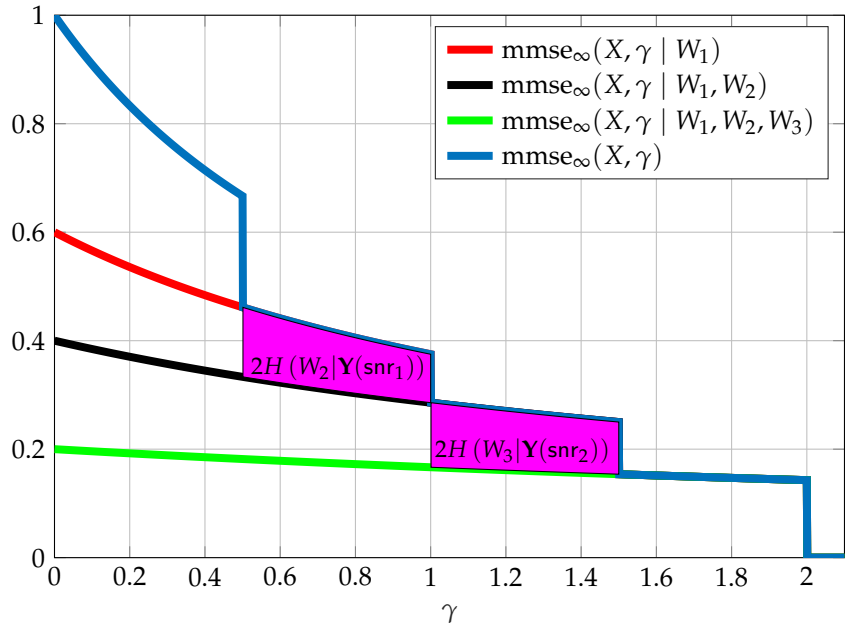


Figure 14. The above figure depicts a general transmission of independent messages (W_1, W_2, W_3), each required to be reliably decoded at the respective SNR $(\mathsf{snr}_1, \mathsf{snr}_2, \mathsf{snr}_3) = (1/2, 1, 3/2)$. Here we denote two equivocation measures $2H(W_2|\mathbf{Y}(\mathsf{snr}_1))$ and $2H(W_3|\mathbf{Y}(\mathsf{snr}_2))$ according to Theorem 15.

Let us now assume complete secrecy, meaning

$$H(W_i|\mathbf{Y}(\mathsf{snr}_{i-1})) = H(W_i). \tag{96}$$

Using Theorems 14 and 15 we have that

$$\begin{aligned}&\int_0^{\mathsf{snr}_i} \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_{i-1}) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_i)d\gamma = \\ &\int_{\mathsf{snr}_{i-1}}^{\mathsf{snr}_i} \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_{i-1}) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_i)d\gamma,\end{aligned} \tag{97}$$

assuming we have reliable decoding of messages $(W_1, W_2, \ldots, W_{i-1})$ at snr_{i-1}. This reduces to

$$\int_0^{\mathsf{snr}_{i-1}} \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_{i-1}) - \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,\ldots,W_i)d\gamma = 0, \tag{98}$$

which due to the non-negativity of the integrand results in

$$\mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,\ldots,W_{i-1}) = \mathrm{mmse}_\infty(\mathbf{X};\gamma|W_1,W_2,\ldots,W_i), \tag{99}$$

for all $\gamma \in [0, \mathsf{snr}_{i-1})$. This is exactly the condition for complete secrecy given in [34]. The important observation here is that to obtain complete secrecy we require that the stripe of the secure message is reduced to the section $[\mathsf{snr}_{i-1}, \mathsf{snr}_i]$, where the eavesdropper is at snr_{i-1} and the legitimate receiver is at snr_i. This reduction in the stripe of the secure message can be interpreted as having been used for the transmission of the camouflaging information required for complying with the secrecy constraint.

The above approach can be further extended and can provide a graphical interpretation for more elaborate settings with additional requirements at the receiver. An immediate such example would be adding "disturbance" constraints in terms of MMSEs. Another extension which has been also considered in [88] is the problem of "secrecy outside the bounded range" [89]. For this setting complete secrecy rate can be enhanced by using the inherent randomness in the message which results from the fact that it contains an additional "unintended" message which is not necessarily reliably decoded. For more details on this problem and its graphical interpretation the reader is referred to [88,89].

8. Interference Channels

A two user interference channel (IC), introduced by Ahlswede in [77], depicted in Figure 15, is a system consisting of two transmitters and two receivers. The goal of a transmitter $i \in [1:2]$ is to reliably transmit the message W_i to receiver i. Transmitter i encodes a message W_i into a transmitted codeword $\mathbf{X}_i$ of length n. Receiver i receives the sequence $\mathbf{Y}_i$ of length n and tries to decode the message W_i from the observed sequence $\mathbf{Y}_i$. An achievable rate pair for the IC is defined as follows:

Definition 10. *A rate pair (R_1, R_2) is said to be achievable, if for a message W_1 of cardinality 2^{nR_1} and a message W_2 of cardinality 2^{nR_2} there exists a sequence of encoding functions*

$$f_{n,1}(W_1) = \mathbf{X}_1,$$
$$f_{n,2}(W_2) = \mathbf{X}_2,$$

and decoding functions

$$\hat{W}_1 = g_{1,n}(\mathbf{Y}_1),$$
$$\hat{W}_2 = g_{2,n}(\mathbf{Y}_2),$$

such that

$$\lim_{n\to\infty} \mathbb{P}[(W_1, W_2) \neq (\hat{W}_1, \hat{W}_2)] = 0, \tag{100}$$

assuming that W_1 and W_2 are uniformly distributed over their respective message sets.

The capacity region is defined as the closure over all achievable rate pairs. In [77] Ahlswede demonstrated a multi-letter capacity expression given in (49). Unfortunately, the capacity expression in (49) is considered "uncomputable" in the sense that we do not know how to explicitly characterize the input distributions that attain its convex closure. Moreover, it is not clear whether there exists an equivalent single-letter form for (49) in general.

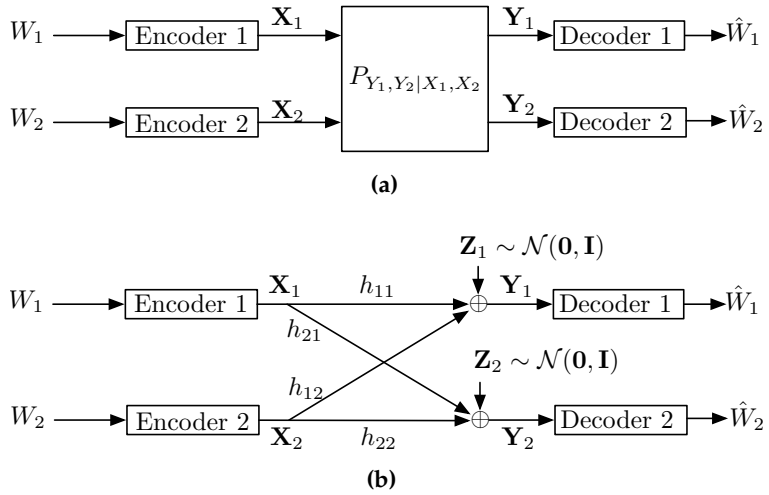


Figure 15. Two user interference channels. (**a**) A general interference channel; (**b**) The Gaussian interference channel.

Because of "uncomputability" the capacity expression in (49) has received little attention, except for the following: in [79] the limiting expression was used to show that limiting to jointly Gaussian distributions is suboptimal; in [72] the limiting expression was used to derive the sum-capacity in the very weak interference regime; and in [90] it was shown that in the high-power regime the limiting expression normalized by the point-to-point capacity (i.e., the degrees of freedom (DoF)) can be single letterized.

Instead, the field has focussed on finding alternative ways to characterize single-letter inner and outer bounds. The best known inner bound is the HK achievable scheme [74], which is:

- capacity achieving in the strong interference regime [75,91,92];
- capacity achieving for a class of injective deterministic channels [93,94];
- approximately capacity achieving for a class of injective semi-deterministic channels [95]; and
- approximately capacity achieving (within 1/2 bit) for a class of Gaussian noise channels (which is a special case of the injective semi-deterministic channel) [76].

It is important to point out that in [96] the HK scheme was shown to be strictly sub-optimal for a class of DMC's. Moreover, the result in [96] suggests that multi-letter achievable strategies might be needed to achieve capacity of the IC.

8.1. Gaussian Interference Channel

In this section we consider the practically relevant scalar G-IC channel, depicted in Figure 15b, with input-output relationship

$$Y_1 = h_{11}X_1 + h_{12}X_2 + Z_1, \tag{101a}$$
$$Y_2 = h_{21}X_1 + h_{22}X_2 + Z_2, \tag{101b}$$

where Z_i is i.i.d. zero-mean unit-variance Gaussian noise. For the G-IC in (101), the maximization in (49) is further restricted to inputs satisfying the power constraint $\mathbb{E}[X_i^2] \leq 1$, $i \in [1:2]$.

For simplicity we will focus primarily on the *symmetric* G-IC defined by

$$|h_{11}|^2 = |h_{22}|^2 = \mathsf{snr} \geq 0, \tag{102a}$$
$$|h_{12}|^2 = |h_{21}|^2 = \mathsf{inr} \geq 0, \tag{102b}$$

and we will discuss how the results for the symmetric G-IC extend to the general asymmetric setting.

In general, little is known about the optimizing input distribution in (49) for the G-IC and only some special cases have been solved. In [71–73] it was shown that i.i.d. Gaussian inputs maximize the sum-capacity in (49) for $\sqrt{\frac{\mathsf{inr}}{\mathsf{snr}}}(1+\mathsf{inr}) \leq \frac{1}{2}$ in the symmetric case. In contrast, the authors of [79] showed that in general multivariate Gaussian inputs do not exhaust regions of the form in (49). The difficulty arises from the competitive nature of the problem [43]: for example, say X_2 is i.i.d. Gaussian; taking X_1 to be Gaussian increases $I(X_1;Y_1)$ but simultaneously decreases $I(X_2;Y_2)$, as Gaussians are known to be the "best inputs" for Gaussian point-to-point power-constrained channels, but are also the "worst noise" (or interference, if it is treated as noise) for a Gaussian input.

So, instead of pursuing exact results, the community has recently focussed on giving performance guarantees on approximations of the capacity region [97]. In [76] the authors showed that the HK scheme with Gaussian inputs and without time-sharing is optimal to within 1/2 bit, irrespective of the channel parameters.

8.2. Generalized Degrees of Freedom

The constant gap result of [76] provides an exact characterization of the generalized degrees of freedom (gDoF) region defined as

$$\mathcal{D}(\alpha) := \left\{(d_1,d_2) : d_i := \lim_{\mathsf{snr}\to\infty} \frac{R_i(\mathsf{snr},\mathsf{inr}=\mathsf{snr}^\alpha)}{\frac{1}{2}\log(1+\mathsf{snr})}, i \in [1:2], (R_1,R_2) \text{ is achievable}\right\}, \tag{103}$$

and where $\mathcal{D}(\alpha)$ was shown to be

$$\mathcal{D}(\alpha) = \left\{ (d_1,d_2) : \begin{array}{ll} d_1 & \leq 1 \\ d_2 & \leq 1 \\ d_1+d_2 & \leq [1-\alpha]^+ + \max(\alpha,1) \\ d_1+d_2 & \leq \max(1-\alpha,\alpha) + \max(1-\alpha,\alpha) \\ 2d_1+d_2 & \leq [1-\alpha]^+ + \max(1,\alpha) + \max(1-\alpha,\alpha) \\ d_1+2d_2 & \leq [1-\alpha]^+ + \max(1,\alpha) + \max(1-\alpha,\alpha) \end{array} \right\}. \tag{104a}$$

The region in (104) is achieved by the HK scheme without time sharing; for the details see [42,76].

The α parameter is the strength of the interference in dB. The gDoF is an important metric that sheds light on the optimal coding strategies in the high SNR regime. The gDoF metric deemphasizes the role of noise in the network and only focuses on the role of signal interactions. Often these strategies can be translated to the medium and low SNR regions. The gDoF is especially useful in analyzing *interference alignment* strategies [98,99] where proper design of the signaling scheme can ensure very high rates. The notion of gDoF has received considerable attention in information theoretic literature and the interested reader is referred to [100] and reference therein.

For our purposes, we will only look at the sum-gDoF of the interference channel given by

$$d_\Sigma(\alpha) = \max_{(d_1,d_2)\in\mathcal{D}(\alpha)} d_1+d_2 = 2\min\left(1, \max\left(\frac{\alpha}{2}, 1-\frac{\alpha}{2}\right), \max(\alpha,1-\alpha)\right). \tag{105}$$

The sum-gDoF in (105) as a function of the parameter α is plotted in Figure 16. The curve in Figure 16 is often called the *W-curve*.

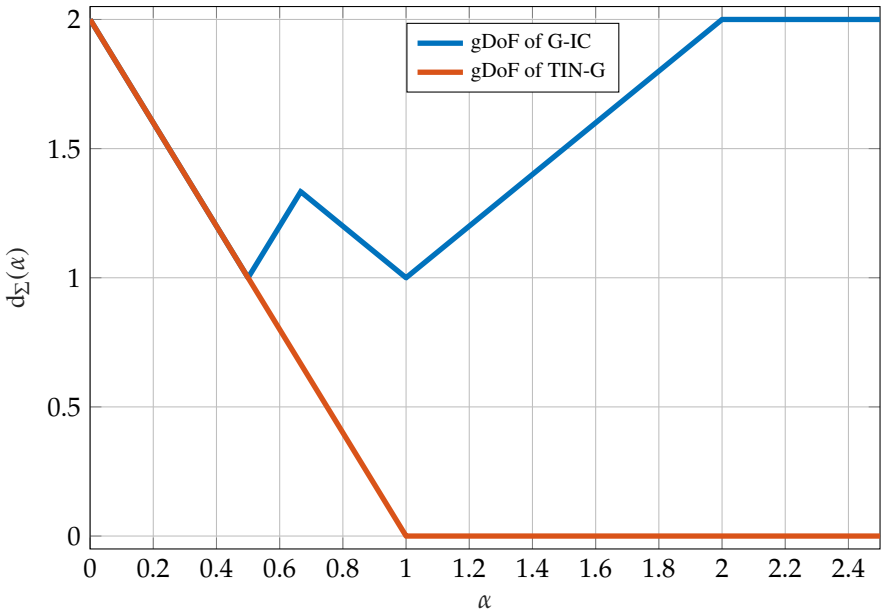


Figure 16. gDoF of the G-IC.

Depending on the parameters $(\mathsf{snr}, \mathsf{inr} = \mathsf{snr}^\alpha)$ we identify the following operational regimes: very strong interference, $\alpha \geq 2$; strong interference, $1 \leq \alpha < 2$; weak interference type I, $2/3 \leq \alpha < 1$; weak interference type II, $1/2 \leq \alpha < 2/3$; very weak interference, $0 \leq \alpha < 1/2$.

8.3. Treating Interference as Noise

An inner bound on the capacity region in (49) can be obtained by considering i.i.d. inputs in (49) thus giving

$$\mathcal{R}_{\text{in}}^{\text{TIN+TS}} = \text{co}\left(\bigcup_{P_{X_1X_2}=P_{X_1}P_{X_2}} \left\{ \begin{array}{l} 0 \leq R_1 \leq I(X_1;Y_1) \\ 0 \leq R_2 \leq I(X_2;Y_2) \end{array} \right\} \right), \tag{106}$$

where the superscript "TIN+TS" reminds the reader that the region is achieved by treating interference and noise and with time sharing (TS), where TS is enabled by the convex hull operation [42]. By further removing the convex hull operation in (106) we arrive at

$$\mathcal{R}_{\text{in}}^{\text{TINnoTS}} = \bigcup_{P_{X_1X_2}=P_{X_1}P_{X_2}} \left\{ \begin{array}{l} 0 \leq R_1 \leq I(X_1;Y_1) \\ 0 \leq R_2 \leq I(X_2;Y_2) \end{array} \right\}. \tag{107}$$

The region in (107) does not allow the users to time-share.

Obviously

$$\mathcal{R}_{\text{in}}^{\text{TINnoTS}} \subseteq \mathcal{R}_{\text{in}}^{\text{TIN+TS}} \subseteq \mathcal{C}.$$

The question of interest in this section is how $\mathcal{R}_{\text{in}}^{\text{TINnoTS}}$ fares compared to $\mathcal{C}$. Note that there are many advantages in using TINnoTS in practice. For example, TINnoTS does not require codeword synchronization, as for example for joint decoding or interference cancellation, and does not require much coordination between users, thereby reducing communications overhead. Therefore, an interesting question that arises is: What are the limits of the TIN region?

By evaluating the TIN region with Gaussian inputs we get an achievable sum-gDoF of

$$d_{\Sigma}^{\text{TIN-G}}(\alpha) = 2\max(0, 1-\alpha), \tag{108}$$

shown by a red curve in Figure 16. Clearly, using Gaussian inputs in the TIN region is gDoF optimal in the very weak interference regime and is otherwise strictly suboptimal. Because Gaussian inputs are often mutual information maximizers, one might think that the expression in (108) is the best that we can hope for. However, this intuition can be very misleading, and despite the simplicity of TIN, in [45] TINnoTS was shown to achieve capacity $\mathcal{C}$ within a gap which also implies that TIN is gDoF optimal. The key observation is to use non-Gaussian inputs, specifically the mixed inputs presented in Section 5.5.

Theorem 16 ([45]). *For the G-IC, as defined in* (102)*, the TINnoTS achievable region in* (107) *is optimal to within a constant gap, or a gap of order* $O(\log\log(\min(\mathsf{snr}, \mathsf{inr})))$*, and it is therefore gDoF optimal.*

Next, we demonstrated the main ideas behind Theorem 16. The key to this analysis is to use mixed inputs, presented in Section 5.5, and given by

$$X_i = \sqrt{1-\delta_i}\, X_{iD} + \sqrt{\delta_i}\, X_{iG},\ i \in [1:2]: \tag{109a}$$

$$X_{iD} \sim \mathsf{PAM}\left(N_i, \sqrt{\frac{12}{N_i^2-1}}\right), \tag{109b}$$

$$X_{iG} \sim \mathcal{N}(0,1), \tag{109c}$$

where the random variables X_{ij} are independent for $i \in [1:2]$ and $j \in \{D, G\}$. Inputs in (109) have four parameters, namely the number of points $N_i \in \mathbb{N}$ and the power split $\delta_i \in [0,1]$, for $i \in [1:2]$, which must be chosen carefully in order to match a given outer bound.

By evaluating the TIN region in (107) with mixed inputs in (109) we arrive at the following achievable region.

Proposition 8. *(TIN with Mixed Inputs [45]) For the G-IC the TINnoTS region in* (107) *contains the region* $\mathcal{R}_{in}$ *defined as*

$$\mathcal{R}_{in} := \bigcup \left\{ \begin{array}{l} 0 \le R_1 \le I\left(S_1, S_1 + Z_1\right) + \frac{1}{2}\log\left(1 + \frac{|h_{11}|^2\delta_1}{1+|h_{12}|^2\delta_2}\right) - \min\left(\log(N_2), \frac{1}{2}\log\left(1 + \frac{|h_{12}|^2(1-\delta_2)}{1+|h_{12}|^2\delta_2}\right)\right) \\ 0 \le R_2 \le I\left(S_2; S_2 + Z_2\right) + \frac{1}{2}\log\left(1 + \frac{|h_{22}|^2\delta_2}{1+|h_{21}|^2\delta_1}\right) - \min\left(\log(N_1), \frac{1}{2}\log\left(1 + \frac{|h_{21}|^2(1-\delta_1)}{1+|h_{21}|^2\delta_1}\right)\right) \end{array} \right\}, \tag{110}$$

where the union is over all possible parameters $[N_1, N_2, \delta_1, \delta_2] \in \mathbb{N}^2 \times [0,1]^2$ *for the mixed inputs in* (109) *and where the equivalent discrete constellations seen at the receivers are*

$$S_1 := \frac{\sqrt{1-\delta_1}h_{11}X_{1D} + \sqrt{1-\delta_2}h_{12}X_{2D}}{\sqrt{1+|h_{11}|^2\delta_1 + |h_{12}|^2\delta_2}}, \tag{111a}$$

$$S_2 := \frac{\sqrt{1-\delta_1}h_{21}X_{1D} + \sqrt{1-\delta_2}h_{22}X_{2D}}{\sqrt{1+|h_{21}|^2\delta_1 + |h_{22}|^2\delta_2}}. \tag{111b}$$

Next, we select the parameters $[N_1, N_2, \delta_1, \delta_2]$ to optimize the region in (110). For simplicity, we focus only on the very strong interference regime ($\alpha \ge 2$). The gDoF optimality of TIN in the very strong interference regime is perhaps the most surprising. The capacity in this regime has been shown by Carleial in [91] who demonstrated that capacity can be achieved with a successive cancellation decoding strategy where the interference is decoded before the desired signal. Unlike the Carleial

scheme TIN only use a point-to-point decoder for non-Gaussian noise and can be classified as a soft-interference-decoding strategy discussed in Section 5.1.

In the very strong interference ($\alpha \geq 2$) regime the sum-gDoF is given by

$$d_{\Sigma}(\alpha) = 2. \tag{112}$$

To show that TIN can achieve the gDoF (112), let the parameters in (110) be given by $N = N_1 = N_2 = \lfloor \sqrt{1+\mathsf{snr}} \rfloor$ and $\delta_1 = \delta_2 = 0$. It is not difficult to see that with this choice of inputs the rate in (110) is given by

$$\begin{aligned} R_i &= I\left(S_i, S_i + Z_i\right) - \min\left(\log(N), \frac{1}{2}\log\left(1+\mathsf{inr}\right)\right) \\ &\geq I\left(S_i, S_i + Z_i\right) - \log(N). \end{aligned}$$

Therefore, the key now is to lower bound $I\left(S_i, S_i + Z_i\right)$. This is done by using the Ozarow-Wyner bound in (35b).

Lemma 2. *Let $N = N_1 = N_2 = \lfloor \sqrt{1+\mathsf{snr}} \rfloor$ and $\delta_1 = \delta_2 = 0$. Then,*

$$I\left(S_i, S_i + Z_i\right) \geq 2\log(N) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{3}\right). \tag{113}$$

Proof. Using the Ozarow-Wyner bound in (35b)

$$\begin{aligned} I\left(S_i, S_i + Z_i\right) &\geq H(S_i) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) - \frac{1}{2}\log\left(1 + \frac{12\mathrm{mmse}(S_i,1)}{d_{\min}(S_i)^2}\right) \\ &\overset{a)}{\geq} H(S_i) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) - \frac{1}{2}\log\left(1 + \frac{12\mathrm{mmse}(S_i,1)}{\mathsf{snr} d_{\min}(X_i)^2}\right) \\ &\overset{b)}{\geq} H(S_i) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) - \frac{1}{2}\log\left(1 + \frac{12(\mathsf{snr}+\mathsf{inr})}{(1+\mathsf{snr}+\mathsf{inr})\mathsf{snr} d_{\min}(X_i)^2}\right) \\ &\overset{c)}{\geq} H(S_i) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) - \frac{1}{2}\log\left(1 + \frac{(\mathsf{snr}+\mathsf{inr})}{(1+\mathsf{snr}+\mathsf{inr})}\right) \\ &\overset{d)}{\geq} H(S_i) - \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{3}\right), \end{aligned}$$

where the (in)-equalities follow from: (a) using the bound $d_{\min}(S_i) \geq \min(\sqrt{\mathsf{snr}} d_{\min}(X_1), \sqrt{\mathsf{inr}} d_{\min}(X_2))$; (b) using the LMMSE bound in (11); (c) using the bound $d_{\min}(X_i)^2 = \frac{12}{N_i^2 - 1} \geq \frac{12}{\mathsf{snr}}$; and (d) using the bound $\frac{(\mathsf{snr}+\mathsf{inr})}{(1+\mathsf{snr}+\mathsf{inr})} \leq 1$.

The proof is concluded by observing that in the very strong interference regime with the choice $N = N_1 = N_2 = \lfloor \sqrt{1+\mathsf{snr}} \rfloor$, the entropy of a sum-set is given by

$$H(S_i) = H(X_1) + H(X_2) = 2\log(N).$$

□

Therefore, the sum-gDoF of TIN is given by

$$\begin{aligned} \lim_{\mathsf{snr}\to\infty} \frac{R_1 + R_2}{\frac{1}{2}\log(1+\mathsf{snr})} &\geq \lim_{\mathsf{snr}\to\infty} \frac{2\log(N) - \log\left(\frac{\pi \mathrm{e}}{3}\right)}{\frac{1}{2}\log(1+\mathsf{snr})} \\ &= \lim_{\mathsf{snr}\to\infty} \frac{2\log(\lfloor \sqrt{1+\mathsf{snr}} \rfloor) - \log\left(\frac{\pi \mathrm{e}}{3}\right)}{\frac{1}{2}\log(1+\mathsf{snr})} = 2. \end{aligned}$$

This concludes the proof of the achievability for the very strong interference regime.

Using the same ideas as in the proof for the very strong interference regime one can extend the optimality of TIN to other regimes.

9. Concluding Remarks and Future Directions

This section concludes this work by summarizing some interesting future directions.

One of the intriguing extensions of the I-MMSE relationship is the gradient formula, obtained by Palomar and Verdú [9]:

$$\nabla_{\mathbf{H}} I(\mathbf{X};\mathbf{HX}+\mathbf{Z}) = \mathbf{H}\mathbb{E}\left[(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid \mathbf{HX}+\mathbf{Z}])(\mathbf{X}-\mathbb{E}[\mathbf{X}\mid \mathbf{HX}+\mathbf{Z}])^{\mathsf{T}}\right]. \tag{114}$$

The expression in (114) has been used to study MIMO wiretap channels [101], extensions of Costa's EPI [102], and the design of optimal precoders for MIMO Gaussian channels [103].

However, much work is needed to attain the same level of maturity for the gradient expression in (114) as for the original I-MMSE results [8]. For example, it is not clear what is the correct extension (or if it exists) of a matrix version of the SCPP in Proposition 4. A matrix version of the SCPP could facilitate a new proof of the converse of the MIMO BC by following the steps of Section 6.1. The reader is referred to [33] where the SCPP type bounds have been extended to several classes of MIMO Gaussian channels.

Estimation theoretic principles have been instrumental in finding a general formula for the DoF for a static scalar K-user Gaussian interference channel [90], based on notions of the information dimension [104] and the MMSE dimension [32]. While the DoF is an important measure of the network performance it would be interesting to see if the approach of [90] could be used to analyze the more robust gDoF measure. Undoubtedly, such an extension will rely on the interplay between estimation and information measures.

"Information bottleneck" type problems [105] are defined as

$$\min_{Y} I(X;Y,Z) \tag{115a}$$

$$\text{s.t. } I(X;Y) = C, \tag{115b}$$

where $X \sim \mathcal{N}(0,1)$, $Z = \sqrt{\mathsf{snr}}X + N$ with $N \sim \mathcal{N}(0,1)$ independent of X and Y. A very elegant solution to (115) can be found by using the I-MMSE, the SCPP, and the argument used in the proof of the converse for the Gaussian BC in Section 6.1. It would be interesting to explore whether other variants of the bottleneck problem can be solved via estimation theoretic tools. For example, it would be interesting to consider

$$\max_{X,Z} I(X;Z) \tag{116a}$$

$$\text{s.t. } I(Y;Z) \leq C, \tag{116b}$$

where $X \leftrightarrow Y \leftrightarrow Z$ and $Y = \sqrt{\mathsf{snr}}X + N$ where $N \sim \mathcal{N}(0,1)$ and independent of X.

The extremal entropy inequality of [106], inspired by the channel enhancement method [107], was instrumental in showing several information theoretic converses in problems such as the MIMO wiretap channel [108], two-user Gaussian interference channel [71–73], and cognitive interference channel [109] to name a few. In view of the successful applications of the I-MMSE relationship to prove EPI type inequalities (e.g., [35,36,38,102]), it would be interesting to see if the extremal inequality presented in [106] can be shown via estimation theoretic arguments. Existence of such a method can reveal a way of deriving a large class of extremal inequalities potentially useful for information theoretic converses.

The extension of the I-MMSE results to cases that allow snr dependency of the input signal have been derived and shown to be useful in [10]. An interesting future direction is to consider the MMPE

while allowing snr dependency of the input signal; such a generalization has the potential of being useful when studying feedback systems as did the generalization of the MMSE in [10].

Another interesting direction is to study sum-rates of arbitrary networks with the use of the Ozarow-Wyner bound in Theorem 4. Note that, the Ozarow-Wyner bound holds for an arbitrary transition probability and the rate of an arbitrary network with n independent inputs and outputs can be lower bounded as

$$\sum_i R_i = I(X_1, ..., X_n; Y_1, ..., Y_n) \geq \sum_i H(X_i) - \mathsf{gap}, \tag{117}$$

where the gap term is explicitly given in Theorem 4 and is a function of the network transition probability.

Acknowledgments: The work of Alex Dytso and H. Vincent Poor was partially supported by the NSF under Grants CNS-1456793 and CCF-1420575. The work of Ronit Bustin and Shlomo Shamai was supported by the European Union's Horizon 2020 Research And Innovation Programme, grant agreement No. 694630. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

Author Contributions: All authors contributed equally to this work.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Proposition 3

Wu et al. [110] have shown precisely and rigorously, using basic inequalities, that the I-MMSE relationship holds for any input of finite variance. The approach in [110] followed by examining the truncated input. Although this approach extends directly for any finite n it is not trivial to extend it to the limit as $n \to \infty$. Note that the truncation argument is used only for the lower bound, whereas the upper bound is obtained directly and can be extended to the limit as shown in the sequel. Thus, our approach is to show this extension indirectly, relying on the existence of the I-MMSE relationship for any n.

Proof. We begin with an upper bound on the quantity

$$\frac{1}{n} I\left(\boldsymbol{X}_{n,\delta}; \sqrt{\delta}\boldsymbol{X}_{n,\delta} + \boldsymbol{N}_n | \boldsymbol{V}_{n,\delta} = \boldsymbol{v}_n\right), \tag{A1}$$

where $\{\boldsymbol{X}_{n,\delta}, \boldsymbol{V}_{n,\delta}\}_{\delta>0}$ is a collection of jointly distributed random vectors and $\boldsymbol{N}_n \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$ is independent of the pair $\{\boldsymbol{X}_{n,\delta}, \boldsymbol{V}_{n,\delta}\}_{\delta>0}$.

For fixed $\delta \in (0,1)$, we have

$$\begin{aligned}
\frac{1}{n} I\left(\boldsymbol{X}_{n,\delta}; \sqrt{\delta}\boldsymbol{X}_{n,\delta} + \boldsymbol{N}_n | \boldsymbol{V}_{n,\delta} = \boldsymbol{v}_n\right) &\leq \frac{1}{n}\frac{1}{2} \log |\mathbf{I}_n + \delta \mathbf{K}_{\delta,\boldsymbol{v}_n}| \\
&= \frac{1}{n}\frac{1}{2} \log \prod_{i=1}^{n} (1 + \delta \lambda_i(\mathbf{K}_{\delta,\boldsymbol{v}_n})) \\
&= \frac{1}{2}\frac{1}{n} \sum_{i=1}^{n} \log (1 + \delta \lambda_i(\mathbf{K}_{\delta,\boldsymbol{v}_n})) \\
&\leq \frac{1}{2} \log \left(1 + \delta \frac{1}{n} \sum_{i=1}^{n} \lambda_i(\mathbf{K}_{\delta,\boldsymbol{v}_n})\right) \\
&\leq \frac{\delta}{2}\frac{1}{n} \sum_{i=1}^{n} \lambda_i(\mathbf{K}_{\delta,\boldsymbol{v}_n}) \\
&= \frac{\delta}{2}\frac{1}{n} \mathrm{Tr}\left(\mathbf{K}_{\delta,\boldsymbol{v}_n}\right),
\end{aligned} \tag{A2}$$

where in the first inequality we have used the fact that the Gaussian distribution maximizes the entropy and where we denote the conditional covariance matrix of $\boldsymbol{X}_{n,\delta}$ given $\boldsymbol{v}_n$ as follows:

$$\mathbf{K}_{\delta,\boldsymbol{v}_n} = \mathbb{E}\left[\left(\boldsymbol{X}_{n,\delta} - \mathbb{E}\left[\boldsymbol{X}_{n,\delta}|\boldsymbol{V}_{n,\delta} = \boldsymbol{v}_n\right]\right)\left(\boldsymbol{X}_{n,\delta} - \mathbb{E}\left[\boldsymbol{X}_{n,\delta}|\boldsymbol{V}_{n,\delta} = \boldsymbol{v}_n\right]\right)^{\mathsf{T}}\right]. \tag{A3}$$

The second inequality uses Jensen's inequality and the last inequality uses $\log(1+x) \leq x$. Thus, we have that

$$\frac{1}{n} I\left(\boldsymbol{X}_{n,\delta}; \sqrt{\delta}\boldsymbol{X}_{n,\delta} + \boldsymbol{N}_n | \boldsymbol{V}_{n,\delta} = \boldsymbol{v}_n\right) \leq \frac{\delta}{2}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\delta,\boldsymbol{v}_n}\right). \tag{A4}$$

Note that we may take the expectation with respect to $\boldsymbol{V}_n$ of both sides of the inequality at any point, either by applying Jensen's inequality, or simply when the right-hand-side is linear in $K_{\delta,\boldsymbol{v}_n}$. We get that

$$\frac{1}{n} I\left(\boldsymbol{X}_{n,\delta}; \sqrt{\delta}\boldsymbol{X}_{n,\delta} + \boldsymbol{N}_n | \boldsymbol{V}_{n,\delta}\right) \leq \frac{\delta}{2}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\delta}\right), \tag{A5}$$

which holds for all n. Our assumption is that the limit of the normalized conditional mutual information quantity exists; however, we have no such assumption over the normalized MMSE quantity on the right-hand-side of the above. Thus, we take the lim inf of both sides of the above inequality, to obtain

$$I_\infty\left(\boldsymbol{X}_{\delta}; \sqrt{\delta}\boldsymbol{X}_{\delta} + \boldsymbol{N} | \boldsymbol{V}_{\delta}\right) \leq \frac{\delta}{2}\liminf_{n\to\infty}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\delta}\right). \tag{A6}$$

A similar assumption to that in [110] is that

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\delta}\right) = \sigma^2_{\inf} < \infty, \tag{A7}$$

for any pair $\{\boldsymbol{X}_{n,\delta}, \boldsymbol{V}_{n,\delta}\}_{\delta>0}$. However, in our setting $\boldsymbol{X}_{n,\delta}$ is independent of δ and $\boldsymbol{V}_{n,\delta} = \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n$ is also independent of δ. Thus, the above convergence requirement reduces simply to

$$\liminf_{n\to\infty}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}\right) = \sigma^2_{\inf}(\mathsf{snr}) < \infty, \tag{A8}$$

where we emphasize the dependence on $\boldsymbol{V}_n$, meaning the dependence on snr.

Now, instead of considering the lower bound on the normalized conditional mutual information as done in [110] going through the truncation argument, we examine the I-MMSE relationship directly:

$$\frac{\mathsf{d}}{\mathsf{d}\mathsf{snr}}\frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) = \frac{1}{2}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\mathsf{snr}}\right), \tag{A9}$$

where

$$\mathbf{K}_{\mathsf{snr}} = \mathbb{E}\left[\left(\boldsymbol{X}_n - \mathbb{E}\left[\boldsymbol{X}_n|\sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right]\right)\left(\boldsymbol{X}_n - \mathbb{E}\left[\boldsymbol{X}_n|\sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right]\right)^{\mathsf{T}}\right]. \tag{A10}$$

As shown in the "incremental proof" of the I-MMSE in [8] using the chain rule for mutual information and data processing arguments, the above is equivalent to

$$\lim_{\delta\to 0}\frac{\frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\delta}\boldsymbol{X}_n + \boldsymbol{N}_{1,n} | \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_{2,n}\right)}{\delta} = \frac{1}{2}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\mathsf{snr}}\right), \tag{A11}$$

where $\boldsymbol{N}_{1,n}$ is independent of $\boldsymbol{N}_{2,n}$ and both are standard Gaussian. The above can also be equivalently written as follows:

$$\frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) = \frac{1}{2}\int_0^{\mathsf{snr}}\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_{\gamma}\right)\mathsf{d}\gamma. \tag{A12}$$

We take the above integral form of this result in order to apply the reverse Fatou lemma. Denote for any $\gamma \geq 0$

$$\limsup_{n\to\infty} \frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) = \sigma^2_{\mathrm{sup}}(\gamma) < \infty, \tag{A13}$$

where the above is bounded again due to the assumption of an average power coonstraint on the input. We have that

$$\limsup_{n\to\infty} \int_0^{\mathsf{snr}} \left|\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) - \sigma^2_{\mathrm{sup}}\right| \mathsf{d}\gamma \leq \int_0^{\mathsf{snr}} \limsup_{n\to\infty} \left|\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) - \sigma^2_{\mathrm{sup}}\right| \mathsf{d}\gamma = 0, \tag{A14}$$

where the inequality is due to the reverse Fatou lemma, where

$$f_n = \left|\frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) - \sigma^2_{\mathrm{sup}}\right| \leq 1 + \sigma^2_{\mathrm{sup}}, \quad \forall n \geq 1, \tag{A15}$$

due to the fact that $\mathbf{K}_\gamma \preceq \mathbb{E}\left[\boldsymbol{X}_n \boldsymbol{X}_n^{\mathsf{T}}\right]$ and the power constraint asusmed on the input sequence. Due to the non-negativity of the integrand we have that

$$\limsup_{n\to\infty} \int_0^{\mathsf{snr}} \frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) \mathsf{d}\gamma = \int_0^{\mathsf{snr}} \sigma^2_{\mathrm{sup}}(\gamma)\mathsf{d}\gamma. \tag{A16}$$

Thus, putting everything together we have that

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) &\overset{a}{=} \limsup_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) \\
&\overset{b}{=} \limsup_{n\to\infty} \int_0^{\mathsf{snr}} \frac{1}{n}\mathrm{Tr}\left(\mathbf{K}_\gamma\right) \mathsf{d}\gamma \\
&\overset{c}{=} \frac{1}{2}\int_0^{\mathsf{snr}} \sigma^2_{\mathrm{sup}}(\gamma)\mathsf{d}\gamma,
\end{aligned} \tag{A17}$$

where a is due to assumption that the limit of the normalized mutual information exists; b is due to (A12); and c is due to (A16), meaning a consequence of Fatou's lemma.

Taking the derivative with respect to snr of both sides we have that

$$\frac{\mathsf{d}}{\mathsf{d}\mathsf{snr}} \lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_n\right) = \frac{1}{2}\sigma^2_{\mathrm{sup}}(\mathsf{snr}). \tag{A18}$$

We can again follow the arguments in the "incremental proof" of the I-MMSE in [8] using the chain rule for mutual information and data processing on the normalized mutual information in the limit to obtain that the above is equivalent to

$$\lim_{\delta\to 0} \frac{\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\delta}\boldsymbol{X}_n + \boldsymbol{N}_{1,n} | \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_{2,n}\right)}{\delta} = \frac{1}{2}\sigma^2_{\mathrm{sup}}(\mathsf{snr}), \tag{A19}$$

where in the second equation we again apply the same steps as in the "incremental channel" proof as in [8] on the normalized mutual information in the limit. Thus, we have that

$$\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\delta}\boldsymbol{X}_n + \boldsymbol{N}_{1,n} | \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_{2,n}\right) = \frac{\delta}{2}\sigma^2_{\mathrm{sup}}(\mathsf{snr}) + o(\delta). \tag{A20}$$

Putting this together with the upper bound that we have obtained, we get that

$$\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\delta}\boldsymbol{X}_n + \boldsymbol{N}_{1,n} | \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_{2,n}\right) = \frac{\delta}{2}\sigma^2_{\mathrm{sup}}(\mathsf{snr}) + o(\delta) \leq \frac{\delta}{2}\sigma^2_{\mathrm{inf}}(\mathsf{snr}) + o(\delta). \tag{A21}$$

But since by definition

$$\sigma^2_{\text{inf}}(\mathsf{snr}) \leq \sigma^2_{\text{sup}}(\mathsf{snr}), \tag{A22}$$

for all $\mathsf{snr} \geq 0$ we have that

$$\lim_{n\to\infty} \frac{1}{n} I\left(\boldsymbol{X}_n; \sqrt{\delta}\boldsymbol{X}_n + \boldsymbol{N}_{1,n} | \sqrt{\mathsf{snr}}\boldsymbol{X}_n + \boldsymbol{N}_{2,n}\right) = \frac{\delta}{2}\sigma^2_{\text{sup}}(\mathsf{snr}) + o(\delta) = \frac{\delta}{2}\sigma^2_{\text{inf}}(\mathsf{snr}) + o(\delta), \tag{A23}$$

and

$$\sigma^2_{\text{inf}}(\mathsf{snr}) = \sigma^2_{\text{sup}}(\mathsf{snr}). \tag{A24}$$

This concludes the proof. □

Appendix B. Proof of Proposition 5

Before showing the proof of Proposition 5 we present several auxiliary results and definitions. We define the conditional MMPE as follows.

Definition A1. *For any* $\mathbf{X}$ *and* $\mathbf{V}$*, the conditional MMPE of* $\mathbf{X}$ *given* $\mathbf{V}$ *is defined as*

$$\text{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{V}) := \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}}, \mathbf{V})\|_p^p. \tag{A25}$$

The conditional MMPE in (A25) reflects the fact that the optimal estimator has been given additional information in the form of $\mathbf{V}$. Note that when $\mathbf{Z}$ is independent of $(\mathbf{X}, \mathbf{V})$ we can write the conditional MMPE for $\mathbf{X_u} \sim P_{\mathbf{X}|\mathbf{V}}(\cdot|\mathbf{v})$ as

$$\text{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{V}) = \int \text{mmpe}(\mathbf{X_v}, \mathsf{snr}, p)\, dP_{\mathbf{V}}(\mathbf{v}). \tag{A26}$$

Since giving extra information does not increase the estimation error, we have the following result.

Proposition A1. (Conditioning reduces the MMPE [25].) *For every* $\mathsf{snr} \geq 0$*, and random variable* $\mathbf{X}$*, we have*

$$\text{mmpe}(\mathbf{X}, \mathsf{snr}, p) \geq \text{mmpe}(\mathbf{X}, \mathsf{snr}, p|\mathbf{V}). \tag{A27}$$

Finally, the following proposition generalizes [23] and states that the MMPE estimation of $\mathbf{X}$ from two observations is equivalent to estimating $\mathbf{X}$ from a single observation with a higher SNR.

Proposition A2 ([25])**.** *For every* $\mathbf{X}$ *and* $p \geq 0$*, let* $\mathbf{V} = \sqrt{\Delta} \cdot \mathbf{X} + \mathbf{Z}_\Delta$ *where* $\mathbf{Z}_\Delta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ *and where* $(\mathbf{X}, \mathbf{Z}, \mathbf{Z}_\Delta)$ *are mutually independent. Then*

$$\text{mmpe}(\mathbf{X}, \mathsf{snr}_0, p|\mathbf{V}) = \text{mmpe}(\mathbf{X}, \mathsf{snr}_0 + \Delta, p). \tag{A28}$$

Proof. For two independent observations $\mathbf{Y}_{\mathsf{snr}_0} = \sqrt{\mathsf{snr}_0}\mathbf{X} + \mathbf{Z}$ and $\mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta$ where $\mathbf{Z}_\Delta$ and $\mathbf{Z}$ are independent, by using maximal ratio combining, we have that

$$\begin{aligned}\mathbf{Y}_{\mathsf{snr}} &= \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0 + \Delta}}\mathbf{Y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0 + \Delta}}\mathbf{Y}_{\mathsf{snr}_0} \\ &= \sqrt{\mathsf{snr}_0 + \Delta}\mathbf{X} + \mathbf{W},\end{aligned}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Next by using the same argument as in [23], we have that the conditional probabilities are

$$p_{\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_\Delta}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}_0},\mathbf{y}_\Delta) = p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}_{\mathsf{snr}}), \tag{A29}$$

for $\mathbf{y}_{\mathsf{snr}} = \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{y}_{\mathsf{snr}_0}$. The equivalence of the posterior probabilities implies that the estimation of $\mathbf{X}$ from $\mathbf{Y}_{\mathsf{snr}}$ is as good as the estimation of $\mathbf{X}$ from $(\mathbf{Y}_{\mathsf{snr}_0},\mathbf{Y}_\Delta)$. This concludes the proof. □

We are now in a position to prove the SCPP bound in Proposition 5.

Proof. Let $\mathsf{snr} = \mathsf{snr}_0 + \Delta$ for $\Delta \geq 0$, and let $\mathbf{Y}_\Delta = \sqrt{\Delta}\mathbf{X} + \mathbf{Z}_\Delta$. Then

$$\begin{aligned}\mathbf{Y}_{\mathsf{snr}} &= \frac{\sqrt{\Delta}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_\Delta + \frac{\sqrt{\mathsf{snr}_0}}{\sqrt{\mathsf{snr}_0+\Delta}}\mathbf{Y}_{\mathsf{snr}_0}\\ &= \sqrt{\mathsf{snr}_0+\Delta}\mathbf{X} + \mathbf{W},\end{aligned}$$

where $\mathbf{W} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$. Next, let

$$m := \mathrm{mmpe}^{\frac{2}{p}}(\mathbf{X},\mathsf{snr}_0,p) = \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})\|_p^2, \tag{A30}$$

and define a suboptimal estimator given $(\mathbf{Y}_\Delta,\mathbf{Y}_{\mathsf{snr}_0})$ as

$$\hat{\mathbf{X}} = \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Y}_\Delta + \gamma f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}), \tag{A31}$$

for some $\gamma \in \mathbb{R}$ to be determined later. Then

$$\mathbf{X} - \hat{\mathbf{X}} = \gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta,$$

and

$$\begin{aligned}\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X},\mathsf{snr},p) &= \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}})\|_p\\ &\overset{a)}{=} \|\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_\Delta,\mathbf{Y}_{\mathsf{snr}_0})\|_p\\ &\overset{b)}{\leq} \|\mathbf{X} - \hat{\mathbf{X}}\|_p = \left\|\gamma(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \frac{(1-\gamma)}{\sqrt{\Delta}}\mathbf{Z}_\Delta\right\|_p\\ &\overset{c)}{=} \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0})) - \sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m},\end{aligned} \tag{A32}$$

where the (in)-equalities follow from: (a) Proposition A2; (b) by using the sub-optimal estimator in (A31); and (c) by choosing $\gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2+\Delta\cdot m}$ for m defined in (A30).

Next, by applying the triangle inequality to (A32) we get

$$\begin{aligned}\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X},\mathsf{snr},p) &\leq \frac{\left\|\|\mathbf{Z}\|_p^2(\mathbf{X} - f_p(\mathbf{X}|\mathbf{Y}_{\mathsf{snr}_0}))\right\|_p + \left\|\sqrt{\Delta}\cdot m\cdot\mathbf{Z}_\Delta\right\|_p}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}\\ &= \frac{\sqrt{m}\|\mathbf{Z}\|_p\cdot(\|\mathbf{Z}\|_p + \sqrt{\Delta}\cdot\sqrt{m})}{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}\\ &\leq \sqrt{2}\frac{\sqrt{m}\|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta\cdot m}},\end{aligned} \tag{A33}$$

where in the last step we have used $(a+b) \leq \sqrt{2}\sqrt{a^2+b^2}$.

Note that for the case $p = 2$, instead of using the triangle inequality in (A33), the term in (A32) can be expanded into a quadratic equation for which it is not hard to see that the choice of $\gamma = \frac{\|\mathbf{Z}\|_p^2}{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}$ is optimal and leads to the bound

$$\mathrm{mmpe}^{\frac{1}{p}}(\mathbf{X}, \mathsf{snr}, p) \leq \frac{\sqrt{m}\|\mathbf{Z}\|_p}{\sqrt{\|\mathbf{Z}\|_p^2 + \Delta \cdot m}}.$$

The proof is concluded by noting that $\beta = \frac{m}{\|\mathbf{Z}\|_p^2 - \mathsf{snr}_0 m}$. □

References

1. Stam, A.J. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112.
2. Esposito, R. On a relation between detection and estimation in decision theory. *Inf. Control* **1968**, *12*, 116–120.
3. Hatsell, C.; Nolte, L. Some geometric properties of the likelihood ratio. *IEEE Trans. Inf. Theor.* **1971**, *17*, 616–618.
4. Duncan, T.E. Evaluation of likelihood functions. *Inf. Control* **1968**, *13*, 62–74.
5. Kadota, T.; Zakai, M.; Ziv, J. Mutual information of the white Gaussian channel with and without feedback. *IEEE Trans. Inf. Theor.* **1971**, *17*, 368–371.
6. Kailath, T. The innovations approach to detection and estimation theory. *Proc. IEEE* **1970**, *58*, 680–695.
7. Duncan, T.E. On the calculation of mutual information. *SIAM J. Appl. Math.* **1970**, *19*, 215–220.
8. Guo, D.; Shamai (Shitz), S.; Verdú, S. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inf. Theor.* **2005**, *51*, 1261–1282.
9. Palomar, D.P.; Verdú, S. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. Inf. Theor.* **2006**, *52*, 141–154.
10. Han, G.; Song, J. Extensions of the I-MMSE Relationship to Gaussian Channels With Feedback and Memory. *IEEE Trans. Inf. Theor.* **2016**, *62*, 5422–5445.
11. Guo, D.; Shamai (Shitz), S.; Verdú, S. Additive non-Gaussian noise channels: Mutual information and conditional mean estimation. In Proceedings of the IEEE International Symposium on Information Theory, Adelaide, Australia, 4–9 September 2005; pp. 719–723.
12. Guo, D.; Shamai (Shitz), S.; Verdú, S. Mutual information and conditional mean estimation in Poisson channels. *IEEE Trans. Inf. Theor.* **2008**, *54*, 1837–1849.
13. Atar, R.; Weissman, T. Mutual information, relative entropy, and estimation in the Poisson channel. *IEEE Trans. Inf. Theor.* **2012**, *58*, 1302–1318.
14. Jiao, J.; Venkat, K.; Weissman, T. Relation between Information and Estimation in Discrete-Time Lévy Channels. *IEEE Trans. Inf. Theor.* **2017**, *63*, 3579–3594.
15. Taborda, C.G.; Guo, D.; Perez-Cruz, F. Information-estimation relationships over binomial and negative binomial models. *IEEE Trans. Inf. Theor.* **2014**, *60*, 2630–2646.
16. Verdú, S. Mismatched estimation and relative entropy. *IEEE Trans. Inf. Theor.* **2010**, *56*, 3712–3720.
17. Guo, D. Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation. In Proceedings of the IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 814–818.
18. Zakai, M. On mutual information, likelihood ratios, and estimation error for the additive Gaussian channel. *IEEE Trans. Inf. Theor.* **2005**, *51*, 3017–3024.
19. Duncan, T.E. Mutual information for stochastic signals and fractional Brownian motion. *IEEE Trans. Inf. Theor.* **2008**, *54*, 4432–4438.
20. Duncan, T.E. Mutual information for stochastic signals and Lévy processes. *IEEE Trans. Inf. Theor.* **2010**, *56*, 18–24.
21. Weissman, T.; Kim, Y.H.; Permuter, H.H. Directed information, causal estimation, and communication in continuous time. *IEEE Trans. Inf. Theor.* **2013**, *59*, 1271–1287.
22. Venkat, K.; Weissman, T. Pointwise relations between information and estimation in Gaussian noise. *IEEE Trans. Inf. Theor.* **2012**, *58*, 6264–6281.

23. Guo, D.; Shamai (Shitz), S.; Verdú, S. *The Interplay Between Information and Estimation Measures*; Now Publishers: Boston, MA, USA, 2013.
24. Ozarow, L.; Wyner, A. On the capacity of the Gaussian channel with a finite number of input levels. *IEEE Trans. Inf. Theor.* **1990**, *36*, 1426–1428.
25. Dytso, A.; Bustin, R.; Tuninetti, D.; Devroye, N.; Poor, H.V.; Shamai (Shitz), S. On the minimum mean p-th error in Gaussian noise channels and its applications. *arXiv* **2016**, arXiv:1607.01461.
26. Sherman, S. Non-mean-square error criteria. *IRE Trans. Inf. Theor.* **1958**, *4*, 125–126.
27. Akyol, E.; Viswanatha, K.B.; Rose, K. On conditions for linearity of optimal estimation. *IEEE Trans. Inf. Theor.* **2012**, *58*, 3497–3508.
28. Bustin, R.; Schaefer, R.F.; Poor, H.V.; Shamai (Shitz), S. On the SNR-evolution of the MMSE function of codes for the Gaussian broadcast and wiretap channels. *IEEE Trans. Inf. Theor.* **2016**, *62*, 2070–2091.
29. Bustin, R.; Poor, H.V.; Shamai (Shitz), S. The effect of maximal rate codes on the interfering message rate. In Proceedings of the IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 91–95.
30. Guo, D.; Wu, Y.; Shamai (Shitz), S.; Verdú, S. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Trans. Inf. Theor.* **2011**, *57*, 2371–2385.
31. Wu, Y.; Verdú, S. Functional properties of minimum mean-square error and mutual information. *IEEE Trans. Inf. Theor.* **2012**, *58*, 1289–1301.
32. Wu, Y.; Verdú, S. MMSE dimension. *IEEE Trans. Inf. Theor.* **2011**, *57*, 4857–4879.
33. Bustin, R.; Payaró, M.; Palomar, D.P.; Shamai (Shitz), S. On MMSE crossing properties and implications in parallel vector Gaussian channels. *IEEE Trans. Inf. Theor.* **2013**, *59*, 818–844.
34. Bustin, R.; Schaefer, R.F.; Poor, H.V.; Shamai (Shitz), S. On MMSE properties of optimal codes for the Gaussian wiretap channel. In Proceedings of the IEEE Information Theory Workshop, Jerusalem, Israel, 26 April–1 May 2015; pp. 1–5.
35. Verdú, S.; Guo, D. A simple proof of the entropy-power inequality. *IEEE Trans. Inf. Theor.* **2006**, *52*, 2165–2166.
36. Guo, D.; Shamai (Shitz), S.; Verdú, S. Proof of entropy power inequalities via MMSE. In Proceedings of the IEEE International Symposium on Information Theory, Seattle, WA, USA, 9–14 July 2006; pp. 1011–1015.
37. Tulino, A.M.; Verdú, S. Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof. *IEEE Trans. Inf. Theor.* **2006**, *52*, 4295–4297.
38. Dytso, A.; Bustin, R.; Poor, H.V.; Shamai (Shitz), S. Comment on the equality condition for the I-MMSE proof of the entropy power inequality. *arXiv* **2017**, arXiv:1703.07442.
39. Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley: Hoboken, NJ, USA, 2006.
40. Shannon, C. A mathematical theory of communication. Available online: http://math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf (accessed on 3 August 2017).
41. Csiszar, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
42. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2012.
43. Abbe, E.; Zheng, L. A coordinate system for Gaussian networks. *IEEE Trans. Inf. Theor.* **2012**, *58*, 721–733.
44. Bennatan, A.; Shamai (Shitz), S.; Calderbank, A. Soft-decoding-based strategies for relay and interference channels: analysis and achievable rates using LDPC codes. *IEEE Trans. Inf. Theor.* **2014**, *60*, 1977–2009.
45. Dytso, A.; Tuninetti, D.; Devroye, N. Interference as noise: Friend or foe? *IEEE Trans. Inf. Theor.* **2016**, *62*, 3561–3596.
46. Moshksar, K.; Ghasemi, A.; Khandani, A. An alternative to decoding interference or treating interference as Gaussian noise. *IEEE Trans. Inf. Theor.* **2015**, *61*, 305–322.
47. Shamai (Shitz), S. From constrained signaling to network interference alignment via an information-estimation perspective. *IEEE Inf. Theor. Soc. Newslett.* **2012**, *62*, 6–24.
48. Ungerboeck, G. Channel coding with multilevel/phase signals. *IEEE Trans. Inf. Theor.* **1982**, *28*, 55–67.
49. Dytso, A.; Tuninetti, D.; Devroye, N. On the two-user interference channel with lack of knowledge of the interference codebook at one receiver. *IEEE Trans. Inf. Theor.* **2015**, *61*, 1257–1276.
50. Dytso, A.; Bustin, R.; Tuninetti, D.; Devroye, N.; Poor, H.V.; Shamai (Shitz), S. New bounds on MMSE and applications to communication with the disturbance constraint. Available online: https://arxiv.org/pdf/1603.07628.pdf (accessed on 3 August 2017).

51. Dong, Y.; Farnia, F.; Özgür, A. Near optimal energy control and approximate capacity of energy harvesting communication. *IEEE J. Selected Areas Commun.* **2015**, *33*, 540–557.
52. Shaviv, D.; Nguyen, P.M.; Özgür, A. Capacity of the energy-harvesting channel with a finite battery. *IEEE Trans. Inf. Theor.* **2016**, *62*, 6436–6458.
53. Bloch, M.; Barros, J.; Rodrigues, M.R.; McLaughlin, S.W. Wireless information-theoretic security. *IEEE Trans. Inf. Theor.* **2008**, *54*, 2515–2534.
54. Dytso, A.; Bustin, R.; Tuninetti, D.; Devroye, N.; Poor, H.V.; Shamai, S. On the applications of the minimum mean p-th error (MMPE) to information theoretic quantities. In Proceedings of the IEEE Information Theory Workshop, Cambridge, UK, 11–14 September 2016; pp. 66–70.
55. Forney, G.D., Jr. On the role of MMSE estimation in approaching the information theoretic limits of linear Gaussian channels: Shannon meets Wiener. In Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 1–3 October 2003.
56. Alvarado, A.; Brannstrom, F.; Agrell, E.; Koch, T. High-SNR asymptotics of mutual information for discrete constellations with applications to BICM. *IEEE Trans. Inf. Theor.* **2014**, *60*, 1061–1076.
57. Wu, Y.; Verdú, S. The impact of constellation cardinality on Gaussian channel capacity. In Proceedings of the 48th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 29 September–11 October 2010; pp. 620–628.
58. Dytso, A.; Tuninetti, D.; Devroye, N. On discrete alphabets for the two-user Gaussian interference channel with one receiver lacking knowledge of the interfering codebook. In Proceedings of the Information Theory and Applications Workshop, San Diego, CA, USA, 9–14 February 2014; pp. 1–8.
59. Dytso, A.; Goldenbaum, M.; Poor, H.V.; Shamai (Shitz), S. A generalized Ozarow-Wyner capacity bound with applications. In Proceedings of the IEEE International Symposium on Information Theory, Aachen, Germany, 25–30 June 2017; pp. 1058–1062.
60. Dytso, A.; Bustin, R.; Poor, H.V.; Shamai (Shitz), S. On additive channels with generalized Gaussian noise. In Proceedings of the IEEE International Symposium on Information Theory, Aachen, Germany, 25–30 June 2017.
61. Dytso, A.; Goldenbaum, M.; Shamai (Shitz), S.; Poor, H.V. Upper and lower bounds on the capacity of amplitude-constrained MIMO channels. In Proceedings of the IEEE Global Communications Conference, Singapore, 4–8 December 2017.
62. Peleg, M.; Sanderovich, A.; Shamai (Shitz), S. On extrinsic information of good codes operating over memoryless channels with incremental noisiness. In Proceedings of the IEEE 24th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, 15–17 November 2006; pp. 290–294.
63. Merhav, N.; Guo, D.; Shamai (Shitz), S. Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions. *IEEE Trans. Inf. Theor.* **2010**, *56*, 1400–1416.
64. Wyner, A.D. The wire-tap channel. *Bell Lab. Tech. J.* **1975**, *54*, 1355–1387.
65. Csiszár, I.; Korner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theor.* **1978**, *24*, 339–348.
66. Leung-Yan-Cheong, S.; Hellman, M. The Gaussian wire-tap channel. *IEEE Trans. Inf. Theor.* **1978**, *24*, 451–456.
67. Massey, J.L. A simplified treatment of Wyner's wire-tap channel. In Proceedings of the 21st Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 3–4 October 1983; pp. 268–276.
68. Bloch, M.; Barros, J. *Physical-Layer Security: From Information Theory to Security Engineering*; Cambridge University Press: Cambridge, UK, 2011.
69. Bandemer, B.; El Gamal, A. Communication with disturbance constraints. *IEEE Trans. Inf. Theor.* **2014**, *60*, 4488–4502.
70. Bustin, R.; Shamai (Shitz), S. MMSE of 'bad' codes. *IEEE Trans. Inf. Theor.* **2013**, *59*, 733–743.
71. Shang, X.; Kramer, G.; Chen, B. A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels. *IEEE Trans. Inf. Theor.* **2009**, *55*, 689–699.
72. Motahari, A.S.; Khandani, A.K. Capacity bounds for the Gaussian interference channel. *IEEE Trans. Inf. Theor.* **2009**, *55*, 620–643.
73. Annapureddy, V.S.; Veeravalli, V.V. Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. *IEEE Trans. Inf. Theor.* **2009**, *55*, 3032–3050.
74. Han, T.; Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theor.* **1981**, *27*, 49–60.

75. Sato, H. The capacity of Gaussian interference channel under strong interference. *IEEE Trans. Inf. Theor.* **1981**, *27*, 786–788.
76. Etkin, R.; Tse, D.; Wang, H. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inf. Theor.* **2008**, *54*, 5534–5562.
77. Ahlswede, R. Multi-way communication channels. In Proceedings of the IEEE International Symposium on Information Theory, Ashkelon, Israel, 25–29 June 1973; pp. 23–52.
78. Gherekhloo, S.; Chaaban, A.; Sezgin, A. Expanded GDoF-optimality regime of treating interference as noise in the M×2 X-channel. *IEEE Trans. Inf. Theor.* **2016**, *63*, 355–376.
79. Cheng, R.; Verdú, S. On limiting characterizations of memoryless multiuser capacity regions. *IEEE Trans. Inf. Theor.* **1993**, *39*, 609–612.
80. Blasco-Serrano, R.; Thobaben, R.; Skoglund, M. Communication and interference coordination. In Proceedings of the Information Theory and Applications Workshop, San Diego, CA, USA, 9–14 February 2014; pp. 1–8.
81. Huleihel, W.; Merhav, N. Analysis of mismatched estimation errors using gradients of partition functions. *IEEE Trans. Inf. Theor.* **2014**, *60*, 2190–2216.
82. Dytso, A.; Bustin, R.; Tuninetti, D.; Devroye, N.; Poor, H.V.; Shamai (Shitz), S. On communications through a Gaussian noise channel with an MMSE disturbance constraint. In Proceedings of the Information Theory and Applications Workshop, San Diego, CA, USA, 1–5 February 2016; pp. 1–8.
83. Cover, T. Broadcast channels. *IEEE Trans. Inf. Theor.* **1972**, *18*, 2–14.
84. Cover, T. Comments on broadcast channels. *IEEE Trans. Inf. Theor.* **1998**, *44*, 2524–2530.
85. Gallager, R.G. Capacity and coding for degraded broadcast channels. *Problemy Peredachi Informatsii* **1974**, *10*, 3–14.
86. Bergmans, P. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Trans. Inf. Theor.* **1974**, *20*, 279–280.
87. Bustin, R.; Schaefer, R.F.; Poor, H.V.; Shamai (Shitz), S. On MMSE properties of "good" and "bad" codes for the Gaussian broadcast channel. In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 14–19.
88. Bustin, R.; Schaefer, R.F.; Poor, H.V.; Shamai (Shitz), S. An I-MMSE based graphical representation of rate and equivocation for the Gaussian broadcast channel. In Proceedings of the IEEE Conference on Communications and Network Security, Florence, Italy, 28–30 September 2015; pp. 53–58.
89. Zou, S.; Liang, Y.; Lai, L.; Shamai (Shitz), S. Degraded broadcast channel: Secrecy outside of a bounded range. In Proceedings of the IEEE Information Theory Workshop, Jerusalem, Israel, 26 April–1 May 2015; pp. 1–5.
90. Wu, Y.; Shamai (Shitz), S.; Verdú, S. Information dimension and the degrees of freedom of the interference channel. *IEEE Trans. Inf. Theor.* **2015**, *61*, 256–279.
91. Carleial, A. A case where interference does not reduce capacity. *IEEE Trans. Inf. Theor.* **1975**, *21*, 569–570.
92. Costa, M.H.; El Gamal, A. The capacity region of the discrete memoryless interference channel with strong interference. *IEEE Trans. Inf. Theor.* **1987**, *33*, 710–711.
93. El Gamal, A.; Costa, M. The capacity region of a class of deterministic interference channels. *IEEE Trans. Inf. Theor.* **1982**, *28*, 343–346.
94. Bresler, G.; Tse, D. The two-user Gaussian interference channel: a deterministic view. *Trans. Emerg. Telecommun. Technol.* **2008**, *19*, 333–354.
95. Telatar, E.; Tse, D. Bounds on the capacity region of a class of interference channels. In Proceedings of the IEEE International Symposium on Information Theory, Toronto, ON, Canada, 6–11 July 2008; pp. 2871–2874.
96. Nair, C.; Xia, L.; Yazdanpanah, M. Sub-optimality of Han-Kobayashi achievable region for interference channels. In Proceedings of the IEEE International Symposium on Information Theory, Hong Kong, China, 14–19 June 2015; pp. 2416–2420.
97. Tse, D. It's easier to approximate. *IEEE Infor. Theor. Soc. Newslett.* **2010**, *60*, 6–11.
98. Cadambe, V.R.; Jafar, S.A. Interference alignment and degrees of freedom of the *K*-user interference channel. *IEEE Trans. Inf. Theor.* **2008**, *54*, 3425–3441.
99. Huang, C.; Cadambe, V.; Jafar, S. On the capacity and generalized degrees of freedom of the X channel. *arXiv* **2008**, arXiv:0810.4741.
100. Jafar, S.A. Interference alignment—A new look at signal dimensions in a communication network. *Found. Trend. Commun. Inf. Theor.* **2011**, *7*, 1–134.

101. Bustin, R.; Liu, R.; Poor, H.V.; Shamai (Shitz), S. An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel. *EURASIP J. Wirel. Commun. Netw.* **2009**, *1*, 370970.
102. Liu, R.; Liu, T.; Poor, H.V.; Shamai (Shitz), S. A vector generalization of Costa's entropy-power inequality with applications. *IEEE Trans. Inf. Theor.* **2010**, *56*, 1865–1879.
103. Pérez-Cruz, F.; Rodrigues, M.R.; Verdú, S. MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. *IEEE Trans. Inf. Theor.* **2010**, *56*, 1070–1084.
104. Wu, Y.; Verdú, S. Optimal phase transitions in compressed sensing. *IEEE Trans. Inf. Theor.* **2012**, *58*, 6241–6263.
105. Chechik, G.; Globerson, A.; Tishby, N.; Weiss, Y. Information bottleneck for Gaussian variables. *J. Mach. Learn. Res.* **2005**, *6*, 165–188.
106. Liu, T.; Viswanath, P. An extremal inequality motivated by multiterminal information-theoretic problems. *IEEE Trans. Inf. Theor.* **2007**, *53*, 1839–1851.
107. Weingarten, H.; Steinberg, Y.; Shamai (Shitz), S. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theor.* **2006**, *52*, 3936–3964.
108. Liu, T.; Shamai (Shitz), S. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. Inf. Theor.* **2009**, *55*, 2547–2553.
109. Rini, S.; Tuninetti, D.; Devroye, N. On the capacity of the Gaussian cognitive interference channel: New inner and outer bounds and capacity to within 1 bit. *IEEE Trans. Inf. Theor.* **2012**, *58*, 820–848.
110. Wu, Y.; Guo, D.; Verdú, S. Derivative of mutual information at zero SNR: The Gaussian-noise case. *IEEE Trans. Inf. Theor.* **2011**, *57*, 7307–7312.