On Communication through a Gaussian Channel with an MMSE Disturbance Constraint

Alex Dytso, Ronit Bustin, Daniela Tuninetti, Natasha Devroye, H.Vincent Poor, Shlomo Shamai (Shitz)

Abstract

This paper considers a Gaussian channel with one transmitter and two receivers. The goal is to maximize the communication rate at the intended/primary receiver subject to a disturbance constraint at the unintended/secondary receiver. The disturbance is measured in terms of minimum mean square error (MMSE) of the interference that the transmission to the primary receiver inflicts on the secondary receiver.

The paper presents a new upper bound for the problem of maximizing the mutual information subject to an MMSE constraint. The new bound holds for vector inputs of any length and recovers a previously known limiting (when the length of vector input tends to infinity) expression from the work of Bustin *et al.* The key technical novelty is a new upper bound on the MMSE. This bound allows one to bound the MMSE for all signal-to-noise ratio (SNR) values *below* a certain SNR at which the MMSE is known (which corresponds to the disturbance constraint). This bound complements the 'single-crossing point property' of the MMSE that upper bounds the MMSE for all SNR values *above* a certain value

Alex Dytso, Daniela Tuninetti and Natasha Devroye are with the department of Electrical and Computer Engineering, University of Illinois at Chicago, IL, Chicago 60607, USA (e-mail: odytso2, danielat, devroye @ uic.edu). Ronit Busitn is with the department of Electrical Engineering - Systems, Tel Aviv University, Tel Aviv 6997801, Israel (email:ronitbustin@post.tau.ac.il). H.Vincent Poor is with the department of Electrical and Computer Engineering, Princeton University, NJ, Princeton 08544, USA (email:poor@princeton.edu). S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: sshlomo@ee.technion.ac.il). The work of Alex Dytso, Daniela Tuninetti and Natasha Devroye was partially funded by NSF under award 1422511. The work of Ronit Bustin was supported in part by the Women Postdoctoral Scholarship of Israel's Council for Higher Education (VATAT) 2014-2015. The work of H. Vincent Poor and Ronit Bustin was partially supported by NSF under awards CCF-1420575 and ECCS-1343210. The work of Shlomo Shamai was supported by the Israel Science Foundation and the S. and N. Grand Research Fund. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies. The work was presented in part in [1].

at which the MMSE value is known. The MMSE upper bound provides a refined characterization of the phase-transition phenomenon which manifests, in the limit as the length of the vector input goes to infinity, as a discontinuity of the MMSE for the problem at hand.

For vector inputs of size $n = 1$, a matching lower bound, to within an additive gap of order $O\left(\log\log \frac{1}{\mathsf{MMSE}}\right)$ (where MMSE is the disturbance constraint), is shown by means of the mixed inputs technique recently introduced by Dytso *et al.*

I. Introduction

Consider a Gaussian noise channel with one transmitter and two receivers:

$$\mathbf{Y} = \sqrt{\mathsf{snr}}\ \mathbf{X} + \mathbf{Z}, \tag{1a}$$

$$\mathbf{Y}_{\mathsf{snr}_0} = \sqrt{\mathsf{snr}_0}\ \mathbf{X} + \mathbf{Z}_0, \tag{1b}$$

where $\mathbf{Z}, \mathbf{Z}_0, \mathbf{X}, \mathbf{Y}, \mathbf{Y}_{\mathsf{snr}_0} \in \mathbb{R}^n$, $\mathbf{Z}, \mathbf{Z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and $\mathbf{X}$ and $(\mathbf{Z}, \mathbf{Z}_0)$ are independent.[1] When it will be necessary to stress the SNR at $\mathbf{Y}$ in (1a) we will denote it by $\mathbf{Y}_{\mathsf{snr}}$.

We denote the mutual information between the input $\mathbf{X}$ and output $\mathbf{Y}$ as

$$I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}, \mathsf{snr}) := \mathbb{E}\left[\log\left(\frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{p_{\mathbf{Y}}(\mathbf{Y})}\right)\right]. \tag{2}$$

We also denote the mutual information normalized by n as

$$I_n(\mathbf{X}, \mathsf{snr}) := \frac{1}{n} I(\mathbf{X}, \mathsf{snr}). \tag{3}$$

We denote the minimum mean squared error (MMSE) of estimating $\mathbf{X}$ from $\mathbf{Y}$ as

$$\mathrm{mmse}(\mathbf{X}|\mathbf{Y}) = \mathrm{mmse}(\mathbf{X}, \mathsf{snr}) := \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})\right]\right), \tag{4}$$

where $\mathbf{Cov}(\mathbf{X}|\mathbf{Y})$ is the conditional covariance matrix of $\mathbf{X}$ given $\mathbf{Y}$ and is defined as

$$\mathbf{Cov}(\mathbf{X}|\mathbf{Y}) := \mathbb{E}\left[\left(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\right)\left(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]\right)^{\mathrm{T}} |\mathbf{Y}\right].$$

Moreover, since the distribution of the noise is fixed, the quantities $I(\mathbf{X}; \mathbf{Y})$ and $\mathrm{mmse}(\mathbf{X}|\mathbf{Y})$ are completely determined by $\mathbf{X}$ and snr, and there is no ambiguity in using the notation $I(\mathbf{X}, \mathsf{snr})$ and $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$.

[1] Since there is no cooperation between receivers the capacity depends on $p_{\mathbf{Y}_1,\mathbf{Y}_2|\mathbf{X}}$ only thorough the marginals $p_{\mathbf{Y}_1|\mathbf{X}}$ and $p_{\mathbf{Y}_2|\mathbf{X}}$.

We consider a scenario in which a message, encoded as $\mathbf{X}$, must be decoded at the primary receiver $\mathbf{Y}_{\mathsf{snr}}$ while it is also seen at the unintended/secondary receiver for which it is an interferer. This scenario 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 structure is neglected. It has been shown to be sum-capacity optimal in the so called very-weak interference regime [2], [3], and [4].
2) *Partial interference cancellation*: by using the Han-Kobayashi (HK) achievable scheme [5], 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. This approach has been shown to be capacity achieving in the strong interference regime [6] and optimal within 1/2 bit per channel per user otherwise [7].
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 [8], [9] and [10], and shown to be optimal to within either a constant or a $O(\log\log(\mathsf{snr}))$ gap in [11].

In this paper we look at a somewhat simplified scenario compared to the G-IC as shown in Fig. 1. We assume that there is only one message for the primary receiver, and the primary user inflicts interference (disturbance) on a secondary receiver. The primary transmitter wishes to maximize its comunication rate, while subject to a constraint on the disturbance it inflicts on the secondary receiver. The disturbance is measured in terms of MMSE. Intuitively, the MMSE disturbance constraint quantifies the remaining interference after partial interference cancellation or soft-decoding have been performed [12], [13]. Formally, we aim to solve the following problem.

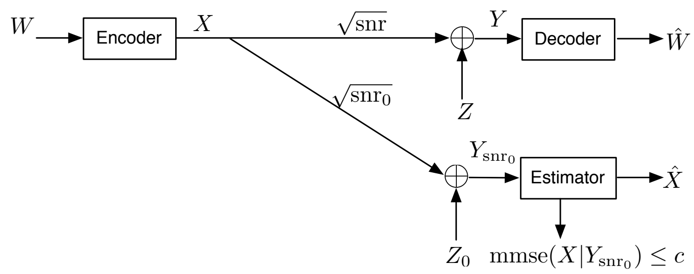


Fig. 1: Channel Model.

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

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} I_n(\mathbf{X}, \mathsf{snr}), \tag{5a}$$

$$\textit{s.t. } \frac{1}{n}\text{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \le 1, \ \textit{power constraint}, \tag{5b}$$

$$\textit{and } \text{mmse}(\mathbf{X}, \mathsf{snr}_0) \le \frac{\beta}{1+\beta\mathsf{snr}_0}, \ \textit{MMSE constraint}. \tag{5c}$$

The subscript n in $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ emphasizes that we seek to find bounds that hold for any input length n. Even though this model is somewhat simplified, compared to the G-IC, it can serve as an important building block towards characterizing the capacity of the G-IC [12], [13].

In [12] the capacity of the channel in Fig. 1 was properly defined and it was shown to be equal to $\lim_{n\to\infty} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$. Note that $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ does not denote the capacity since the MMSE does not 'single-letterize.' Finally, in [13, Sec. VI.3] and [12, Sec. VIII] it was conjectured that the optimal input for $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ is discrete.

A. Notation

Throughout the paper we adopt the following notational conventions: deterministic scalar quantities are denoted by lowercase letters and deterministic vector quantities are denoted by lowercase bold letters; matrices are denoted by bold uppercase letters; random variables are denoted by uppercase letters and random vectors are denoted by bold uppercase letters; all logarithms are taken to be base e; we denote the support of a random variable A by $\mathsf{supp}(A)$; $X \sim \mathsf{PAM}\,(N)$ denotes the pulse-amplitude modulation (PAM) constellation, i.e., the uniform probability mass function over a zero-mean constellation with $|\mathsf{supp}(X)| = N$ points, minimum distance $d_{\min(X)}$, and therefore average energy $\mathbb{E}[X^2] = d_{\min(X)}^2 \frac{N^2-1}{12}$; ordering notation $\mathbf{A} \succeq \mathbf{B}$

implies that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix; we denote the Fisher information matrix of the random vector $\mathbf{A}$ by $\mathbf{J}(\mathbf{A})$; for $x \in \mathbb{R}$ we let $[x]^+ := \max(x, 0)$ and $\log^+(x) := [\log(x)]^+$; we use the Landau notation $f(x) = O(g(x))$ to mean that for some $c > 0$ there exists an x_0 such that $f(x) \le cg(x)$ for all $x \ge x_0$.

B. On Presentation of Results

Throughout the paper we will plot normalized quantities, where the normalization is with respect to the same quantity when the input is $\mathcal{N}(\mathbf{0}, \mathbf{I})$. For example, for mutual information $I_n(\mathbf{X}, \mathsf{snr})$ in (3) we will plot

$$d(\mathbf{X}, \mathsf{snr}) := \frac{I_n(\mathbf{X}, \mathsf{snr})}{\frac{1}{2}\log(1+\mathsf{snr})}, \tag{6}$$

while for MMSE in (4) we will plot

$$D(\mathbf{X}, \mathsf{snr}) := \frac{\mathrm{mmse}(\mathbf{X}, \mathsf{snr})}{\frac{1}{1+\mathsf{snr}}} = (1+\mathsf{snr}) \cdot \mathrm{mmse}(\mathbf{X}, \mathsf{snr}). \tag{7}$$

In particular, at high snr the quantity in (6) is commonly referred to as the *degrees of freedom* [14] and the quantity in (7) as the *MMSE dimension* [15]. Moreover, it is well known that under the block-power constraint in (5b), a Gaussian input maximizes both the mutual information and the MMSE [16], and thus the quantities $d(\mathbf{X}, \mathsf{snr})$, $D(\mathbf{X}, \mathsf{snr})$ have a natural meaning of multiplicative loss of the inputs $\mathbf{X}$ compared to the Gaussian input. Fig. 2 compares normalized and unnormalized quantities.

II. Past Work and Paper Contributions

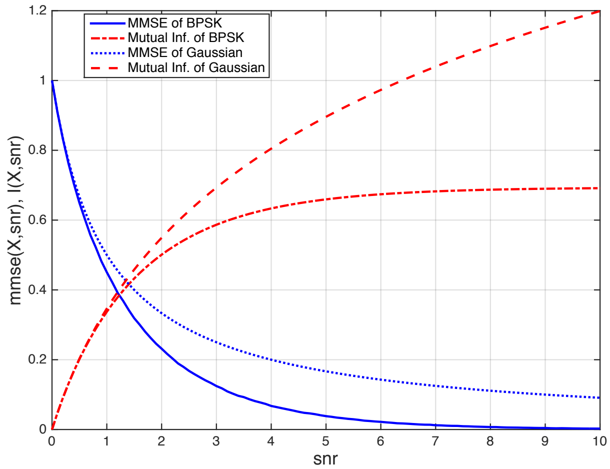
The mutual information and the MMSE are related, for any input $\mathbf{X}$, via the so called *I-MMSE relationship* [17, Theorem 1].

Proposition 1. *(*I-MMSE relationship *[17].) The I-MMSE relationship is given by the derivative relationship*

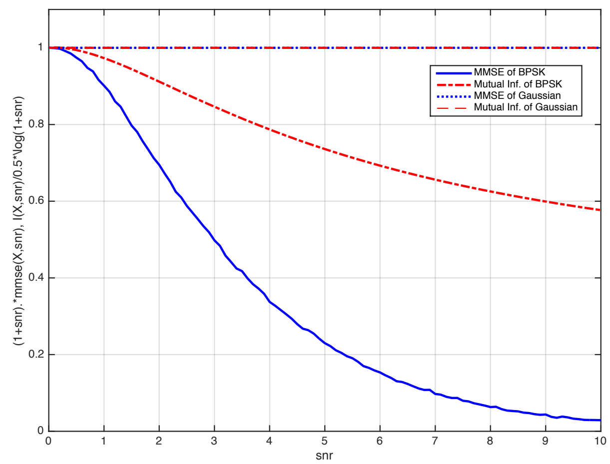
$$\frac{d}{d\mathsf{snr}} I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{8a}$$

or the integral relationship [17, Eq.(47)]

$$I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, t)dt. \tag{8b}$$



(a) Unnormalized plot.



(b) Normalized plot (dashed red and dotted blue lines overlap).

Fig. 2: Comparing mutual informations and MMSE's for BPSK and Gaussian inputs. Fig. 2b clearly shows the multiplicative loss of BPSK, for both mutual information and MMSE, compared to a Gaussian input.

In order to develop bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ we require bounds on the MMSE. An important bound on the MMSE is the following *linear MMSE (LMMSE) upper bound.*

Proposition 2. *(*LMMSE bound [17].*) For any* $\mathbf{X}$ *and* $\mathsf{snr} > 0$ *it holds that*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{1}{\mathsf{snr}}. \tag{9a}$$

If $\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \le \sigma^2$*, then for any* $\mathsf{snr} \ge 0$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{\sigma^2}{1+\sigma^2\mathsf{snr}}, \tag{9b}$$

where equality in (9b) *is achieved iff* $\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

Another important bound for the MMSE is the *single-crossing point property* (SCPP) bound developed in [18] for $n = 1$ and extended in [19] to any $n \ge 1$.

Proposition 3. (SCPP [19].) *For any fixed* $\mathbf{X}$*, suppose that* $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta \mathsf{snr}_0}$*, for some fixed* $\beta \geq 0$*. Then for all* $\mathsf{snr} \in [\mathsf{snr}_0, \infty)$ *we have that*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \frac{\beta}{1+\beta \mathsf{snr}}, \tag{10a}$$

and for all $\mathsf{snr} \in [0, \mathsf{snr}_0)$

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \geq \frac{\beta}{1+\beta \mathsf{snr}}. \tag{10b}$$

In words, Proposition 3 means that if we know that the value of MMSE at snr_0 is given by $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) = \frac{\beta}{1+\beta \mathsf{snr}_0}$ then for all higher SNR values ($\mathsf{snr}_0 \leq \mathsf{snr}$) we have the upper bound in (10a) and for all lower SNR values ($\mathsf{snr} \leq \mathsf{snr}_0$) we have the lower bound in (10b). Unfortunately, Proposition 3 does not provide an upper bound on $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ for $\mathsf{snr} \in [0, \mathsf{snr}_0)$ and one of the goals of this paper is to fill this gap. Note that upper bounds on the MMSE are useful, thanks to the I-MMSE relationship, as tools to derive converse results, and have been used in [20], [18], [19], and [21] to name a few.

Motivated by the search for the complementary upper bound to the SCPP we define the following problem.

Definition 2. *(*max-MMSE problem.*) For some* $\beta \in [0, 1]$

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{11a}$$

$$\textit{s.t.}\ \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \leq 1, \tag{11b}$$

$$\textit{and}\ \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta \mathsf{snr}_0}. \tag{11c}$$

Clearly, $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \leq \mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ for all finite n. Observe that the max-MMSE problem in (11) and the max-I problem in (5) have different objective functions but have the same constraints. This is also a good place to point out that neither of the max-MMSE and max-I problems falls under the category of convex optimization. This follows from the fact that the MMSE is a strictly concave function in the input distribution [22]. Therefore, the set of input distributions, defined by (11b) and (11c), over which we are optimizing, might not be convex.

Note that Proposition 3 gives a solution to the max-MMSE problem in (11) for $\mathsf{snr} \geq \mathsf{snr}_0$ and any $n \geq 1$ as follows:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta \mathsf{snr}},\ \text{ for } \mathsf{snr} \geq \mathsf{snr}_0, \tag{12}$$

achieved by $\mathbf{X} \sim \mathcal{N}(0, \beta\mathbf{I})$. Therefore in the rest of the paper the treatment of the max-MMSE problem will focus only on the regime $\mathsf{snr} \le \mathsf{snr}_0$.

The case $n \to \infty$ of the max-MMSE problem in (11) was solved for random codes using statistical physics in [23, Section V-C] and generalized in [12, Theorem 2] as follows:

$$\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} < \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr} \ge \mathsf{snr}_0, \end{cases}, \tag{13}$$

achieved by using superposition coding with Gaussian codebooks. For other recent links between random codes, the MMSE and statistical physics see [24].

Clearly there is a discontinuity in (13) 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* [23]. It is also well known that, for any finite n, $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ is a continuous function of snr [18]. 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:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} \le \mathsf{snr}_L, \\ T_n(\mathsf{snr}, \mathsf{snr}_0, \beta), & \mathsf{snr}_L \le \mathsf{snr} \le \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \le \mathsf{snr}, \end{cases} \tag{14}$$

for some snr_L. In this paper we seek to characterize snr_L in (14) and the continuous function $T_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ such that

$$T_n(\mathsf{snr}_L, \mathsf{snr}_0, \beta) = \frac{1}{1+\mathsf{snr}_L}, \tag{15a}$$

$$T_n(\mathsf{snr}_0, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta\mathsf{snr}_0}, \tag{15b}$$

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

$$W_n := \mathsf{snr}_0 - \mathsf{snr}_L. \tag{16}$$

Back to the max-I problem in (5). Clearly $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ is a non-decreasing function of n.

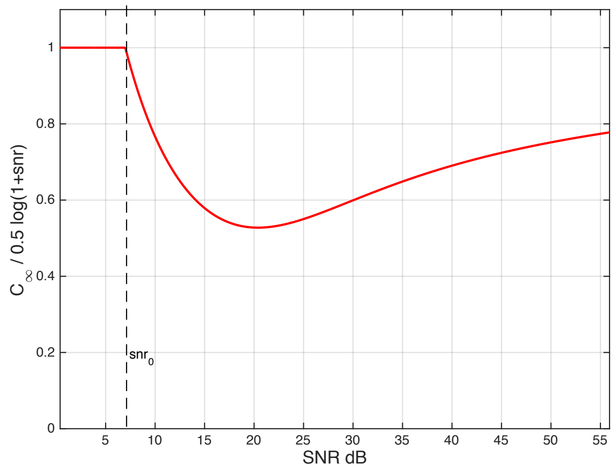


Fig. 3: Plot of $\frac{\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta)}{\frac{1}{2}\log(1+\mathsf{snr})}$ vs. snrdB, for $\beta = 0.01$, $\mathsf{snr}_0 = 5 = 6.989$ dB.

In [12, Theorem. 3] it was shown that

$$\begin{aligned}\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta) &= \lim_{n\to\infty} \mathcal{C}_n(\mathsf{snr},\mathsf{snr}_0,\beta), \\ &= \begin{cases} \frac{1}{2}\log(1+\mathsf{snr}), & \mathsf{snr} \le \mathsf{snr}_0, \\ \frac{1}{2}\log(1+\beta\mathsf{snr}) + \frac{1}{2}\log\left(1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}\right), & \mathsf{snr} \ge \mathsf{snr}_0, \end{cases} \\ &= \frac{1}{2}\log^+\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(1+\min(\mathsf{snr},\mathsf{snr}_0)\right), \end{aligned} \tag{17}$$

which is achieved by using superposition coding with Gaussian codebooks. Fig. 3 shows a plot of $\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta)$ normalized by the capacity of the point-to-point channel $\frac{1}{2}\log(1+\mathsf{snr})$. The region $\mathsf{snr} \le \mathsf{snr}_0$ (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 $\mathsf{snr} \ge \mathsf{snr}_0$ (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 can not be fully decoded (recall that in this regime superposition coding appears to be the best achievable strategy for the G-IC, but it is unknown whether it achieves capacity [7]).

The importance of studying models of communication systems with disturbance constraints has been recognized previously. For example, in [25] Bandemer *et al.* studied the following problem related to the max-I problem in (5).

Definition 3. *(*Bandemer *et al.* problem.*) For some* $R \geq 0$

$$\mathcal{I}_n(\mathsf{snr}, \mathsf{snr}_0, R) := \max_{\mathbf{X}} I_n(\mathbf{X}, \mathsf{snr}), \tag{18a}$$

$$\textit{s.t. } \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \leq 1, \tag{18b}$$

$$\textit{and } I_n(\mathbf{X}, \mathsf{snr}_0) \leq R. \tag{18c}$$

In [25] it was shown that the optimal solution for $\mathcal{I}_n(\mathsf{snr}, \mathsf{snr}_0, R)$, for any n, is attained by $\mathbf{X} \sim \mathcal{N}(0, \alpha \mathbf{I})$ where $\alpha = \min\left(1, \frac{\mathrm{e}^{2R}-1}{\mathsf{snr}_0}\right)$; here α is such that the most stringent constraint between (18b) and (18c) is satisfied with equality. In other words, the optimal input is i.i.d. Gaussian with power reduced such that the disturbance constraint in (18c) is not violated.

Observe that the max-I problem in (5) and the one in (18) have the same objective function but have different constraints. The relationship between the constraints in (5c) and (18c) can be explained as follows. The constraint in (5c) imposes a maximum value on the function $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ at $\mathsf{snr} = \mathsf{snr}_0$, while the constraint in (18c), via the integral I-MMSE relationship in (8), imposes a constraint on the area below the function $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ in the range $\mathsf{snr} \in [0, \mathsf{snr}_0]$.

Measuring the disturbance with the mutual information as in (18), in contrast to the MMSE as in (5), suggests that it is always optimal to use Gaussian codebooks with the reduced power without any rate splitting. Moreover, while the mutual information constraint in (18) limits the amount of information transmitted to the unintended receiver, it may not be the best choice when one models the interference, since any information that can be reliably decoded is not really interference. For this reason, it has been argued in [12] and [13] that the max-I problem in (5) with the MMSE disturbance constraint is a more suitable building block to study the G-IC and understand the key role of rate splitting.

A. Contributions and Paper Outline

The main contributions of the paper are as follows. In Section III we summarize our main results:

- Theorem 1, our main technical result, provides new upper bounds for the max-MMSE problem for arbitrary n that complement the SCPP bound.

- Proposition 4 provides a lower bound on the width of the phase transition region of the order of $\frac{1}{n}$.
- Proposition 5 provides a new upper bound for the max-I problem for arbitrary n.
- Proposition 8 shows that, for the case of $n = 1$, superposition of discrete and Gaussian inputs, termed *mixed input* inputs in [11], achieves the proposed upper bound on the max-I problem from Proposition 5 to within an additive gap of order $\log\log \frac{1}{\mathrm{mmse}(X,\mathsf{snr}_0)}$.

In Section IV we develop bounds on the derivative of MMSE, which we use to prove Theorem 1:

- Proposition 9 considerably refines existing bounds on the derivative of MMSE for $n = 1$ and generalizes them to any n.
- In Section IV-A, by using Proposition 9, we prove Theorem 1.

In Section V we explore whether the MMSE constraint implies a power constraint:

- Proposition 12 demonstrates that there exists an input distribution that can transmit at full power while satisfying any MMSE constraint.
- Proposition 14 develops new bounds on the MMSE under the assumption that the derivative of the MMSE exists at $\mathsf{snr} = 0^+$. This assumption is also a necessary and sufficient condition for the MMSE constraint to imply a power constraint.

Most proofs can be found in the Appendix.

III. Main Results

A. *max-MMSE problem: upper bounds on* $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$

We start by giving bounds on the phase transition region of $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ defined in (14). The bound in Theorem 1 is referred to as the D-bound because it was derived through the technique of bounding the derivative of the MMSE.

Theorem 1. *(D-Bound.) For any* $\mathbf{X}$ *and* $0 < \mathsf{snr} \leq \mathsf{snr}_0$*, let* $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta\mathsf{snr}_0}$ *for some* $\beta \in [0, 1]$*. Then*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \leq \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) + k_n \left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0} \right) - \Delta, \tag{19a}$$

$$k_n \leq n + 2, \ \Delta = 0. \tag{19b}$$

If $\mathbf{X}$ *is such that* $\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \le 1$ *then*

$$\Delta := \Delta_{(19c)} = \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{1}{\gamma^2(1+\gamma)^2} d\gamma.$$
$$= 2\log\left(\frac{1+\mathsf{snr}_0}{1+\mathsf{snr}}\right) - 2\log\left(\frac{\mathsf{snr}_0}{\mathsf{snr}}\right) + \frac{1}{1+\mathsf{snr}} - \frac{1}{1+\mathsf{snr}_0} + \frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}. \tag{19c}$$

Proof: See Section IV-A. ■

The bound on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (19a) is depicted in Fig. 4a, where:

- the red solid line is the $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ upper bound on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$, and
- the blue dashed-dotted line is the new upper bound on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ from Theorem 1.

Observe that the new bound provides a tighter and continuous upper bound on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ than the trivial upper bound given by $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$.

We next show how fast the phase transition region shrinks with n as $n \to \infty$.

Proposition 4. *The bound in* (19a)*, with* $\Delta = 0$*, from Theorem 1 intersects the LMMSE bound in* (9a) *from Proposition 2 at*

$$\mathsf{snr}_L = \mathsf{snr}_0 \frac{1+\beta\mathsf{snr}_0}{\frac{k_n}{k_n-1} + \beta\mathsf{snr}_0} = O\left(\left(1-\frac{1}{n}\right)\mathsf{snr}_0\right). \tag{20a}$$

Thus, the width of the phase transition region is given, for k_n *in* (19b)*, by*

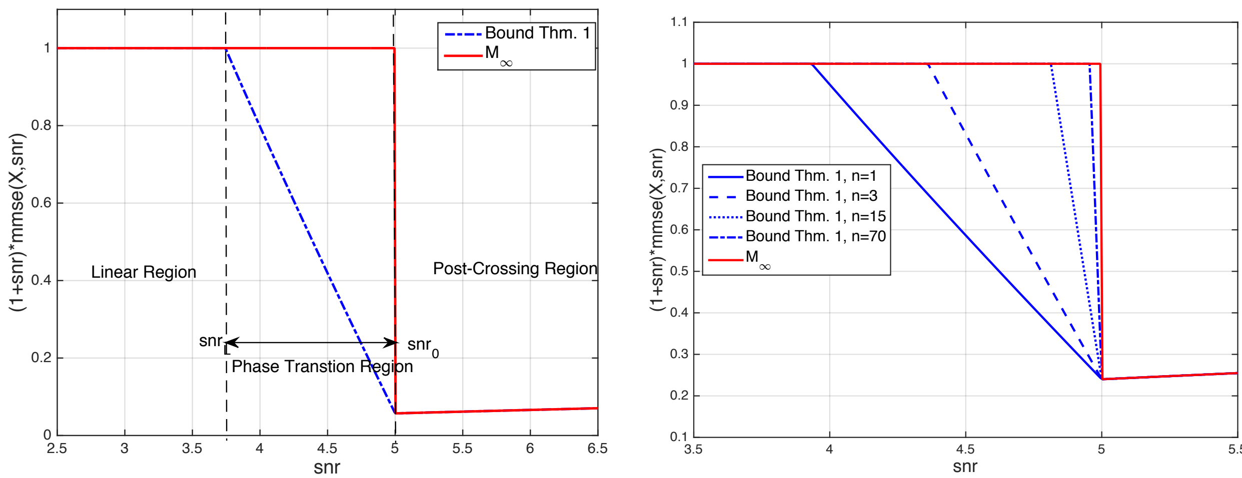
$$W_n = \frac{1}{k_n-1} \frac{\mathsf{snr}_0}{\frac{k_n}{k_n-1} + \beta\mathsf{snr}_0} = O\left(\frac{1}{n}\right). \tag{20b}$$

Proof: See Appendix A. ■

In Proposition 4 we found the intersection between the LMMSE bound $\frac{1}{\mathsf{snr}}$ in (9a) and the bound in (19a) from Theorem 1. Unfortunately, for the power constraint case, the intersection of the LMMSE bound $\frac{1}{1+\mathsf{snr}}$ in (9b) and the bound in (19c) cannot be found analytically. However, the solution can be computed efficiently by using numerical methods. Moreover, the asymptotic behavior of the phase transition region is still given by $O\left(\frac{1}{n}\right)$. The bound in Theorem 1 for several values of n is shown in Fig. 4b, where:

- the red line is the $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ bound on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$, and
- the blue line is the bound on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ from Theorem 1 for $n = 1, 3, 15$ and 70.

We observe that the new bound provides a refined characterization of the phase transition phenomenon for finite n and, in particular, it recovers the bound in (13) as $n \to \infty$.



(a) For $n = 1$, $\mathsf{snr}_0 = 5$ and $\beta = 0.01$.

(b) For several values of n, $\mathsf{snr}_0 = 5$ and $\beta = 0.05$.

Fig. 4: Bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr.

B. max-I problem: upper bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$

Using the previous novel bound on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in Theorem 1 we can find new upper bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ by integration as follows:

$$\begin{aligned} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) &\leq \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{M}_n(t, \mathsf{snr}_0, \beta) dt \\ &= \frac{1}{2}\log(1+\mathsf{snr}_L) + \frac{1}{2}\int_{\mathsf{snr}_L}^{\mathsf{snr}_0} T_n(t, \mathsf{snr}_0, \beta) dt + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right), \text{ for } \mathsf{snr}_0 \leq \mathsf{snr}, \end{aligned} \tag{21}$$

and

$$\begin{aligned} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) &\leq \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{M}_n(t, \mathsf{snr}_0, \beta) dt \\ &\leq \frac{1}{2}\log(1+\min(\mathsf{snr}_L, \mathsf{snr})) + \frac{1}{2}\int_{\min(\mathsf{snr}_L, \mathsf{snr})}^{\mathsf{snr}} T_n(t, \mathsf{snr}_0, \beta) dt, \text{ for } \mathsf{snr}_0 \geq \mathsf{snr}. \end{aligned} \tag{22}$$

By using Theorem 1 (with finite power assumption) to bound $T_n(t, \mathsf{snr}_0, \beta)$ we get the following upper bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$.

Proposition 5. *For any $0 \leq \mathsf{snr}_0$, $\beta \in [0,1]$, and snr_L given in Proposition 4, we have that for $\mathsf{snr}_0 \leq \mathsf{snr}$*

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \leq \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) - \Delta_{(25)}, \tag{23}$$

and for $\mathsf{snr}_0 \geq \mathsf{snr}$

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \leq \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) - \Delta_{(26)}, \tag{24}$$

where

$$0 \leq \Delta_{(25)} = \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_0}{1+\mathsf{snr}_L}\right) - \frac{1}{2}\frac{\beta(\mathsf{snr}_0 - \mathsf{snr}_L)}{1+\beta\mathsf{snr}_0} - \frac{(n+2)}{2}\log\left(\frac{\mathsf{snr}_0}{\mathsf{snr}_L}\right) + \frac{(n+2)(\mathsf{snr}_0 - \mathsf{snr}_L)}{2\mathsf{snr}_0}$$
$$+ \frac{1}{2}\left((2\mathsf{snr}_L + 1)\log\left(\frac{\mathsf{snr}_0(1+\mathsf{snr}_L)}{\mathsf{snr}_L(1+\mathsf{snr}_0)}\right) - \frac{\mathsf{snr}_0 - \mathsf{snr}_L}{1+\mathsf{snr}_0} - \frac{\mathsf{snr}_0 - \mathsf{snr}_L}{\mathsf{snr}_0}\right) = O\left(\frac{1}{n}\right), \tag{25}$$

and

$$0 \leq \Delta_{(26)} = \frac{1}{2}\log\left(\frac{1+\mathsf{snr}}{1+\min(\mathsf{snr}_L, \mathsf{snr})}\right) - \frac{\beta(\mathsf{snr} - \min(\mathsf{snr}_L, \mathsf{snr}))}{2(1+\beta\mathsf{snr}_0)}$$
$$- \frac{(n+2)}{2}\log\left(\frac{\mathsf{snr}}{\min(\mathsf{snr}_L, \mathsf{snr})}\right) + \frac{(n+2)(\mathsf{snr} - \min(\mathsf{snr}_L, \mathsf{snr}))}{2\mathsf{snr}_0}$$
$$+ \frac{1}{2}\Bigg((2\min(\mathsf{snr}_L, \mathsf{snr}) + 1)\log\left(\frac{1+\min(\mathsf{snr}_L, \mathsf{snr})}{\min(\mathsf{snr}_L, \mathsf{snr})}\right) - (2\mathsf{snr}+1)\log\left(\frac{1+\mathsf{snr}}{\mathsf{snr}}\right)$$
$$+ 2(\mathsf{snr} - \min(\mathsf{snr}_L, \mathsf{snr}))\log\left(\frac{1+\mathsf{snr}_0}{\mathsf{snr}_0}\right) - \frac{\mathsf{snr} - \min(\mathsf{snr}_L, \mathsf{snr})}{\mathsf{snr}_0} - \frac{\mathsf{snr} - \min(\mathsf{snr}_L, \mathsf{snr})}{1+\mathsf{snr}_0}\Bigg)$$
$$= O\left(\frac{1}{n}\right). \tag{26}$$

Fig. 5 compares the bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (17) from Proposition 5 with $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ for several values of n. The figure shows how the new bounds in Proposition 5 improve on the trivial $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ bound for finite n.

C. max-MMSE problem: achievability of $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$

In this section we propose an input that will be used in the achievable strategy for both the max-I problem and the max-MMSE problem with input length $n = 1$. This input is referred to as *mixed input* [11] and is defined as

$$X_{\text{mix}} := \sqrt{1-\delta}X_D + \sqrt{\delta}X_G, \ \ \delta \in [0, 1], \tag{27}$$

where X_G and X_D are independent, $X_G \sim \mathcal{N}(0, 1)$, $\mathbb{E}[X_D^2] \leq 1$, and where the distribution of X_D and the parameter δ are to be optimized over. The input X_{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 X_D and X_G components, albeit at different SNR values.

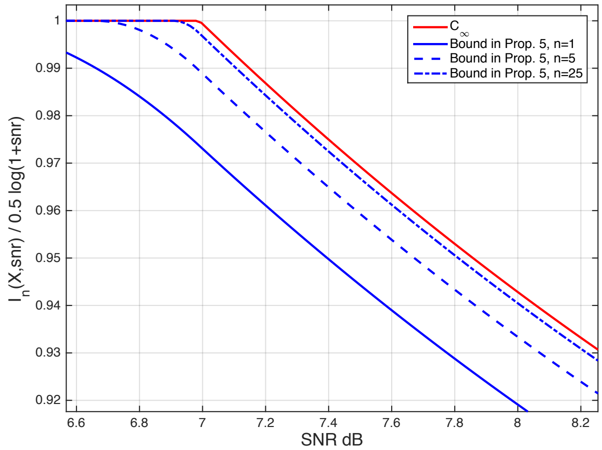


Fig. 5: Bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.1$ and $\mathsf{snr}_0 = 5 = 6.9897$ dB.

Proposition 6. *For X_{mix} defined in* (27) *we have that*

$$I(X_{\text{mix}}, \mathsf{snr}) = I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + I(X_G, \mathsf{snr}\ \delta), \tag{28a}$$

$$\text{mmse}(X_{\text{mix}}, \mathsf{snr}) = \frac{1-\delta}{(1+\mathsf{snr}\delta)^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \delta\ \text{mmse}(X_G, \mathsf{snr}\ \delta). \tag{28b}$$

Proof: See Appendix B. ∎

Observe that Proposition 6 implies that, in order for mixed inputs (with $\delta < 1$) to comply with the MMSE constraint in (5c) and (11c), the MMSE of X_D must satisfy

$$\text{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) \leq \frac{(\beta-\delta)(1+\delta\mathsf{snr}_0)}{(1-\delta)(1+\beta\mathsf{snr}_0)}. \tag{29}$$

The bound in (29) will be helpful in choosing the parameter δ later on.

When X_D is a discrete random variable with $\mathsf{supp}(X_D) = N$ we use the following bounds from [26, App. C] and [11, Rem. 2].

Proposition 7. *([26], [11]) For a discrete random variable X_D such that $p_i = \Pr(X_D = x_i)$, for $i \in [1 : N]$, we have that*

$$\text{mmse}(X_D, \mathsf{snr}) \leq d_{\max}^2 \sum_{i=1}^{N} p_i \mathrm{e}^{-\frac{\mathsf{snr}}{8} d_i^2}, \tag{30a}$$

$$I(X_D, \mathsf{snr}) \geq H(X_D) - \frac{1}{2}\log\left(\frac{\pi}{6}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\text{mmse}(X_D, \mathsf{snr})\right), \tag{30b}$$

where

$$d_\ell := \min_{x_i \in \mathsf{supp}(X_D): i \neq \ell} |x_\ell - x_i|, \tag{30c}$$

$$d_{\min} := \min_{\ell \in [1:N]} d_\ell, \tag{30d}$$

$$d_{\max} := \max_{x_k, x_i \in \mathsf{supp}(X_D)} |x_k - x_i|. \tag{30e}$$

Proposition 6 and Proposition 7 are particularly useful because they will allow us to design Gaussian and discrete components of the mixed input independently.

Fig. 6 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 (13) (solid red line) ;
- The upper bound from Theorem 1 with finite power (dashed cyan line);
- The Gaussian-only input lower bound (green line), with $X \sim \mathcal{N}(0, \beta)$, where the power has been reduced to meet the MMSE constraint;
- The mixed input lower bound (blue dashed line), with the input in (27). We used Proposition 6 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 (29). That is, we chose δ such that the power of X_G is approximately β while the MMSE constraint on X_D in (29) is not equal to zero. The input X_D used in Fig. 6 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 lower bound (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 lower bound (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, Fig. 6 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, as seen in (30a)) 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 in (30a). This further motivates determining whether the MMSE constraint can imply a power constraint, which we shall investigate in Section V. 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, Fig. 6 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 (29). Note that the difference between Discrete 1 and Discrete 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 the Discrete 2 was found as an optimal discrete input without a Gaussian component (i.e., $\delta = 0$).

The insight gained from analyzing different lower bounds on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ will be crucial to show an approximately optimal input for $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$, which we consider next.

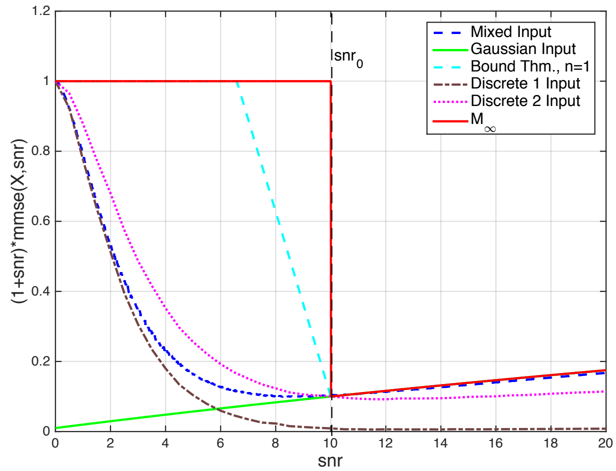


Fig. 6: Upper and lower bounds on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.01$, $\mathsf{snr}_0 = 10$.

TABLE I: Parameters of the mixed input in (27) used in the proof of Proposition 8.

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 ($\mathsf{snr} \leq \mathsf{snr}_0$)	$N = \left\lfloor \sqrt{1 + c_2 \mathsf{snr}} \right\rfloor$, $c_2 = \frac{3}{2 \log\left(\frac{12(1+\beta \mathsf{snr}_0)}{\beta} \right)}$, $\delta = 0$.

D. max-I problem: achievability of $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$

In this section we demonstrate that an inner bound on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ with the mixed input in (27) is to within an additive gap of the outer bound in Proposition 5.

Proposition 8. *A lower bound on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ with the mixed input in (27), with $X_D \sim \mathsf{PAM}(N)$ and with input parameters as specified in Table I, is to within $O\left(\log \log (\frac{1}{\mathrm{mmse}(X, \mathsf{snr}_0)}) \right)$ of the outer bound in Proposition 5 with the exact gap value given by*

$$\mathsf{snr} \geq \mathsf{snr}_0 \geq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\mathrm{mix}}, \mathsf{snr}) := \mathsf{gap}_1, \tag{31a}$$

$$\mathsf{snr}_0 \geq \mathsf{snr} \geq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\mathrm{mix}}, \mathsf{snr}) := \mathsf{gap}_2, \tag{31b}$$

$$\mathsf{snr} \leq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\mathrm{mix}}, \mathsf{snr}) := \mathsf{gap}_3, \tag{31c}$$

where

$$\mathsf{gap}_1 \le \frac{1}{2}\log\left(\frac{2}{3}\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0)}{\beta}\right)+\frac{6\beta}{1+\beta\mathsf{snr}_0}\right)+\frac{1}{2}\log\left(\frac{4\pi}{3}\right)-\Delta_{(25)}, \quad (31d)$$

$$\mathsf{gap}_2 \le \frac{1}{2}\log\left(1+\frac{2}{3}\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)\right)+\frac{1}{2}\log\left(\frac{4\pi}{6}\right)-\Delta_{(26)}, \quad (31e)$$

$$\mathsf{gap}_3 \le \frac{1}{2}\log(2). \quad (31f)$$

and $\Delta_{(25)}$ *and* $\Delta_{(26)}$ *are given in* (25) *and* (26)*, respectively.*

Proof: See Appendix C. ■

Please note that the gap result in Proposition 8 is constant in snr (i.e., independent of snr) but not in snr_0.

Fig. 7 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 8 to:

- The $\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta)$ upper bound in (17), (solid red line);
- The upper bound from Proposition 5 (dashed blue line); and
- The inner bound with $X \sim \mathcal{N}(0,\beta)$, where the reduction in power is necessary to satisfy the MMSE constraint $\mathrm{mmse}(X,\mathsf{snr}_0) \le \frac{\beta}{1+\beta\mathsf{snr}_0}$ (dotted green line).

Fig. 7 shows that Gaussian inputs are sub-optimal and that mixed inputs achieve large degrees of freedom compared to Gaussian inputs. Interestingly, in the regime $\mathsf{snr} \le \mathsf{snr}_0$, 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 [12] that discrete inputs may be optimal for $n = 1$ and $\mathsf{snr} \le \mathsf{snr}_0$. For the case $\mathsf{snr} \ge \mathsf{snr}_0$ our result partially refute the conjecture by excluding the possibility of discrete inputs with finitely many points from being optimal.

The above discussion completes the presentation of our bounds on max-I and max-MMSE problems. The remainder of the paper contains the proof of Theorem 1 and a discussion of when the MMSE constraint necessarily implies a power constraint.

IV. Properties of the first derivative of MMSE

A key element in the proof of the SCPP in Proposition 3 was the characterization of the first derivative of the MMSE as

$$-\frac{d\mathrm{mmse}(\mathbf{X},\mathsf{snr})}{d\mathsf{snr}} = \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) := \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X},\mathsf{snr})\right]\right), \quad (32)$$

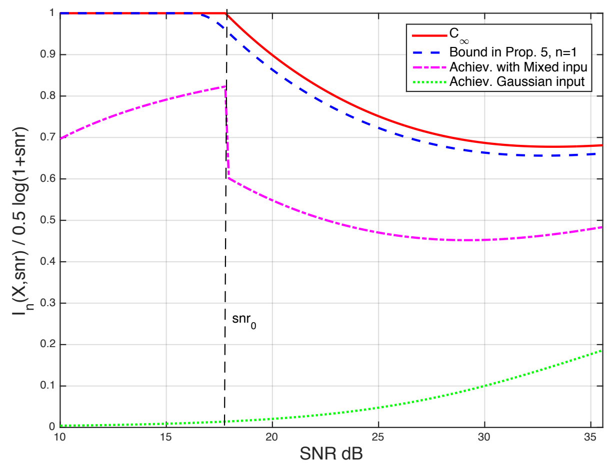


Fig. 7: Upper and lower bounds on $\mathcal{C}_{n=1}(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.001$ and $\mathsf{snr}_0 = 60 = 17.6815$ dB.

which was given in [18, Proposition 9] for $n = 1$ and in [19, Lemma 3] for $n \geq 1$. The first derivative in (32) turns out to be instrumental in proving Theorem 1 as well.

For ease of presentation, in the rest of the section, instead of focusing on the derivative we will focus on $\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})]\right)$. The quantity $\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})]\right)$ is well defined for any $\mathbf{X}$. Moreover, for the case of $n = 1$ it has been shown [18, Proposition 5] that

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \leq \frac{k_1}{\mathsf{snr}^2}, \text{ where } k_1 \leq 3 \cdot 2^4. \tag{33}$$

Before using (32) in the proof of Theorem 1, we will need to sharpen the existing constant for $n = 1$ in (33) (given by $k_1 \leq 3 \cdot 2^4$) and generalize the bound to any $n \geq 1$, which to the best of our knowledge has not been considered before.

Proposition 9. *For any* $\mathbf{X}$ *and* $\mathsf{snr} > 0$ *we have*

$$\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})]\right) \leq \frac{k_n}{\mathsf{snr}^2}, \tag{34a}$$

where

$$k_n \leq \frac{n(n+2) - n \ \mathrm{mmse}(\mathbf{Z}\mathbf{Z}^T|\mathbf{Y}) - \mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right)}{n} \leq n + 2. \tag{34b}$$

Proof: See Appendix D. ■

In Proposition 9 the bound on k_1 in (33) has been tightened from $k_1 \le 3 \cdot 2^4$ in (33) to $k_1 \le 3$. This improvement will result in tighter bounds in what follows.

The following tightens k_n for power constrained inputs.

Proposition 10. *If* $\mathbf{X}$ *is such that* $\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right]\right) \le 1$*, then*

$$\mathrm{Tr}(\mathbf{J}^2(\mathbf{Y})) \ge \frac{n}{(1+\mathsf{snr})^2}. \tag{35}$$

Equality in (35) *is achieved when* $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Proof: See Appendix E. ∎

Observe that, by using the bound in (34) from Proposition 9 together with the lower bound on the Fisher information in Proposition 10, the bound on the constant k_n in (34b) can be tightened to

$$k_n \le \frac{n(n+2) - \frac{n}{(1+\mathsf{snr})^2}}{n} = n + 2 - \frac{1}{(1+\mathsf{snr})^2}. \tag{36}$$

By further assuming that $\mathbf{X}$ has a finite fourth moment we can arrive at the following bound that does not blow up around $\mathsf{snr} = 0^+$, as opposed to the bound in (34a).

Proposition 11. *If* $\mathbf{X}$ *such that* $\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\left(\mathbf{X}\mathbf{X}^T\right)^2\right]\right) < \infty$ *then*

$$\begin{aligned} &\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})]\right) \\ &\le \min\left(\frac{\mathrm{Tr}\left(\mathbb{E}\left[\left(\left(\mathbf{X} - \sqrt{\mathsf{snr}}\mathbf{Z}\right)\left(\mathbf{X} - \sqrt{\mathsf{snr}}\mathbf{Z}\right)^T\right)^2\right]\right)}{(1+\mathsf{snr})^4}, \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[\mathbf{X}\mathbf{X}^T|\mathbf{Y}\right]\right]\right)\right), \end{aligned} \tag{37a}$$

where we can further bound

$$\mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[\mathbf{X}\mathbf{X}^T|\mathbf{Y}\right]\right]\right) \le \mathrm{Tr}\left(\mathbb{E}\left[\left(\mathbf{X}\mathbf{X}^T\right)^2\right]\right). \tag{37b}$$

Proof: See Appendix F. ∎

Note that evaluation of the first term of the minimum in (37a) requires only the knowledge of second and fourth moments of $\mathbf{X}$.

We are now ready to prove our main result.

A. *Proof of Theorem 1*

The proof of Theorem 1 relies on the fact that the MMSE is an infinitely differentiable function of snr [18, Proposition 7] and therefore can be written as the difference of two MMSE functions using the fundamental theorem of calculus

$$\begin{aligned}
&\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) - \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0)\\
&= -\int_{\mathsf{snr}}^{\mathsf{snr}_0} \mathrm{mmse}'(\mathbf{X}, \gamma) d\gamma\\
&\overset{a)}{=} \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{1}{n} \mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}, \gamma)]\right) d\gamma\\
&\overset{b)}{\leq} \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{(n+2)}{\gamma^2} d\gamma = (n+2)\left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}\right) - \Delta,\ \Delta = 0,
\end{aligned}$$

where the (in)-equalities follow by using: a) (32), and b) the bound in Proposition 9 with $k_n \leq n+2$. If we further assume that $\mathbf{X}$ has finite power, instead of bounding $k_n \leq n+2$, we can use (36), to obtain

$$0 \leq \Delta = \Delta_{(19c)} = \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{1}{\gamma^2 (1+\gamma)^2} d\gamma.$$

This concludes the proof of Theorem 1.

V. When does an MMSE constraint imply a power constraint

In this section we try to determine whether the MMSE constraint may imply a power constraint. For simplicity we focus on the case of $n = 1$. This question is motivated by the following limit, which exists iff $\mathbb{E}[X^2] < \infty$:

$$\lim_{\mathsf{snr} \to 0^+} \mathrm{mmse}(X, \mathsf{snr}) = \mathbb{E}[X^2]. \tag{38}$$

The limit in (38) raises the question of whether the MMSE constraint at snr_0 around zero would imply a power constraint. In other words, are we required to reduce power to meet the MMSE constraint for very small snr_0? Surprisingly, the answer to this question is no.

Proposition 12. *There exists an input distribution X with maximum power as in* (5b) *that satisfies the MMSE constraint in* (5c) *for any $\mathsf{snr}_0 > 0$ and any $\beta > 0$.*

Proof: Consider an input distribution given by

$$X_a = [-a, 0, a], \quad P_{X_a} = \left[\frac{1}{2a^2}, 1 - \frac{1}{a^2}, \frac{1}{2a^2}\right], \tag{39}$$

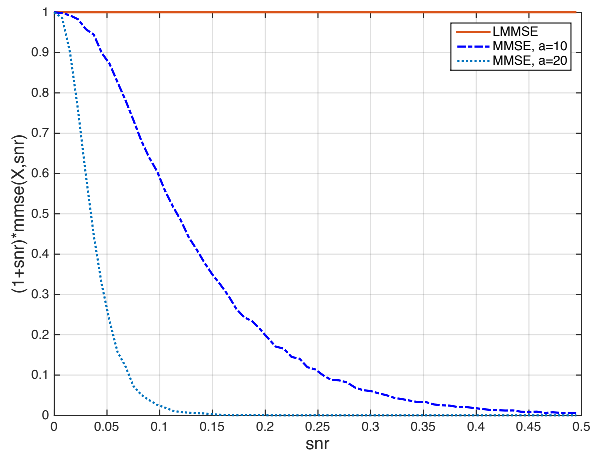


Fig. 8: $\mathrm{mmse}(X_a, \mathsf{snr})$ vs. snr, for $a = 10$ and $a = 20$.

for any $a \geq 1$. Note that for the input distribution in (39) $\mathbb{E}[X_a^2] = 1$ for any a. The MMSE of X_a can be upper bounded by

$$\mathrm{mmse}(X_a, \mathsf{snr}) \leq \min\left(1, 4(a^2+1)e^{-\frac{a^2\mathsf{snr}}{8}}\right), \tag{40}$$

where the upper bound in (40) follows by applying the upper bound in Proposition 7 together with the bound $\mathrm{mmse}(X_a, \mathsf{snr}) \leq \mathbb{E}[X_a^2] = 1$. Therefore, by choosing a large enough, any MMSE constraint can be met while transmitting at full power. This concludes the proof. ■

The MMSE of X_a is shown and compared to the LMMSE in Fig. 8. Here are some other properties of X_a that are easy to verify.

Proposition 13. *The random variable X_a has the following properties*

- $\lim_{a\to\infty} X_a = 0$ *almost surely (a.s.),*
- $\mathbb{E}[|X_a - 0|^n] = a^n p = E[X_a^2]a^{n-2} = a^{n-2}$.

The random variable X_a serves as a counterexample that shows that a.s. convergence does not imply L^p convergence.

An interesting question is whether we can characterize a family of input distributions for which the MMSE constraint implies a power constraint under some non-trivial condition. In other words, we want to find a family of input distributions such that the power constraint can be related to the MMSE constraint at some snr_0, that is

$$\mathbb{E}[X^2] = f(\text{mmse}(X, \mathsf{snr}_0)) \leq 1. \tag{41}$$

Towards this end we have the following:

Proposition 14. *For any X and any $\mathsf{snr}_0 \geq \mathsf{snr} > 0$, we have that*

$$\text{mmse}(X, \mathsf{snr}) = \text{mmse}(X, \mathsf{snr}_0) + k \cdot (\mathsf{snr}_0 - \mathsf{snr}), \tag{42}$$

where k is defined by some $\mathsf{snr}_c \in (\mathsf{snr}, \mathsf{snr}_0]$ as follows:

$$k = \mathbb{E}\left[\text{Cov}^2(X, \mathsf{snr}_c)\right] \leq \sup_{\gamma \in (\mathsf{snr}, \mathsf{snr}_0)} \mathbb{E}[\text{Cov}^2(X, \gamma)] \leq \mathbb{E}[X^4]. \tag{43}$$

Moreover, for $\mathsf{snr} = 0^+$ the equality in (42) is valid iff

$$\lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\text{Cov}^2(X, \mathsf{snr})] < \infty. \tag{44}$$

Proof: The result easily follows by applying the mean value theorem

$$\begin{aligned} \text{mmse}(X, \mathsf{snr}) - \text{mmse}(X, \mathsf{snr}_0) &= \int_{\mathsf{snr}}^{\mathsf{snr}_0} \mathbb{E}[\text{Cov}^2(X, \gamma)] d\gamma \\ &= \mathbb{E}[\text{Cov}^2(X, \mathsf{snr}_c)](\mathsf{snr}_0 - \mathsf{snr}). \end{aligned} \tag{45}$$

for some $\mathsf{snr}_c \in (\mathsf{snr}, \mathsf{snr}_0)$. Note that for $\mathsf{snr} > 0$ the quantity $\mathbb{E}[\text{Cov}^2(X, \gamma)]$ is finite due to Proposition 9. Therefore, we focus on the case when $\mathsf{snr} = 0^+$.

Therefore, if $\lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\text{Cov}^2(X, \mathsf{snr})] = K < \infty$ for some $K > 0$, by Jensen's inequality we have that

$$K = \lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\text{Cov}^2(X, \mathsf{snr})] \geq (\mathbb{E}[X^2])^2 = (\text{mmse}(X, 0))^2. \tag{46}$$

So, in other words the existence of the derivative at $\mathsf{snr} = 0^+$ implies the existence of the power constraint and the integration in (45) holds for $\mathsf{snr} = 0^+$.

Conversely, if the integration in (45) is finite for $\mathsf{snr} = 0^+$ we have that $\lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\text{Cov}^2(X, \mathsf{snr})] < \infty$.

Therefore, the bound in (42) holds iff $\lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\text{Cov}^2(X, \mathsf{snr})] < \infty$. This concludes the proof. ■

From Proposition 14 we see that necessary and sufficient conditions for the MMSE at snr_0 to imply a reduction in power (i.e., $\mathbb{E}[X^2] < 1$) are

$$1)\ \mathrm{mmse}(X, \mathsf{snr}_0) + \mathsf{snr}_0 \cdot \mathbb{E}[\mathrm{Cov}^2(X, \mathsf{snr}_c)] < 1,$$

$$\Leftrightarrow \mathbb{E}[\mathrm{Cov}^2(X, \mathsf{snr}_c)] < \frac{1 - \mathrm{mmse}(X, \mathsf{snr}_0)}{\mathsf{snr}_0}, \tag{47a}$$

$$2)\ \lim_{\mathsf{snr} \to 0^+} \mathbb{E}[\mathrm{Cov}^2(X, \mathsf{snr})] < \infty, \tag{47b}$$

where snr_c is defined in Proposition 14.

Since snr_c might be difficult to compute, the following slightly stronger (i.e., sufficient condition) can be useful:

$$\sup_{\gamma \in (0, \mathsf{snr}_0)} \mathbb{E}[\mathrm{Cov}^2(X, \gamma)] < \frac{1 - \mathrm{mmse}(X, \mathsf{snr}_0)}{\mathsf{snr}_0}. \tag{48}$$

Finally, observe that $\lim_{a \to \infty} X_a$ does not satisfy this moment condition since

$$\lim_{a \to \infty} \mathbb{E}[\mathrm{Cov}^2(X_a|Y)] = \begin{cases} \infty & \mathsf{snr} = 0, \\ 0 & \mathsf{snr} > 0. \end{cases} \tag{49}$$

VI. Conclusion

In this paper we have considered a Gaussian channel with one transmitter and two receivers in which the maximization of the input-output mutual information at the primary/intended receiver is subject to a disturbance constraint measured by the MMSE at the secondary/unintended receiver. We have derived new upper bounds on the input-output mutual information of this channel that hold for vector inputs of any length. For the case of scalar inputs we have demonstrated a matching lower bound that is to within an additive gap of the order $O\left(\log\log \frac{1}{\mathrm{mmse}(X, \mathsf{snr}_0)}\right)$ of the upper bound. At the heart of our proof is a new upper bound on the MMSE that complements the SCPP of the MMSE and might be of independent interest.

Appendix A
Proof of Proposition 4

In order to find the point of intersection snr_L between (9a) and (19a) we must solve the following equation:

$$\frac{1}{\mathsf{snr}} - \frac{k_n}{\mathsf{snr}} + \frac{k_n}{\mathsf{snr}_0} - \frac{\beta}{1 + \beta \mathsf{snr}_0} = 0 \Rightarrow \frac{1}{\mathsf{snr}} - \frac{k_n}{\mathsf{snr}} + A = 0$$

where $A = \frac{k_n}{\mathsf{snr}_0} - \frac{\beta}{1+\beta\mathsf{snr}_0}$ contains all quantities that do not depend on snr. By solving for snr we find that

$$\mathsf{snr}_L = \frac{k_n - 1}{A} = \frac{\mathsf{snr}_0(1+\beta\mathsf{snr}_0)(k_n-1)}{k_n + (k_n-1)\beta\mathsf{snr}_0} = \mathsf{snr}_0 \frac{1+\beta\mathsf{snr}_0}{\frac{k_n}{k_n-1} + \beta\mathsf{snr}_0},$$

and the width of the phase transition is given by

$$\mathsf{snr}_0 - \mathsf{snr}_L = \mathsf{snr}_0 \left(1 - \frac{1+\beta\mathsf{snr}_0}{\frac{k_n}{k_n-1} + \beta\mathsf{snr}_0}\right) = \frac{1}{k_n-1} \frac{\mathsf{snr}_0}{\frac{k_n}{k_n-1} + \beta\mathsf{snr}_0},$$

as claimed in (20b). This concludes the proof.

Appendix B
Proof of Proposition 6

We first show the decomposition for mutual information with mixed inputs in (27)

$$\begin{aligned} I(X_{\text{mix}}, \mathsf{snr}) &= I(X_{\text{mix}}; Y) = I(X_G, X_D; Y) \\ &= I(X_D; Y) + I(X_G; Y|X_D) \\ &= I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + I(X_G, \mathsf{snr}\delta). \end{aligned} \tag{50}$$

Next we take the derivative of both sides of (50) with respect to snr. On the left side we get $\frac{d}{d\mathsf{snr}} I(X_{\text{mix}}, \mathsf{snr}) = \frac{1}{2}\text{mmse}(X_{\text{mix}}, \mathsf{snr})$ and on the right we get

$$\begin{aligned} &\text{mmse}(X_{\text{mix}}, \mathsf{snr}) \\ &= 2\frac{d}{d\mathsf{snr}} I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + 2\frac{d}{d\mathsf{snr}} I(X_G, \mathsf{snr}\delta) \\ &= \text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) \cdot \frac{d}{d\mathsf{snr}}\left(\frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \text{mmse}(X_G, \mathsf{snr}\delta) \cdot \frac{d}{d\mathsf{snr}}(\mathsf{snr}\delta) \\ &= \frac{1-\delta}{(1+\delta\mathsf{snr})^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \text{mmse}(X_G, \mathsf{snr}\delta)\delta \\ &= \frac{1-\delta}{(1+\delta\mathsf{snr})^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \frac{\delta}{1+\delta\mathsf{snr}}, \end{aligned}$$

as claimed in (28). This concludes the proof.

APPENDIX C
PROOF OF PROPOSITION 8

By letting $X_D \sim \mathsf{PAM}(N)$, given the bound in Proposition 7 and the requirement in (29) we further constrain the MMSE of X_D to satisfy

$$\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) \le d_{\max}^2 \mathrm{e}^{-\frac{\frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}}{8} d_{\min}^2} \le \frac{(1+\mathsf{snr}_0\delta)(\beta-\delta)}{(1-\delta)(1+\beta\mathsf{snr}_0)}, \tag{51}$$

which ensures that the MMSE constraint in (5c) is met. Since, the minimum distance of PAM is given by $d_{\min}^2 = \frac{12}{N^2-1}$, solving for N we have that

$$N \le \left\lfloor \sqrt{1 + c_1 \frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}} \right\rfloor, \tag{52a}$$

$$c_1 = \frac{3}{2\log^+\left(\frac{d_{\max}^2(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)} \le \frac{3}{2\log^+\left(\frac{12(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)}, \tag{52b}$$

where the last inequality is due to the fact that for PAM

$$d_{\max}^2 = (N-1)^2 d_{\min}^2 = 12\frac{(N-1)^2}{N^2-1} = 12\frac{N-1}{N+1} \le 12. \tag{53}$$

For the case of $\mathsf{snr}_0 \le \mathsf{snr}$ we choose the number of points to satisfy (52) with equality and choose $\delta = \beta\frac{\mathsf{snr}_0}{1+\mathsf{snr}_0} := \beta c_2$.

Next we compute the gap between the outer bound in Proposition 5 with the achievable mutual information of a mixed input in Proposition 6, where $I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)$ is lower bounded by Proposition 7 we have

We obtain

$$
\begin{aligned}
&\mathsf{gap}_1 + \Delta_{(25)} \\
&= \mathcal{C}_\infty - I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) - I(X_G, \mathsf{snr}\ \delta) && (54)\\
&= C_\infty - \left(\log(N) - \frac{1}{2}\log\left(\frac{\pi}{6}\right) - \frac{1}{2}\log\left(1+\frac{12}{d_{\min}^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right)\right) - \frac{1}{2}\log(1+\delta\mathsf{snr})\\
&\overset{a)}{\leq} C_\infty - \left(\frac{1}{2}\log\left(1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}\right) - \log(2) - \frac{1}{2}\log\left(\frac{\pi}{6}\right)\right.\\
&\left.-\frac{1}{2}\log\left(1+\frac{12}{d_{\min}^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right) + \frac{1}{2}\log(1+\delta\mathsf{snr})\right)\\
&= \frac{1}{2}\log\left(\frac{1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}}{1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}}\right) + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}}\right) + \frac{1}{2}\log\left(1+\frac{12}{d_{\min}^2}\text{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right)\\
&+\frac{1}{2}\log\left(\frac{4\pi}{6}\right), && (55)
\end{aligned}
$$

where inequality in a) follows from getting an extra one bit gap from dropping the floor operation.

We next bound each term in (55) individually. The first term in (55) can be bounded as follows:

$$
\begin{aligned}
\frac{1}{2}\log\left(\frac{1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}}{1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}}\right) &= \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0-\beta c_1 c_2\mathsf{snr}_0)}\right)\\
&\overset{b)}{\leq} \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0-\beta c_1\mathsf{snr}_0)}\right)\\
&= \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+(1-\beta)c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0)}\right)\\
&\overset{c)}{\leq} \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)}{(1+(1-\beta)c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0)}\right)\\
&\overset{d)}{\leq} \frac{1}{2}\log\left(\max\left(\frac{(1+\mathsf{snr}_0)}{(1+c_1\mathsf{snr}_0)}, \frac{(1+\mathsf{snr}_0)}{(1+c_2\mathsf{snr}_0)}\right)\right)\\
&\overset{e)}{\leq} \frac{1}{2}\log\left(\max\left(\frac{1}{c_1}, 2\right)\right), && (56)
\end{aligned}
$$

where the inequalities follow from the facts: b) $c_2 = \frac{\mathsf{snr}_0}{1+\mathsf{snr}_0} \leq 1$; c) used that $\frac{1+c_2\beta\mathsf{snr}_0}{1+\beta\mathsf{snr}_0} \leq 1$ since $c_2 \leq 1$; d) the denominator term $1+(1-\beta)c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0$ achieves its minimum at either $\beta = 0$ or $\beta = 1$; and e) $\frac{(1+\mathsf{snr}_0)}{(1+c_2\mathsf{snr}_0)} \leq \frac{1}{c_2} = \frac{1+\mathsf{snr}_0}{\mathsf{snr}_0} \leq 2$ for $\mathsf{snr}_0 \geq 1$.

The second term in (55) can be bounded as follows:

$$\frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}}\right) \le \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_0}{\mathsf{snr}_0}\right) \le \frac{1}{2}\log(2), \tag{57}$$

where the inequalities follow from using $\delta = \beta\frac{\mathsf{snr}_0}{1+\mathsf{snr}_0}$ and $\frac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}} \le \frac{\beta}{\delta} = \frac{1+\mathsf{snr}_0}{\mathsf{snr}_0} \le 2$ for $\mathsf{snr} \ge \mathsf{snr}_0 \ge 1$.

The third term in (55) can be bounded as follows

$$\begin{aligned}
&\frac{1}{2}\log\left(1+\frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right)\\
&\overset{f)}{\le} \frac{1}{2}\log\left(1+\frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right)\right)\\
&\overset{g)}{\le} \frac{1}{2}\log\left(1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right)\right)\\
&\overset{h)}{\le} \frac{1}{2}\log\left(1+c_1\frac{(\beta-\delta)\mathsf{snr}_0}{1+\beta\mathsf{snr}_0}\right)\\
&\overset{i)}{\le} \frac{1}{2}\log\left(1+c_1\frac{\beta}{1+\beta\mathsf{snr}_0}\right),
\end{aligned} \tag{58}$$

where the (in)-equalities follow from: f) the fact that the MMSE is a decreasing function of SNR and $\frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}} \ge \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}$; g) using the bound on $d_{\min}^2 = \frac{12}{N^2-1}$ from (52); h) using the bound in (51); and i) using $\delta = \frac{\beta\mathsf{snr}_0}{1+\mathsf{snr}_0} \le \beta$ and therefore $(\beta-\delta)\mathsf{snr}_0 = \frac{\beta\ \mathsf{snr}_0}{1+\mathsf{snr}_0} \le \beta$.

By combining the bounds in (56), (57), and (58) we get

$$\begin{aligned}
\mathsf{gap}_2 + \Delta_{(25)} &\le \frac{1}{2}\log\left(\max\left(\frac{1}{c_1},2\right)\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right) + \frac{1}{2}\log\left(1+c_1\frac{\beta}{1+\beta\mathsf{snr}_0}\right)\\
&= \frac{1}{2}\log\left(\max\left(\frac{1}{c_1},2\right) + 2\max(1,2c_1)\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right)\\
&\overset{j)}{\le} \frac{1}{2}\log\left(\max\left(\frac{1}{c_1},2\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right)\\
&\overset{k)}{=} \frac{1}{2}\log\left(\max\left(\frac{2\log\left(\frac{12(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)}{3},2\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right)\\
&\overset{l)}{\le} \frac{1}{2}\log\left(\max\left(\frac{2}{3}\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0}{\beta}\right),2\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right)\\
&\overset{m)}{=} \frac{1}{2}\log\left(\frac{2}{3}\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0}{\beta}\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right),
\end{aligned}$$

where the inequalities follow from: j) the fact that $c_1 \le \frac{3}{2}$; k) using the value of c_1 in (52); l) using $\delta = \beta\frac{\mathsf{snr}_0}{1+\mathsf{snr}_0}$ and $\frac{1+\beta\mathsf{snr}_0}{1+\delta\mathsf{snr}_0} \le \frac{1+\mathsf{snr}_0}{\mathsf{snr}_0} \le 2$ for $\mathsf{snr}_0 \ge 1$; and m) the fact that $\max\left(\frac{2\log\left(\frac{24(1+\beta\mathsf{snr}_0)}{\beta}\right)}{3},2\right) =$

$\frac{2\log\left(\frac{24(1+\beta\mathsf{snr}_0)}{\beta}\right)}{3}$.

This concludes the proof of the gap result for the $\mathsf{snr} \geq \mathsf{snr}_0$ regime.

We next focus on the $1 \leq \mathsf{snr} \leq \mathsf{snr}_0$ regime. We use only the discrete part of the mixed input and set $\delta = 0$. From (52) we have that the input parameters must satisfy

$$N \leq \left\lfloor \sqrt{1 + c_3\mathsf{snr}_0} \right\rfloor, \tag{59a}$$

$$c_3 \leq \frac{3}{2\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)}, \tag{59b}$$

in order to comply with the MMSE constraint in (5c). However, instead of choosing the number of points as in (59) we choose it to be

$$N = \left\lfloor \sqrt{1 + c_3\mathsf{snr}} \right\rfloor \leq \left\lfloor \sqrt{1 + c_3\mathsf{snr}_0} \right\rfloor. \tag{60}$$

The reason for this choice will be apparent from the gap derivation next.

Similarly to the previous case, we compute the gap between the outer bound in Proposition 5 and the achievable mutual information of the mixed input in Proposition 6, where $I\left(X_D, \mathsf{snr}\right)$ is lower bounded using Proposition 7. We have,

$$\begin{aligned}
\mathsf{gap}_2 + \Delta_{(26)} &\leq C_\infty - \log(N) + \frac{1}{2}\log\left(\frac{\pi \mathrm{e}}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}(X_D, \mathsf{snr})\right) \\
&\overset{n)}{\leq} \frac{1}{2}\log\left(\frac{1+\mathsf{snr}}{1+c_3\mathsf{snr}}\right) + \frac{1}{2}\log\left(\frac{4\pi \mathrm{e}}{6}\right) + \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}(X_D, \mathsf{snr})\right) \\
&\overset{o)}{\leq} \frac{1}{2}\log\left(\frac{1+\mathsf{snr}}{1+c_3\mathsf{snr}}\right) + \frac{1}{2}\log\left(\frac{4\pi \mathrm{e}}{6}\right) + \frac{1}{2}\log\left(1 + \frac{c_3\mathsf{snr}}{1+\mathsf{snr}}\right) \\
&= \frac{1}{2}\log\left(\frac{1+(1+c_3)\mathsf{snr}}{1+c_3\mathsf{snr}}\right) + \frac{1}{2}\log\left(\frac{4\pi \mathrm{e}}{6}\right) \\
&\overset{p)}{\leq} \frac{1}{2}\log\left(1 + \frac{1}{c_3}\right) + \frac{1}{2}\log\left(\frac{4\pi \mathrm{e}}{6}\right) \\
&\overset{r)}{=} \frac{1}{2}\log\left(1 + \frac{2}{3}\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)\right) + \frac{1}{2}\log\left(\frac{4\pi \mathrm{e}}{6}\right),
\end{aligned}$$

where the (in)-equalities follow from: n) getting an extra one bit gap by dropping the floor operation; o) using the bound on $d_{\min}^2 = \frac{12}{N^2-1}$ from (60) and bound $\mathrm{mmse}(X, \mathsf{snr}) \leq \frac{1}{1+\mathsf{snr}}$; p) using that $\frac{1+(1+c_3)\mathsf{snr}}{1+c_3\mathsf{snr}} \leq \frac{1+c_3}{c_3} = 1 + \frac{1}{c_3}$; and r) using the value of c_3 from (59).

This concludes the proof for the case $1 \leq \mathsf{snr} \leq \mathsf{snr}_0$.

Finally, note that for the case $\mathsf{snr} \leq 1$ the gap is trivially given by

$$\mathsf{gap}_3 \leq \mathcal{C}(\beta, \mathsf{snr}, \mathsf{snr}_0) - I(X_{\text{mix}}, \mathsf{snr}) \leq \mathcal{C}(\beta, \mathsf{snr}, \mathsf{snr}_0) \leq \frac{1}{2}\log(1+\mathsf{snr}) \leq \frac{1}{2}\log(2). \tag{61}$$

This concludes the proof.

Appendix D
Proof of Proposition 9

We will need the following identities for the proof:

$$\mathsf{snr} \cdot \mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbf{Cov}(\mathbf{Z}|\mathbf{Y})], \tag{62a}$$

$$\mathsf{snr}^2 \cdot \mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})], \tag{62b}$$

which follow since

$$\sqrt{\mathsf{snr}}\mathbf{X} + \mathbf{Z} = \mathbf{Y} = \mathbb{E}[\mathbf{Y}|\mathbf{Y}] = \sqrt{\mathsf{snr}}\mathbb{E}[\mathbf{X}|\mathbf{Y}] + \mathbb{E}[\mathbf{Z}|\mathbf{Y}],$$

and therefore

$$\sqrt{\mathsf{snr}}(\mathbf{X} - \mathbb{E}[\mathbf{X}|\mathbf{Y}]) = (\mathbf{Z} - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]).$$

Next, Observe that

$$\mathbf{Cov}(\mathbf{Z}|\mathbf{Y}) = \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}])(\mathbb{E}[\mathbf{Z}|\mathbf{Y}])^{\mathrm{T}},$$

and so we have that

$$\begin{aligned}
\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y}) &= \left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\right)^2 \\
&= (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] \\
&\quad - \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}} + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{a)}{=} (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - 2\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] \\
&\quad + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{b)}{\preceq} (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - 2\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}} \\
&\quad + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&= (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{c)}{=} \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}] - \mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}) - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2,
\end{aligned} \tag{63}$$

where the order operations follow from: a) the fact that $\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathsf{T}}$ and $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}]$ are symmetric matrices; b) using $\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathsf{T}} \preceq \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}]$ (from the positive semi-definite property of the conditional covariance matrix); and c) the fact that, since $\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}) = \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}})^{\mathsf{T}}|\mathbf{Y}] - \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}](\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}])^{\mathsf{T}}$ and by symmetry of $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}]$, we have that $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}](\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}])^{\mathsf{T}} = (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}])^2$. By using the monotonicity of the trace, properties of the expected value, and the inequality in (63), we have that

$$\begin{aligned}
\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})]\right) &\leq \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}})^{\mathsf{T}}|\mathbf{Y}] - \mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}) - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathsf{T}})^2\right]\right) \\
&= \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}})^{\mathsf{T}}|\mathbf{Y}]\right]\right) - \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y})\right]\right) \\
&- \mathrm{Tr}\left(\mathbb{E}\left[(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathsf{T}})^2\right]\right). \qquad (64)
\end{aligned}$$

We next focus on each term of the right hand side of (64) individually. The first term can be computed as follows:

$$\begin{aligned}
\mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}})^{\mathsf{T}}|\mathbf{Y}]\right]\right) &\stackrel{d)}{=} \mathrm{Tr}\left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}]\right) \\
&\stackrel{e)}{=} \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} Z_i^2\right)^2\right] \\
&\stackrel{f)}{=} n(n+2), \qquad (65)
\end{aligned}$$

where the (in)-equalities follow from: d) using the law of total expectation; e) since expectation is a linear operator and using fact that the trace can be exchanged with linear operators; and f) observing that $S = \sum_{i=1}^{n} Z_i^2$ is a chi-square distribution of degree n and hence $\mathbb{E}[S] = n(n+2)$.

For the second term in (64), by definition of the MMSE, we have

$$\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y})\right]\right) = n\mathrm{mmse}(\mathbf{Z}\mathbf{Z}^{\mathsf{T}}|\mathbf{Y}). \qquad (66)$$

The third term in (64) satisfies

$$\begin{aligned}
\mathrm{Tr}\left(\mathbb{E}\left[(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2\right]\right) &\overset{g)}{\geq} \mathrm{Tr}\left(\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\right]\right)^2\right) \\
&= \mathrm{Tr}\left(\left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}] - \mathbb{E}[\mathbf{Cov}(\mathbf{Z}|\mathbf{Y})]\right)^2\right) \\
&\overset{h)}{=} \mathrm{Tr}\left((\mathbf{I} - \mathsf{snr}\ \mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})])^2\right) \\
&\overset{i)}{=} \mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right) \qquad (67)
\end{aligned}$$

where the (in)-equalities follow from: g) using Jensen's inequality; h) using the property: $\mathsf{snr} \cdot \mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbf{Cov}(\mathbf{Z}|\mathbf{Y})]$ in (62); and i) using identity [18]

$$\mathbf{I} - \mathsf{snr}\ \mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})] = \mathbf{J}(\mathbf{Y}).$$

By putting (65), (66), and (67) together, we have that

$$\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})\right] \leq k_n := \frac{n(n+2) - n\ \mathrm{mmse}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}) - \mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right)}{n}.$$

Finally, using the identity $\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})\right] = \mathsf{snr}^2 \cdot \mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]$ in (62) concludes the proof.

Appendix E
Proof of Proposition 10

Using the Cramer-Rao lower bound [27, Theorem 20] we have that

$$\begin{aligned}
\mathbf{J}(\mathbf{Y}) &\succeq \mathbf{Cov}^{-1}(\mathbf{Y}) \\
&= \left(\mathsf{snr}\mathbb{E}[\mathbf{X}\mathbf{X}^T] + \mathbf{I}\right)^{-1} \\
&= \mathbf{V}^{-1}\mathbf{\Lambda}^{-1}\mathbf{V},
\end{aligned}$$

where $\mathbf{\Lambda}$ is the eigen-matrix of $\mathsf{snr} \cdot \mathbb{E}[\mathbf{X}\mathbf{X}^T] + \mathbf{I}$, which is a diagonal matrix with the following values along the diagonal: $\lambda_i = \mathsf{snr}\sigma_i + 1$, and σ_i is the i-th eigenvalue of matrix $\mathbb{E}[\mathbf{X}\mathbf{X}^T]$. Therefore,

$$\begin{aligned}
\mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right) &\geq \mathrm{Tr}\left(\mathbf{V}^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}\left(\mathbf{V}^{-1}\mathbf{\Lambda}^{-1}\mathbf{V}\right)^T\right) \\
&= \mathrm{Tr}(\mathbf{\Lambda}^{-2}) \\
&= \sum_{i=1}^{n} \frac{1}{(1+\mathsf{snr}\sigma_i)^2} \\
&\geq \frac{n}{(1+\mathsf{snr})^2},
\end{aligned}$$

where the last inequality comes from minimizing $\sum_{i=1}^{n} \frac{1}{(1+\mathsf{snr}\sigma_i)^2}$ subject to the constraint that $\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) = \sum_{i=1}^{n} \sigma_i \le n$ and where the minimum is attained with $\sigma_i = 1$ for all i.

Finally, note that all inequalities are equalities if $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, (1+\mathsf{snr})\mathbf{I})$ or equivalently if $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. This concludes the proof.

Appendix F
Proof of Proposition 11

First observe that since the conditional expectation is the best estimator under a squared cost function

$$\begin{aligned}\mathbf{Cov}(\mathbf{X}|\mathbf{Y}=\mathbf{y}) &= \mathbb{E}\left[(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])(\mathbf{X}-\mathbb{E}[\mathbf{X}|\mathbf{Y}])^{\mathrm{T}}|\mathbf{Y}=\mathbf{y}\right] \\ &\preceq \mathbb{E}\left[(\mathbf{X}-f(\mathbf{Y}))(\mathbf{X}-f(\mathbf{Y}))^{\mathrm{T}}|\mathbf{Y}=\mathbf{y}\right], \end{aligned} \tag{68}$$

for any deterministic function $f(\cdot)$. Therefore, the first bound in (37a) follows by choosing $f(\mathbf{Y}) = \frac{\sqrt{\mathsf{snr}}\mathbf{Y}}{1+\mathsf{snr}}$ in (68)

$$\begin{aligned}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) &\le \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[\left(\mathbf{X}-\frac{\sqrt{\mathsf{snr}}\mathbf{Y}}{1+\mathsf{snr}}\right)\left(\mathbf{X}-\frac{\sqrt{\mathsf{snr}}\mathbf{Y}}{1+\mathsf{snr}}\right)^{\mathrm{T}}|\mathbf{Y}\right]\right]\right) \\ &= \frac{1}{(1+\mathsf{snr})^4}\mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[\left(\mathbf{X}-\sqrt{\mathsf{snr}}\mathbf{Z}\right)\left(\mathbf{X}-\sqrt{\mathsf{snr}}\mathbf{Z}\right)^{\mathrm{T}}|\mathbf{Y}\right]\right]\right) \\ &\le \frac{1}{(1+\mathsf{snr})^4}\mathrm{Tr}\left(\mathbb{E}\left[\left(\left(\mathbf{X}-\sqrt{\mathsf{snr}}\mathbf{Z}\right)\left(\mathbf{X}-\sqrt{\mathsf{snr}}\mathbf{Z}\right)^{\mathrm{T}}\right)^2\right]\right),\end{aligned}$$

where the last inequality is due to Jensen's inequality.

The second bound in (37a) follows by choosing $f(\mathbf{Y}) = 0$ in (68)

$$\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) \le \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[(\mathbf{X}-\mathbf{0})(\mathbf{X}-\mathbf{0})^{\mathrm{T}}|\mathbf{Y}\right]\right]\right) = \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}^2\left[\mathbf{X}\mathbf{X}^{\mathrm{T}}|\mathbf{Y}\right]\right]\right).$$

This concludes the proof.

References

[1] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai, and H. V. Poor, "On communications through a Gaussian channel with a MMSE disturbance constraint," in *Proc. Workshop on Info. Theory and Applications*, Feb 2016.

[2] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, 2009.

[3] A. S. Motahari and A. K. Khandani, "Capacity bounds for the gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, 2009.

[4] V. S. Annapureddy and V. V. Veeravalli, “Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region,” *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, 2009.

[5] T. Han and K. Kobayashi, “A new achievable rate region for the interference channel,” *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 49–60, Jan. 1981.

[6] H. Sato, “The capacity of Gaussian interference channel under strong interference,” *IEEE Trans. Inf. Theory*, vol. IT-27, no. 6, pp. 786–788, Nov. 1981.

[7] R. Etkin, D. Tse, and H. Wang, “Gaussian interference channel capacity to within one bit,” *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[8] A. Bennatan, S. Shamai, and A. Calderbank, “Soft-decoding-based strategies for relay and interference channels: Analysis and achievable rates using ldpc codes,” *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 1977–2009, April 2014.

[9] K. Moshksar, A. Ghasemi, and A. Khandani, “An alternative to decoding interference or treating interference as Gaussian noise,” *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 305–322, Jan 2015.

[10] A. Dytso, D. Tuninetti, and N. Devroye, “On the two-user interference channel with lack of knowledge of the interference codebook at one receiver,” *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1257–1276, March 2015.

[11] ——, “Interference as noise: Friend or foe?” To appear in *IEEE Trans. Inf. Theory,* http://arxiv.org/abs/1506.02597, 2015.

[12] R. Bustin and S. Shamai, “MMSE of ‘bad’ codes,” *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 733–743, Feb 2013.

[13] S. Shamai, “From constrained signaling to network interference alignment via an information-estimation perspective,” *IEEE Information Theory Society Newsletter*, vol. 62, no. 7, pp. 6–24, September 2012.

[14] V. Cadambe and S. Jafar, “Interference alignment and the degrees of freedom for the K-user interference channel,” *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[15] W. Yihong and S. Verdú, “MMSE dimension,” *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug 2011.

[16] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Wiley, 2006.

[17] D. Guo, S. Shamai, and S. Verdú, “Mutual information and minimum mean-square error in Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[18] D. Guo, Y. Wu, S. Shamai, and S. Verdú, “Estimation in Gaussian noise: Properties of the minimum mean-square error,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, April 2011.

[19] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai, “On MMSE crossing properties and implications in parallel vector Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 818–844, Feb 2013.

[20] D. Guo, S. Shamai, and S. Verdú, *The Interplay Between Information and Estimation Measures*. now Publishers Incorporated, 2013.

[21] R. Bustin, R. F. Schaefer, H. V. Poor, and S. Shamai, “On MMSE properties of optimal codes for the Gaussian wiretap channel,” in *Proc. IEEE Inf. Theory Workshop*, April 2015, pp. 1–5.

[22] Y. Wu and S. Verdú, “Functional properties of minimum mean-square error and mutual information,” *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, March 2012.

[23] N. Merhav, D. Guo, and S. Shamai, “Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transitions,” *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1400–1416, March 2010.

[24] W. Huleihel and N. Merhav, “Analysis of mismatched estimation errors using gradients of partition functions,” *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2190–2216, 2014.

[25] B. Bandemer and A. El Gamal, “Communication with disturbance constraints,” *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4488–4502, Aug 2014.

[26] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[27] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov 1991.