On Communication Through a Gaussian Channel With an MMSE Disturbance Constraint

Alex Dytso, Ronit Bustin, *Member, IEEE*, Daniela Tuninetti, Natasha Devroye, H. Vincent Poor, *Fellow, IEEE*, and 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 the minimum mean square error (MMSE) of the interference that the transmission to the primary receiver inflicts on the secondary receiver. This 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 the 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). The bound also complements the "single-crossing point property" of the MMSE that upper bounds the MMSE for all SNR values *above* a certain value 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 $O(\log\log(1/\mathsf{MMSE}))$ (where MMSE is the disturbance constraint), is shown by means of the mixed inputs technique recently introduced by Dytso *et al.

***Index Terms*—MMSE, discrete inputs.**

Manuscript received March 24, 2016; revised December 29, 2016; accepted August 7, 2017. Date of publication August 31, 2017; date of current version December 20, 2017. A. Dytso was supported in part by U.S. National Science Foundation (NSF) under Grants CNS-1456793, CCF-1420575, and CCF-1422511. R. Bustin was supported in part by a Women Postdoctoral Scholarship of Israel's Council for Higher Education (VATAT) 2014-2015 and in part by the European Union's Horizon 2020 Research and Innovation Programme, under Grant 694630. D. Tuninetti and N. Devroye were supported by the NSF under Grant CCF-1422511. H. V. Poor was supported by the NSF under Grants CNS-1456793 and CCF-1420575. S. Shamai (Shitz) was supported in part by the Israel Science Foundation, in part by the S. and N. Grand Research Fund, and in part by the European Union's Horizon 2020 Research and Innovation Programme, under Grant 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. This paper was presented in part in [1].

A. Dytso and H. V. Poor are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: adytso@princeton.edu; poor@princeton.edu).

R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel (e-mail: ronitbustin@post.tau.ac.il; sshlomo@ee.technion.ac.il).

D. Tuninetti and N. Devroye are with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: danielat@uic.edu; devroye@uic.edu).

Communicated by S. S. Pradhan, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2017.2747556

I. Introduction

A. Problem Definition: the Max-I Problem

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 $\mathbf{Z}, \mathbf{Z}_0, \mathbf{X}, \mathbf{Y}$ and $\mathbf{Y}_{\mathsf{snr}_0} \in \mathbb{R}^n$; $\mathbf{Z}, \mathbf{Z}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$; 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) in 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

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

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 $\mathrm{mmse}(\mathbf{X},\mathsf{snr})$.

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, as shown in Fig. 1.

We assume that there is only one message for the primary receiver, and the primary transmitter inflicts interference (disturbance) on a secondary receiver. The primary transmitter wishes to maximize its communication rate, while subject to a constraint on the disturbance it inflicts on the secondary receiver. The disturbance is measured in terms of the MMSE. Intuitively, the MMSE disturbance constraint quantifies the

[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 $p_{\mathbf{Y}_2|\mathbf{X}}$.

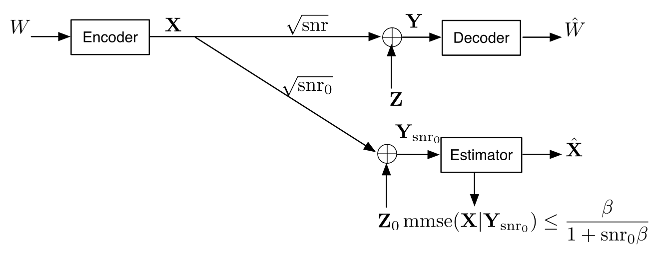


Fig. 1. Channel model.

remaining interference after partial interference cancellation or soft-decoding has been performed [2], [3]. Now consider the following problem:

Definition 1 (Max-I Problem): For some $\beta \in [0, 1]$

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

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

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

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 Fig. 1 is captured when $n \to \infty$ in the Max-I problem, in which case the objective function has a meaning of reliable achievable rate. In [2, Theorem 3] the capacity of the channel in Fig. 1 was properly defined and it was shown to be equal to $\lim_{n\to\infty} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ where

$$\begin{aligned}
&\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) \\
&= \lim_{n\to\infty} \mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta), \\
&= \begin{cases} \frac{1}{2}\log(1+\mathsf{snr}), & \mathsf{snr} \le \mathsf{snr}_0, \\ \frac{1}{2}\log(1+\beta\mathsf{snr}) + \frac{1}{2}\log\left(1 + \frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}\right), & \mathsf{snr} \ge \mathsf{snr}_0, \end{cases} \\
&= \frac{1}{2}\log^+\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(1+\min(\mathsf{snr}, \mathsf{snr}_0)\right),
\end{aligned} \tag{6}$$

which is achieved by using superposition coding with Gaussian codebooks. Fig. 2 shows a plot of $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (6) normalized by the capacity of the point-to-point channel $\frac{1}{2}\log(1+\mathsf{snr})$. The region $\mathsf{snr} \le \mathsf{snr}_0$ (the 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 $\mathsf{snr} \ge \mathsf{snr}_0$ (the 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 (G-IC), but it is unknown whether it achieves capacity [4]).

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

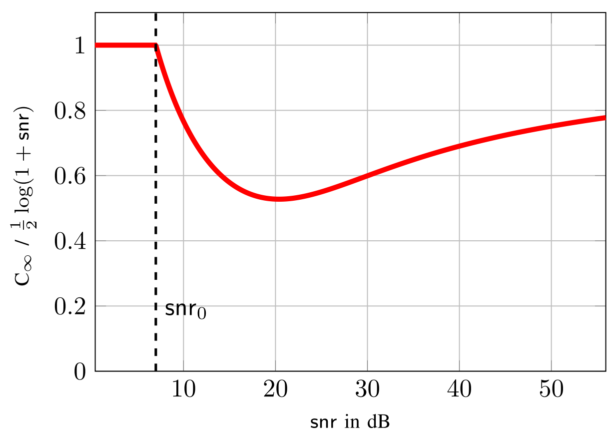


Fig. 2. Plot of $\frac{\mathcal{C}_\infty(\mathsf{snr},\mathsf{snr}_0,\beta)}{\frac{1}{2}\log(1+\mathsf{snr})}$ vs. snr in dB, for $\beta = 0.01$ and $\mathsf{snr}_0 = 5 = 6.989$ dB.

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 [5]–[7].
2) *Partial Interference Cancellation*: By using the Han-Kobayashi (HK) achievable scheme [8], 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 [9] and optimal within 1/2 bit per channel per user otherwise [4].
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 [10]–[12], and shown to be optimal to within either a constant or a $O(\log\log(\mathsf{snr}))$ gap for all regimes in [13].

Even though the Max-I problem is somewhat simplified, compared to the G-IC, it can serve as an important building block towards characterizing the capacity of the G-IC [2], [3], especially in light of the known (but currently uncomputable) capacity limit expression [14]

$$\mathcal{C}_\infty^{\mathrm{IC}} = \lim_{n\to\infty} \mathsf{co} \bigcup_{P_{\mathbf{X}_1\mathbf{X}_2} = P_{\mathbf{X}_1}P_{\mathbf{X}_2}} \left\{ \begin{matrix} 0 \le R_1 \le I_n(\mathbf{X}_1; \mathbf{Y}_1) \\ 0 \le R_2 \le I_n(\mathbf{X}_2; \mathbf{Y}_2) \end{matrix} \right\}, \tag{7}$$

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

$$\mathcal{C}_n^{\mathrm{IC}} \subset \mathcal{C}_\infty^{\mathrm{IC}}, \tag{8}$$

where

$$\mathcal{C}_n^{\rm IC} = \mathsf{co} \bigcup_{P_{\mathbf{X}_1\mathbf{X}_2} = P_{\mathbf{X}_1} P_{\mathbf{X}_2}} \left\{ \begin{array}{l} 0 \le R_1 \le I_n(\mathbf{X}_1; \mathbf{Y}_1) \\ 0 \le R_2 \le I_n(\mathbf{X}_2; \mathbf{Y}_2) \end{array} \right\}. \tag{9}$$

The inner bound $\mathcal{C}_n^{\rm IC}$ will be referred to as the treating-interference-as-noise inner bound. Finding the input distributions $P_{\mathbf{X}_1}$ and $P_{\mathbf{X}_2}$ that exhaust the achievable region in $\mathcal{C}_n^{\rm IC}$ is an important open problem. Recently, for the special case of $n = 1$, $\mathcal{C}_1^{\rm IC}$ has been shown to be within a constant or $O(\log\log(\mathsf{snr}))$ from the capacity $\mathcal{C}_\infty^{\rm IC}$ [13]. Therefore, the Max-I problem, denoted by $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (5), can serve as an important step in characterizing the structure of optimal input distributions for $\mathcal{C}_n^{\rm IC}$. We also note that in [3, Sec. VI.3] and [2, Sec. VIII] 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 treating interference as noise region in (9), we refer the reader to [11], [12], [15]–[17] and the references therein.

The importance of studying models of communication systems with disturbance constraints has been recognized previously. For example, Bandemer and El Gamal [18] studied the following problem related to the Max-I problem in (5).

Definition 2 (Bandemer et al. Problem): For some $R \ge 0$

$$\mathcal{I}_n(\mathsf{snr}, \mathsf{snr}_0, R) := \max_{\mathbf{X}} I_n(\mathbf{X}, \mathsf{snr}), \tag{10a}$$

$$\text{s.t. } \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \le 1, \text{ power constraint,} \tag{10b}$$

$$\text{and } I_n(\mathbf{X}, \mathsf{snr}_0) \le R, \text{ disturbance constraint.} \tag{10c}$$

In [18] 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 $\alpha = \min\left(1, \frac{e^{2R}-1}{\mathsf{snr}_0}\right)$; here α is such that the most stringent constraint between (10b) and (10c) 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 (10c) is not violated.

Measuring the disturbance with the mutual information as in (10), in contrast to the MMSE as in (5), suggests that it is always optimal to use Gaussian codebooks with reduced power without any rate splitting. Moreover, while the mutual information constraint in (10) limits the amount of information transmitted to the unintended receiver, it may not be the best choice for modeling the interference, since any information that can be reliably decoded by the unintended receiver is not really interference. For this reason, and since the MMSE constraint accounts for the interference and 'depicts' the key role of rate splitting, it has been argued in [2] and [3] that the Max-I problem in (5) with the MMSE disturbance constraint is a more suitable building block to study the G-IC.

We also refer the reader to [19] 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 [19] 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.

B. The I-MMSE Identity

The basis for the analysis of the Max-I problem in [2] is the fundamental relationship between information theory and estimation theory, also known as the Guo, Shamai and Verdú I-MMSE relationship.

Proposition 1 (I-MMSE Relationship [20, Th. 1]): The I-MMSE relationship is given by the derivative relationship

$$\frac{d}{d\mathsf{snr}} I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{11a}$$

or the integral relationship

$$I_n(\mathbf{X}, \mathsf{snr}) = \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{mmse}(\mathbf{X}, \gamma)d\gamma. \tag{11b}$$

Observe that the Max-I problem in (5) and the one in (10) have the same objective function but have different constraints. The relationship between the constraints in (5c) and (10c) 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 (10c), via the integral I-MMSE relationship in (11), imposes a constraint on the area below the function $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ in the range $\mathsf{snr} \in [0, \mathsf{snr}_0]$.

C. Bounds on the MMSE

Upper bounds on the MMSE are useful, thanks to the I-MMSE relationship, as tools to derive converse results, and have been used in [21]–[24] to name a few. The key bound to show the converse result for $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ are the *linear MMSE (LMMSE) upper bound* and *single-crossing point property* (SCPP) bound presented next.

Proposition 2 (LMMSE Bound [20]): For any $\mathbf{X}$ and $\mathsf{snr} > 0$ it holds that

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{1}{\mathsf{snr}}. \tag{12a}$$

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

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{\sigma^2}{1 + \sigma^2\mathsf{snr}}, \tag{12b}$$

where equality in (12b) is achieved iff $\mathbf{X} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

Another important bound for the MMSE is the SCPP bound developed in [22] for $n = 1$ and extended in [23] to any $n \ge 1$.

Proposition 3 (SCPP [23]): For any fixed $\mathbf{X}$, suppose that $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta\mathsf{snr}_0}$, for some fixed $\beta \ge 0$. Then for all $\mathsf{snr} \in [\mathsf{snr}_0, \infty)$ we have that

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \frac{\beta}{1 + \beta\mathsf{snr}}, \tag{13a}$$

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

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \ge \frac{\beta}{1 + \beta\mathsf{snr}}. \tag{13b}$$

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 ($\mathsf{snr}_0 \le \mathsf{snr}$) we have the upper bound in (13a) and for all lower SNR values ($\mathsf{snr} \le \mathsf{snr}_0$) we have the lower bound in (13b).

D. Max-MMSE Problem

Motivated by the I-MMSE relationship the approach of [2] was to examine the limiting behavior of the following optimization problem.

Definition 3 (Max-MMSE Problem): For some $\beta \in [0, 1]$

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) := \sup_{\mathbf{X}} \mathrm{mmse}(\mathbf{X}, \mathsf{snr}), \tag{14a}$$

$$\textit{s.t. } \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{X}\mathbf{X}^T]\right) \leq 1, \textit{ power constraint,} \tag{14b}$$

$$\textit{and } \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}, \textit{ MMSE constraint.} \tag{14c}$$

The authors of [2] and [25] proved that

$$\begin{aligned}\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) &= \lim_{n\to\infty} \mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \\ &= \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} < \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr} \geq \mathsf{snr}_0, \end{cases}\end{aligned} \tag{15}$$

achieved by superposition coding with Gaussian codebooks. Clearly there is a discontinuity in (15) 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* [25]. For other recent links between random codes, the MMSE and statistical physics the reader is referred to [26].

The LMMSE bound provides the converse solution for $\mathrm{M}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (15) 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.

We note that, through the I-MMSE relation, integration of $\mathrm{M}_\infty(\gamma, \mathsf{snr}_0, \beta)$ over $\gamma \in [0, \mathsf{snr}]$ gives $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$. However, 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. 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 be 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 Max-MMSE solution for every γ.

One of the main objectives of this paper is to develop bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ and then use the I-MMSE relationship to bound $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$. 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 (14) 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 the Max-MMSE or the Max-I problem falls under the category of convex optimization. This follows from the fact that the MMSE is a strictly concave function in the input distribution [27]. Therefore, the set of input distributions, defined by (14b) and (14c), over which we are optimizing, might not be convex. It is also a good place to show that the set of permissible input distribution is not empty.

Proposition 4: There exists an input distribution $\mathbf{X}$ *with maximum power as in* (5b) *and* (14b) *that satisfies the MMSE constraint in* (5c) *and* (14c) *for any* $\mathsf{snr}_0 > 0$ *and any* $\beta > 0$.

Proof: See Appendix A. □

Note that Proposition 3 gives a solution to the Max-MMSE problem in (14) for $\mathsf{snr} \geq \mathsf{snr}_0$ and any $n \geq 1$ as follows:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta\mathsf{snr}}, \text{ for } \mathsf{snr} \geq \mathsf{snr}_0, \tag{16}$$

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

However, for the case $\mathsf{snr} \leq \mathsf{snr}_0$ the LMMSE bound in (12b) is no longer tight. Therefore, in the rest of the paper, the treatment of the Max-MMSE problem will focus only on the regime $\mathsf{snr} \leq \mathsf{snr}_0$. We refer to the upper bounds in the regime $\mathsf{snr} \leq \mathsf{snr}_0$ as the *complementary SCPP bounds.*

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 $\mathsf{snr} \leq \mathsf{snr}_0$. It is also well known that, for any finite n, $\mathrm{mmse}(\mathbf{X}, \mathsf{snr})$ is a continuous function of snr [22]. Therefore, $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ must be of the following form:

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) = \begin{cases} \frac{1}{1+\mathsf{snr}}, & \mathsf{snr} \leq \mathsf{snr}_L, \\ T_n(\mathsf{snr}, \mathsf{snr}_0, \beta), & \mathsf{snr}_L \leq \mathsf{snr} \leq \mathsf{snr}_0, \\ \frac{\beta}{1+\beta\mathsf{snr}}, & \mathsf{snr}_0 \leq \mathsf{snr}, \end{cases} \tag{17}$$

for some snr_L. In this paper we seek to characterize snr_L in (17) and the continuous function $T_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ such that

$$T_n(\mathsf{snr}_L, \mathsf{snr}_0, \beta) = \frac{1}{1+\mathsf{snr}_L}, \tag{18a}$$

$$T_n(\mathsf{snr}_0, \mathsf{snr}_0, \beta) = \frac{\beta}{1+\beta\mathsf{snr}_0}, \tag{18b}$$

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

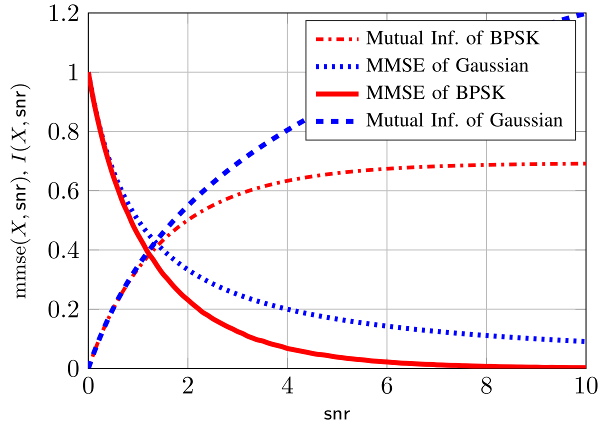
$$W_n := \mathsf{snr}_0 - \mathsf{snr}_L. \tag{19}$$

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.

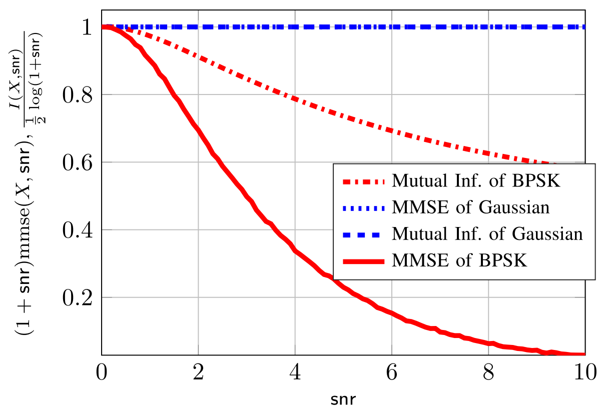
E. Contributions and Paper Outline

The main contributions of the paper are as follows. In Section II 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 6 provides a lower bound on the width of the phase transition region defined in (19) of the order of $\frac{1}{n}$.
- Proposition 7 provides a new upper bound for the Max-I problem for arbitrary n.
- Proposition 10 shows that, for the case of $n = 1$, superposition of discrete and Gaussian inputs, termed *mixed inputs* in [13], achieves the proposed upper bound on the Max-I problem from Proposition 7 to within an



(a)



(b)

Fig. 3. Comparing mutual informations and MMSE's for binary phase shift keying (BPSK) and Gaussian inputs. Fig. 3b clearly shows the multiplicative loss of BPSK, for both mutual information and MMSE, compared to a Gaussian input. (a) Unnormalized plot. (b) Normalized plot (dashed red and dotted blue lines overlap).

additive gap of order $\log\log\frac{1}{\mathrm{mmse}(X,\mathsf{snr}_0)}$. We note that strictly speaking the gap result is only a constant with respect to snr but not snr_0.

- Proposition 11 shows that, as $n \to \infty$, superposition of a lattice constellation and Gaussian inputs exactly achieves the upper bound on the Max-I problem, recovering the result of [2].
- Section II-F discusses how the result can be extended to an arbitrary finite n.

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

- Proposition 17 considerably refines existing bounds on the derivative of MMSE for $n = 1$ and generalizes them to any n.
- In Section III-A, by using Proposition 17, we prove Theorem 1.

Most proofs can be found in the appendix.

F. 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 the 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 equally spaced 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 ordering notation $\mathbf{A} \succeq \mathbf{B}$ implies that $\mathbf{A}-\mathbf{B}$ is a positive semidefinite matrix; for $\mathbf{x} \in \mathbb{R}^n$ the Euclidian norm is denoted by $\|\mathbf{x}\|$; 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 $x \ge x_0$; we denote the upper incomplete gamma function and the gamma function, respectively, as

$$\Gamma(x; a) := \int_a^\infty t^{x-1}e^{-t}dt,\ \ x \in \mathbb{R}, a \in \mathbb{R}^+, \tag{20a}$$

$$\Gamma(x) := \Gamma(x; 0). \tag{20b}$$

G. On the 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 the mutual information $I_n(\mathbf{X}, \mathsf{snr})$ in (3) we will plot

$$d(\mathbf{X}, \mathsf{snr}) := \frac{I_n(\mathbf{X}, \mathsf{snr})}{\frac{1}{2}\log(1+\mathsf{snr})}, \tag{21}$$

while for the MMSE in (4) we will plot

$$D(\mathbf{X}, \mathsf{snr}) := \frac{\mathrm{mmse}(\mathbf{X}, \mathsf{snr})}{\frac{1}{1+\mathsf{snr}}} = (1+\mathsf{snr})\cdot\mathrm{mmse}(\mathbf{X}, \mathsf{snr}). \tag{22}$$

In particular, at high snr the quantity in (21) is commonly referred to as the *degrees of freedom* [28] and the quantity in (22) as the *MMSE dimension* [29]. Moreover, it is well known that under the block-power constraint in (5b), a Gaussian input maximizes both the mutual information and the MMSE [30], and thus the quantities $d(\mathbf{X}, \mathsf{snr})$ and $D(\mathbf{X}, \mathsf{snr})$ have natural meanings as the multiplicative losses of the input $\mathbf{X}$ compared to the Gaussian input. Fig. 3 compares normalized and unnormalized quantities.

II. 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 (17). 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.

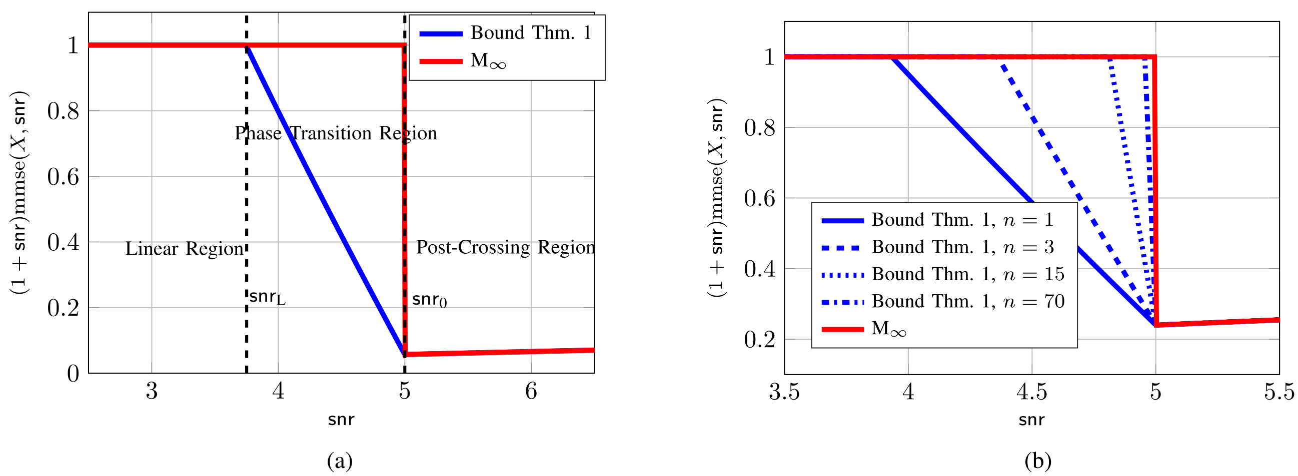


Fig. 4. Bounds on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr. (a) For $n = 1$, $\mathsf{snr}_0 = 5$ and $\beta = 0.01$. (b) For several values of n, $\mathsf{snr}_0 = 5$ and $\beta = 0.05$.

Theorem 1 (D-Bound): For any $\mathbf{X}$ *and* $0 < \mathsf{snr} \le \mathsf{snr}_0$, *let* $\mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) = \frac{\beta}{1+\beta\mathsf{snr}_0}$ *for some* $\beta \in [0, 1]$; *then we have*

$$\mathrm{mmse}(\mathbf{X}, \mathsf{snr}) \le \mathrm{mmse}(\mathbf{X}, \mathsf{snr}_0) + k_n \left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0} \right) - \Delta, \tag{23a}$$

$$k_n \le n + 2, \quad \Delta = 0. \tag{23b}$$

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

$$\Delta := \Delta_{(23c)} = \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{1}{\gamma^2(1+\gamma)^2} d\gamma = 2\log\left(\frac{1+\mathsf{snr}_0}{1+\mathsf{snr}}\right) - 2\log\left(\frac{\mathsf{snr}_0}{\mathsf{snr}}\right) + \frac{1}{1+\mathsf{snr}} - \frac{1}{1+\mathsf{snr}_0} + \frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0}. \tag{23c}$$

Proof: See Section III-A. ∎

Since, the bound in Theorem 1 holds for any $\mathbf{X}$ we also get the following bound.

Proposition 5: For $\mathsf{snr} \le \mathsf{snr}_0$

$$\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \le \frac{\beta}{1+\beta\mathsf{snr}_0} + k_n \left(\frac{1}{\mathsf{snr}} - \frac{1}{\mathsf{snr}_0} \right) - \Delta. \tag{24}$$

The bound on $\mathrm{M}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (24) 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 in (24) provides a continuous upper bound on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ which is tighter 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 6: The bound in (23a), *with* $\Delta = 0$, *from Theorem 1 intersects with the LMMSE bound in* (12a) *from Proposition 2 at*

$$\mathsf{snr}_L \ge \mathsf{snr}_0 \frac{1+\beta\mathsf{snr}_0}{\frac{k_n}{k_n-1}+\beta\mathsf{snr}_0} = O\left(\left(1-\tfrac{1}{n}\right)\mathsf{snr}_0\right). \tag{25a}$$

Thus, the width of the phase transition region is upper bounded, for k_n *in* (23b), *by*

$$W_n \le \frac{1}{k_n-1} \frac{\mathsf{snr}_0}{\frac{k_n}{k_n-1}+\beta\mathsf{snr}_0} = O\left(\tfrac{1}{n}\right). \tag{25b}$$

Proof: See Appendix B. ∎

In Proposition 6 we found the intersection between the LMMSE bound $\frac{1}{\mathsf{snr}}$ in (12a) and the bound in (23a) from Theorem 1. Unfortunately, for the power constraint case, the intersection of the LMMSE bound $\frac{1}{1+\mathsf{snr}}$ in (12b) and the bound in (23c) cannot be found analytically. However, the solution can be computed efficiently by using numerical methods. Moreover, the asymptotic behavior of the width 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 solid 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 lines are the bounds 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 (15) as $n \to \infty$.

B. Max-I Problem: Upper Bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$

By using Theorem 1 (with the 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 7: For any $0 \le \mathsf{snr}_0$, $\beta \in [0, 1]$, *and* snr_L *given in* (25), *we have that for* $\mathsf{snr}_0 \le \mathsf{snr}$

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \le \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) - \Delta_{(26a)}, \tag{27a}$$

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

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \le \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) - \Delta_{(26b)}, \tag{27b}$$

where $\Delta_{(26a)}$ *and* $\Delta_{(26b)}$ *are given in* (26) *shown at the bottom of the next page.*

Proof: 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

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \le \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{M}_n(t, \mathsf{snr}_0, \beta)dt = \frac{1}{2}\log(1+\mathsf{snr}_L)+\frac{1}{2}\int_{\mathsf{snr}_L}^{\mathsf{snr}_0} T_n(t, \mathsf{snr}_0, \beta)dt + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right), \text{ for } \mathsf{snr}_0 \le \mathsf{snr}, \quad (28)$$

and

$$\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) \le \frac{1}{2}\int_0^{\mathsf{snr}} \mathrm{M}_n(t, \mathsf{snr}_0, \beta)dt \le \frac{1}{2}\log(1+\min(\mathsf{snr}_L, \mathsf{snr})) + \frac{1}{2}\int_{\min(\mathsf{snr}_L,\mathsf{snr})}^{\mathsf{snr}} T_n(t, \mathsf{snr}_0, \beta)dt, \text{ for } \mathsf{snr}_0 \ge \mathsf{snr}. \quad (29)$$

We only show steps leading to (27a) and (26a), as shown at the bottom of this page, since the proof of (27b) and (26b) follows similarly. From (28) we have that

$$\begin{aligned}\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta) &\le \frac{1}{2}\log(1+\mathsf{snr}_L) + \frac{1}{2}\int_{\mathsf{snr}_L}^{\mathsf{snr}_0} T_n(t, \mathsf{snr}_0, \beta)dt + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\beta\mathsf{snr}_0}\right)\\ &= \mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta) - \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_0}{1+\mathsf{snr}_L}\right) + \frac{1}{2}\int_{\mathsf{snr}_L}^{\mathsf{snr}_0} T_n(t, \mathsf{snr}_0, \beta)dt. \quad (30)\end{aligned}$$

Next by using Theorem 1 (30) can be bounded as follows:

$$\int_{\mathsf{snr}_L}^{\mathsf{snr}_0} T_n(t, \mathsf{snr}_0, \beta)dt \le \int_{\mathsf{snr}_L}^{\mathsf{snr}_0} \frac{\beta}{1+\beta\mathsf{snr}_0} + k_n\left(\frac{1}{t} - \frac{1}{\mathsf{snr}_0}\right)dt - \int_{\mathsf{snr}_L}^{\mathsf{snr}_0}\int_t^{\mathsf{snr}_0} \frac{1}{\gamma^2(1+\gamma)^2}d\gamma\, dt, \quad (31)$$

where the integration of (31) leads to (26a). This concludes the proof. ■

Fig. 5 compares the bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ from Proposition 7 with $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ in (6) for several values of n. The figure shows how the new bounds in Proposition 7 improve on the trivial $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ bound for finite n.

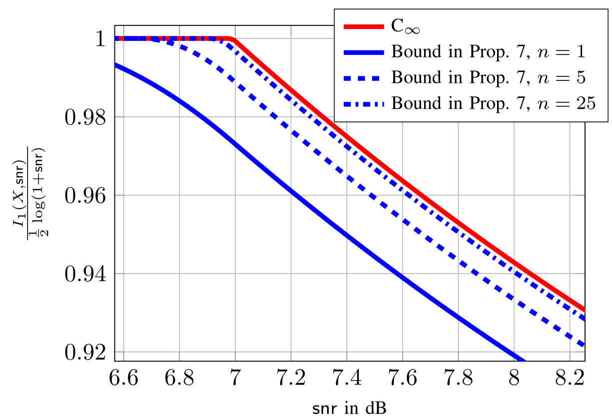


Fig. 5. Bounds on $\mathcal{C}_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$ from Proposition 7 vs. snr, for $\beta = 0.1$ and $\mathsf{snr}_0 = 5 = 6.9897$ dB.

C. Max-MMSE Problem: Achievability of $\mathrm{M}_I(\mathsf{snr}, \mathsf{snr}_0, \beta)$

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. This input is referred to as a *mixed input* [13] and is defined as

$$\mathbf{X}_{\text{mix}} := \sqrt{1-\delta}\mathbf{X}_D + \sqrt{\delta}\mathbf{X}_G, \ \delta \in [0, 1]: \quad (32)$$

$$\mathbf{X}_G \sim \mathcal{N}(0, \mathbf{I}), \ \mathbb{E}[\|\mathbf{X}_D\|^2] \le n, \ \frac{1}{n}H(\mathbf{X}_D) < \infty, \quad (33)$$

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 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 8: For $\mathbf{X}_{\text{mix}}$ defined in (32) we have that

$$I(\mathbf{X}_{\text{mix}}, \mathsf{snr}) = I\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + I(\mathbf{X}_G, \mathsf{snr}\ \delta), \quad (34a)$$

$$\mathrm{mmse}(\mathbf{X}_{\text{mix}}, \mathsf{snr}) = \frac{1-\delta}{(1+\mathsf{snr}\delta)^2}\mathrm{mmse}\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + \delta\ \mathrm{mmse}(\mathbf{X}_G, \mathsf{snr}\ \delta). \quad (34b)$$

Proof: See Appendix C. ■

Observe that Proposition 8 implies that, in order for mixed inputs (with $\delta < 1$) to comply with the MMSE constraint

$$0 \le \Delta_{(26a)} = \frac{1}{2}\log\left(\frac{1+\mathsf{snr}_0}{1+\mathsf{snr}_L}\right) - \frac{1}{2}\frac{\beta(\mathsf{snr}_0-\mathsf{snr}_L)}{1+\beta\mathsf{snr}_0} - \frac{(n+2)}{2}\log\left(\frac{\mathsf{snr}_0}{\mathsf{snr}_L}\right) + \frac{(n+2)(\mathsf{snr}_0-\mathsf{snr}_L)}{2\mathsf{snr}_0} + \frac{1}{2}\left((2\mathsf{snr}_L+1)\log\left(\frac{\mathsf{snr}_0(1+\mathsf{snr}_L)}{\mathsf{snr}_L(1+\mathsf{snr}_0)}\right) - \frac{\mathsf{snr}_0-\mathsf{snr}_L}{1+\mathsf{snr}_0} - \frac{\mathsf{snr}_0-\mathsf{snr}_L}{\mathsf{snr}_0}\right) = O\left(\frac{1}{n}\right), \quad (26a)$$

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

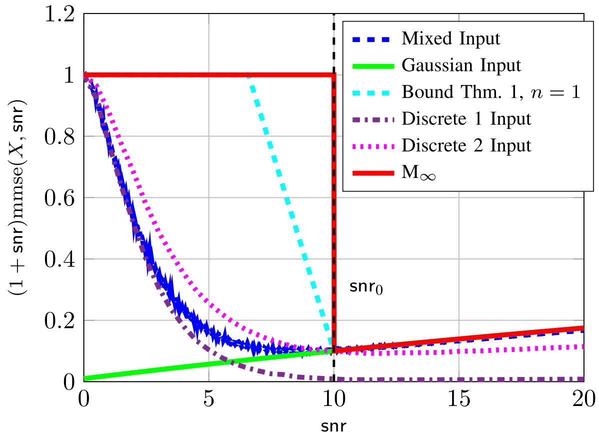


Fig. 6. Upper and lower bounds on $\mathrm{M}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ vs. snr, for $\beta = 0.01$ and $\mathsf{snr}_0 = 10$.

in (5c) and (14c), the MMSE of $\mathbf{X}_D$ must satisfy

$$\mathrm{mmse}\left(\mathbf{X}_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) \le \frac{(\beta-\delta)(1+\delta\mathsf{snr}_0)}{(1-\delta)(1+\beta\mathsf{snr}_0)}. \quad (35)$$

The bound in (35) will be helpful in choosing the parameter δ later on.

When X_D is a scalar discrete random variable with $\mathsf{supp}(X_D) = N$ we use the following bounds from [31, Appendix C] and [13, Remark 2].

Proposition 9 [13], [31]: For a discrete random variable X_D such that $p_i = \Pr(X_D = x_i)$, for $i \in [1:N]$, we have that

$$\mathrm{mmse}(X_D, \mathsf{snr}) \le d_{\max}^2 \sum_{i=1}^{N} p_i \mathrm{e}^{-\frac{\mathsf{snr}}{8} d_i^2}, \quad (36a)$$

$$I(X_D, \mathsf{snr}) \ge H(X_D) - \frac{1}{2}\log\left(\frac{\pi}{6}\right) - \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}(X_D, \mathsf{snr})\right), \quad (36b)$$

where

$$d_\ell := \min_{x_i \in \mathsf{supp}(X_D): i \ne \ell} |x_\ell - x_i|, \quad (36c)$$

$$d_{\min} := \min_{\ell \in [1:N]} d_\ell, \quad (36d)$$

$$d_{\max} := \max_{x_k, x_i \in \mathsf{supp}(X_D)} |x_k - x_i|. \quad (36e)$$

Proposition 8 and Proposition 9 are particularly useful because they will allow us to design the 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 (15) (solid red line);
- The upper bound from Theorem 1 in (23c) with finite power (dashed cyan line);
- The Gaussian-only input lower bound (solid 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 (32). To obtain this bound we used Proposition 8 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 (35). That is, we chose δ such that the power of X_G is approximately β while the MMSE constraint on X_D in (35) 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 $X_D = [-1.8412, -1.7386, 0.5594]$ and $P_X = [0.1111, 0.1274, 0.7615]$, 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 $X_D = [-1.4689, -1.1634, 0.7838]$ and $P_X = [0.1282, 0.2542, 0.6176]$, 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 $\mathsf{snr} \ge \mathsf{snr}_0$, 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 $\mathsf{snr} \le \mathsf{snr}_0$, we see that discrete-only inputs (brown dashed-dotted line and magenta dotted line) achieve higher MMSE values 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 (36a)) to comply with the MMSE constraint. However, for $\mathsf{snr} \ge \mathsf{snr}_0$, 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 (36a). 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 $\mathsf{snr} \ge \mathsf{snr}_0$ (without any reduction in power) and the behavior of discrete-only inputs for $\mathsf{snr} \le \mathsf{snr}_0$. 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 $\mathsf{snr} \le \mathsf{snr}_0$ but lower than the mixed input for $\mathsf{snr} \ge \mathsf{snr}_0$. This is again due to the tail behavior of the MMSE of discrete inputs. The

TABLE I
PARAMETERS OF THE MIXED INPUT IN (32) USED IN THE PROOF OF PROPOSITION 10

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

reason this second discrete input is not used as a component of the mixed input is because this choice would violate the MMSE constraint on X_D in (35). 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 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.

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 (32) is to within an additive gap of the outer bound in Proposition 7. The case $n > 1$ is more involved and will be treated in Section II-E and Section II-F.

Proposition 10: A lower bound on $\mathcal{C}_1(\mathsf{snr}, \mathsf{snr}_0, \beta)$ *with the mixed input in* (32)*, with* $X_D \sim \mathsf{PAM}(N)$ *and with input parameters as specified in Table I, is to within* $O\left(\log\log\left(\frac{1}{\mathrm{mmse}(X,\mathsf{snr}_0)}\right)\right)$ *of the outer bound in Proposition 7 with the exact gap value given by*

$$\mathsf{snr} \geq \mathsf{snr}_0 \geq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\text{mix}}, \mathsf{snr}) \leq \mathsf{gap}_{1,1}, \tag{37a}$$

$$\mathsf{snr}_0 \geq \mathsf{snr} \geq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\text{mix}}, \mathsf{snr}) \leq \mathsf{gap}_{1,2}, \tag{37b}$$

$$\mathsf{snr} \leq 1 : C_1(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_1(X_{\text{mix}}, \mathsf{snr}) \leq \mathsf{gap}_{1,3}, \tag{37c}$$

where

$$\mathsf{gap}_{1,1} = \frac{1}{2}\log\left(\frac{2}{3}\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0)}{\beta}\right) + \frac{6\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right) - \Delta_{(26a)}, \tag{37d}$$

$$\mathsf{gap}_{1,2} = \frac{1}{2}\log\left(1 + \frac{2}{3}\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)\right) + \frac{1}{2}\log\left(\frac{4\pi}{6}\right) - \Delta_{(26b)}, \tag{37e}$$

$$\mathsf{gap}_{1,3} = \frac{1}{2}\log(2), \tag{37f}$$

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

Proof: See Appendix D. ■

Please note that the gap result in Proposition 10 is constant in snr (i.e., independent of snr) but not in snr_0.

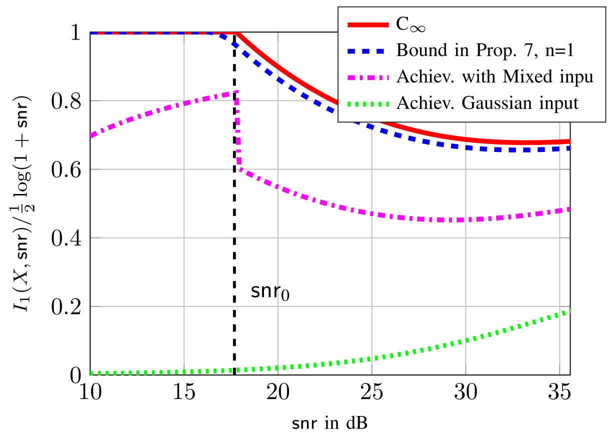


Fig. 7. Upper and lower bounds on $\mathcal{C}_{n=1}(\mathsf{snr}, \mathsf{snr}_0, \beta)$ from Proposition 7 vs. snr, for $\beta = 0.001$ and $\mathsf{snr}_0 = 60 = 17.6815$ dB.

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 10 to

- The $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ upper bound in (6), (solid red line);
- The upper bound from Proposition 7 (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 $\mathrm{mmse}(X, \mathsf{snr}_0) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}$ (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 $\mathsf{snr} \leq \mathsf{snr}_0$, 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 [2] 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.

Next we focus on the case of $n \to \infty$.

E. Max-I Problem: Achievability for $\mathcal{C}_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$

Before examining the general case of $n > 1$ we first focus on the easier case of $n \to \infty$. To extend our achievable result for $n = 1$ to $n > 1$ we need to extend the bounds in Proposition 9. The extension of the bounds in Proposition 9 was recently done in [32] and is given next.

Proposition 11 [32]: For any discrete random vector $\mathbf{X}_D \in \mathbb{R}^n$ *we have that*

$$\mathrm{mmse}(\mathbf{X}_D, \mathsf{snr}) \leq \frac{d_{\max}^2(\mathbf{X}_D)}{n} P_e^{(n)}(\mathsf{snr}), \tag{38a}$$

where $P_e^{(n)}$ is the probability of decoding error

$$P_e^{(n)}(\mathsf{snr}) = \mathbb{P}[\mathbf{X}_D \neq \hat{\mathbf{X}}_D]. \tag{38b}$$

For the mutual information we have

$$I_n(\mathbf{X}_D, \mathsf{snr}) \geq \frac{1}{n} H(\mathbf{X}_D) - G_1(\mathbf{X}_D, \mathsf{snr}) - G_2, \tag{38c}$$

where

$$G_1(\mathbf{X}_D, \mathsf{snr}) = \frac{1}{2}\log\left(1 + \frac{4(2+n)\cdot \mathrm{mmse}(\mathbf{X}_D, \mathsf{snr})}{d_{\min}^2(\mathbf{X}_D)}\right), \tag{38d}$$

$$G_2 \leq \frac{1}{2}\log\left(\frac{2\mathrm{e}}{n}\Gamma^{\frac{2}{n}}\left(\frac{n}{2}+1\right)\right) = O\left(\frac{1}{n}\log(n)\right). \tag{38e}$$

The bound in (38c) is called the Ozarow-Wyner bound [33]. For recent applications of the bound in (38c) to non-Gaussian and MIMO channels the reader is referred to [34]–[36].

By using Proposition 11 and mimicking the proof of Proposition 10 we have the following:

Proposition 12: For

$$\mathsf{snr}_0 \leq \mathsf{snr} : C_n(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_n(\mathbf{X}_{\mathrm{mix}}, \mathsf{snr}) \leq \mathsf{gap}_{n,1}, \tag{39a}$$

where $\mathbf{X}_D$ and δ are chosen to satisfy the MMSE constraint in (5c), we have that

$$\mathsf{gap}_{n,1} = \mathsf{gap}_{e,1} + G_1\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + G_2 - \Delta_{(26\mathrm{a})}, \tag{39b}$$

$$\mathsf{gap}_{e,1} = \frac{1}{2}\log\left(1 + \frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}\right) - \frac{1}{n}H(\mathbf{X}_D) + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}}\right). \tag{39c}$$

For

$$\mathsf{snr} \leq \mathsf{snr}_0 : C_n(\mathsf{snr}, \mathsf{snr}_0, \beta) - I_n(\mathbf{X}_D, \mathsf{snr}) \leq \mathsf{gap}_{n,2}, \tag{39d}$$

where $\mathbf{X}_D$ (note that we have set $\delta = 0$ in $\mathbf{X}_{\mathrm{mix}}$) is chosen to satisfy the MMSE constraint in (5c), we have that

$$\mathsf{gap}_{n,2} = \mathsf{gap}_{e,2} + G_1(\mathbf{X}_D, \mathsf{snr}) + G_2 - \Delta_{(26\mathrm{a})}, \tag{39e}$$

$$\mathsf{gap}_{e,2} = \frac{1}{2}\log(1+\mathsf{snr}) - \frac{1}{n}H(\mathbf{X}_D). \tag{39f}$$

Proof: The proof follows by taking the difference between the upper bound in Proposition 7 and the inner bound in Proposition 8 where the $I_n(\mathbf{X}_D, \gamma)$ term has been lower bounded using Proposition 11. ■

We see that the key term in Proposition 12 is $G_1(\mathbf{X}_D, \gamma)$ in (39b) and (39e). This is so because $\mathsf{gap}_{e,i}$ in (39b) and (39f) depends only on the size of the support of $\mathbf{X}_D$ but not on the support itself (i.e., the positions of the points are irrelevant). Moreover, G_2 and $\Delta_{(26\mathrm{a})}$ are bounded and vanish for large n. However, unlike $\mathsf{gap}_{e,i}$, G_2 and $\Delta_{(26\mathrm{a})}$, the term $G_1(\mathbf{X}_D, \gamma)$ is highly sensitive to the geometry of the input $\mathbf{X}_D$ through the minimum distance and the MMSE. In particular, by using the bound in (38a) we have that

$$\mathrm{e}^{2G_1(\mathbf{X}_D,\gamma)} \leq 1 + \frac{4(2+n)d_{\max}^2(\mathbf{X}_D)}{n\cdot d_{\min}^2(\mathbf{X}_D)} P_e^{(n)}(\gamma). \tag{40}$$

TABLE II

Parameters of the Mixed Input in (32) Used in the Proof of Proposition 13

Regime	Input Parameters
Weak Interference ($\mathsf{snr} \geq \mathsf{snr}_0$)	$N = \sqrt{1 + \frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}}$, $\delta = \beta$.
Strong Interference ($\mathsf{snr} \leq \mathsf{snr}_0$)	$N = \sqrt{1+\mathsf{snr}}$, $\delta = 0$.

Next we show that if the geometry of $\mathbf{X}_D$ has a lattice structure then we can achieve the upper bound in (6) as $n \to \infty$.

Proposition 13: By taking the $\mathbf{X}_D$ part of $\mathbf{X}_{\mathrm{mix}}$ to be a lattice constellation with input parameters given in Table II we have that

$$\lim_{n\to\infty} \mathsf{gap}_{n,i} = 0,\ i \in [1:2]. \tag{41}$$

That is, as $n \to \infty$ the mixed input achieves the capacity upper bound in (6).

Proof: Let $\mathbf{X}_D$ be a lattice constellation with a codebook given by $\mathcal{C}$ and $d_{\max} = 2\sqrt{n}$. From [37] we know that for any $\gamma > 0$ there exists $\mathbf{X}_D$ such that

$$\lim_{n\to\infty} \frac{1}{n}H(\mathbf{X}_D) = \frac{1}{2}\log|\mathcal{C}| = \frac{1}{2}\log(1+\gamma), \tag{42}$$

$$\lim_{n\to\infty} P_e^{(n)}(\gamma) = 0. \tag{43}$$

Therefore, for the case of $\mathsf{snr} \leq \mathsf{snr}_0$ by using Proposition 12 and inputs as in Table II, by (40) and (43) we have that

$$\lim_{n\to\infty} G_1(\mathbf{X}_D, \mathsf{snr}) = 0. \tag{44}$$

Moreover, by using the fact that the MMSE is a decreasing function of snr and the bound in (38a) we have that

$$\lim_{n\to\infty} \mathrm{mmse}(\mathbf{X}_D, \mathsf{snr}_0) \leq \lim_{n\to\infty} \mathrm{mmse}(\mathbf{X}_D, \mathsf{snr}) \leq \lim_{n\to\infty} \frac{d_{\max}^2(\mathbf{X}_D)}{n} P_e^{(n)}(\mathsf{snr}) = 0, \tag{45}$$

which implies that the MMSE constraint in (5c) is always satisfied. This demonstrates that $\lim_{n\to\infty} \mathsf{gap}_{n,1} = 0$.

Using similar reasoning for the case of $\mathsf{snr} \geq \mathsf{snr}_0$ we know that there exists $\mathbf{X}_D$ with input parameters specified in Table II such that

$$\lim_{n\to\infty} P_e^{(n)}\left(\frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) = 0, \tag{46}$$

where in (43) we have taken $\gamma = \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}$. Moreover, since $\mathsf{snr} \geq \mathsf{snr}_0$ we have that

$$\lim_{n\to\infty} P_e^{(n)}\left(\frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) \leq \lim_{n\to\infty} P_e^{(n)}\left(\frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) = 0, \tag{47}$$

and therefore by the bounds in (38a) and (40) we have that

$$\lim_{n\to\infty} \text{mmse}\left(\mathbf{X}_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) = 0, \tag{48}$$

$$\lim_{n\to\infty} G_1\left(\mathbf{X}_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) = 0, \tag{49}$$

this implies that the MMSE constraint in (5c) is satisfied and demonstrates that $\lim_{n\to\infty} \mathsf{gap}_{n,2} = 0$. This concludes the proof. ■

The above result demonstrates that when $\mathbf{X}_D$ has a lattice structure, the upper bound $C_\infty(\mathsf{snr}, \mathsf{snr}_0, \beta)$ is achievable. The result in Proposition 13 also gives an alternative proof of the result in [2].

F. Max-I Problem: Achievability for $C_n(\mathsf{snr}, \mathsf{snr}_0, \beta)$

In this section we discuss how our results can be extended to an arbitrary and finite n. For simplicity we focus only on the case $\mathsf{snr}_0 \geq \mathsf{snr}$. To that end we will need the following bound on the MMSE from [32].

Proposition 14 [32, Proposition 14]: For any $\mathbf{X}_D$

$$\text{mmse}(\mathbf{X}_D, \mathsf{snr}) \leq \frac{d^2_{\max}(\mathbf{X}_D)}{n} P_e^{(n)}(\mathsf{snr}), \tag{50a}$$

$$P_e^{(n)}(\mathsf{snr}) \leq \bar{Q}\left(\frac{n}{2}; \frac{\mathsf{snr}\ d^2_{\min}(\mathbf{X}_D)}{8}\right), \tag{50b}$$

where

$$\bar{Q}(x; a) := \frac{\Gamma(x; a)}{\Gamma(x)}. \tag{50c}$$

In particular, by using the bounds in (40) and (50) we have that

$$e^{2G_1(\mathbf{X}_D,\mathsf{snr})} \leq 1 + \frac{4(n+2)d^2_{\max}(\mathbf{X}_D)}{n\ d^2_{\min}(\mathbf{X}_D)} \bar{Q}\left(\frac{n}{2}; \frac{\mathsf{snr}\ d^2_{\min}(\mathbf{X}_D)}{8}\right). \tag{51}$$

By recalling the following well known limits [38] on $\bar{Q}(x, a)$ for any $p \in \mathbb{R}$:

$$\lim_{x\to\infty} x^p \bar{Q}(x; (1+\epsilon)x) = \lim_{x\to\infty} x^p \frac{\Gamma(x; (1+\epsilon)x)}{\Gamma(x)} = 0, \tag{52a}$$

$$\lim_{x\to\infty} \bar{Q}(x; (1-\epsilon)x) = \lim_{x\to\infty} \frac{\Gamma(x; (1-\epsilon)x)}{\Gamma(x)} = 1, \tag{52b}$$

we see that in order to force $G_1(\mathbf{X}_D, \mathsf{snr})$ in (38d) to be small, it is sufficient to simultaneously satisfy the following two constraints:

$$\frac{d^2_{\max}(\mathbf{X}_D)}{n \cdot d^2_{\min}(\mathbf{X}_D)} = O(n^p), \quad p \in \mathbb{R}, \tag{53}$$

$$\frac{n}{2} \leq \frac{\mathsf{snr}\ d^2_{\min}(\mathbf{X}_D)}{8}. \tag{54}$$

Remark 1: Note that unlike the case of $n = 1$, for the case $n > 1$, using a cubic constellation, which is the Cartesian product of a PAM constellation with itself n-times, will not work well. This is so because if $X_D \sim \mathsf{PAM}(N)$, then for $\mathbf{X}_D$, which is an n fold Cartesian product of X_D, we have that

$$d_{\min}(\mathbf{X}_D) = d_{\min}(X_D),$$
$$d_{\max}(\mathbf{X}_D) = \sqrt{n(N-1)}d_{\min}(X_D),$$

which implies that $d_{\min}(\mathbf{X}_D)$ is independent of n and we cannot satisfy the condition in (54). The above discussion suggests that a lattice structure on $\mathbf{X}_D$ as in Proposition 13 might be necessary to satisfy the condition in (54).

Proposition 15 (Minkowski-Hlawka-Sigel Theorem [37]): For every n and N, there exists a lattice constellation $\mathbf{X}_D$ in $\mathbb{R}^n$ of size N contained in the ball of radius r centered at the origin such that

$$d_{\min}(\mathbf{X}_D) \geq \frac{r}{N^{\frac{2}{n}}}. \tag{55}$$

From Proposition 15 it is not hard to see that by taking $r = \sqrt{n}$, to comply with the power constraint, we have that

$$d_{\min}(\mathbf{X}_D) \geq \frac{\sqrt{n}}{N^{\frac{2}{n}}}, \tag{56}$$

and therefore, with an appropriately chosen N we can satisfy (54) and make $G_1(\mathbf{X}_D, \mathsf{snr})$ in (38d) as small as we want. This intuition is made clear in the following result.

Proposition 16: There exits an input $\mathbf{X}_{\text{mix}}$ with

$$N = \lfloor (c\ \mathsf{snr})^{\frac{n}{2}} \rfloor,\ \delta = 0,\ d^2_{\max}(\mathbf{X}_D) \leq 4n, \tag{57a}$$

where c is chosen to satisfy

$$4\bar{Q}\left(\frac{n}{2}; \frac{n}{8c}\right) \leq 4\bar{Q}\left(\frac{n}{2}; \frac{n\ \mathsf{snr}_0}{8\ \mathsf{snr} c}\right) \leq \frac{\beta}{1+\beta\mathsf{snr}_0}, \tag{57b}$$

such that for $1 \leq \mathsf{snr} \leq \mathsf{snr}_0$

$$\mathsf{gap}_1 \leq \mathsf{gap}_{e,1} + G_1(\mathbf{X}_D, \mathsf{snr}) + G_2, \tag{57c}$$

$$\mathsf{gap}_{e,1} \leq \frac{1}{2}\log(2) + \frac{\log(2)}{n} - \frac{1}{2}\log(c), \tag{57d}$$

$$G_1(\mathbf{X}_D, \mathsf{snr}) \leq \frac{1}{2}\log\left(1 + \frac{16(2+n)\ c\ \mathsf{snr}\ \bar{Q}\left(\frac{n}{2}; \frac{n}{8c}\right)}{n}\right). \tag{57e}$$

Proof: The choice of c in (57b) ensures that the MMSE constraint in (5c) is satisfied. Using Proposition 12 we have that

$$\begin{aligned}
\mathsf{gap}_{e,1} &= \frac{1}{2}\log(1+\mathsf{snr}) - \frac{1}{n}H(\mathbf{X}_D) \\
&= \frac{1}{2}\log(1+\mathsf{snr}) - \frac{1}{n}\log(\lfloor (c\ \mathsf{snr})^{\frac{n}{2}} \rfloor) \\
&\overset{a)}{\leq} \frac{1}{2}\log\left(\frac{1+\mathsf{snr}}{\mathsf{snr}}\right) - \frac{1}{2}\log(c) + \frac{\log(2)}{n} \\
&\overset{b)}{\leq} \frac{1}{2}\log(2) - \frac{1}{2}\log(c) + \frac{\log(2)}{n},
\end{aligned}$$

where the inequalities follow from: a) using the bound $\lfloor x \rfloor \geq \frac{x}{2}$ for $x \geq 1$; and b) using $\frac{1+\mathsf{snr}}{\mathsf{snr}} \geq 2$ for $\mathsf{snr} \geq 1$.

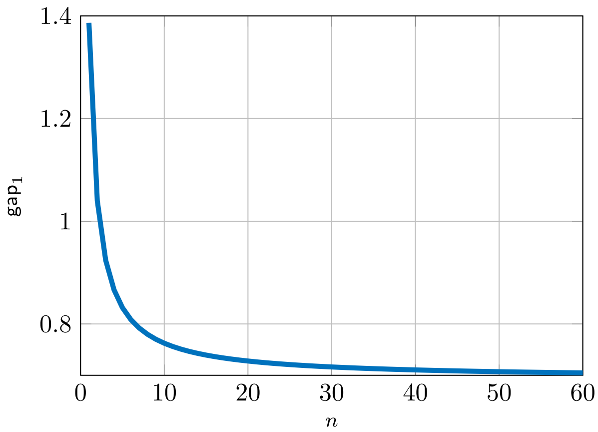


Fig. 8. Plot of the gap_1 in Proposition 16 vs. block length n for $\mathsf{snr} = 5$, $\mathsf{snr}_0 = 10$ and $\beta = 0.5$.

Moreover,

$$\begin{aligned} & e^{2G_1(\mathbf{X}_D,\mathsf{snr})} \\ & \le 1 + \frac{4(n+2)d_{\max}^2(\mathbf{X}_D)}{n\ d_{\min}^2(\mathbf{X}_D)} \bar{Q}\left(\frac{n}{2}; \frac{\mathsf{snr}\ d_{\min}^2(\mathbf{X}_D)}{8}\right) \\ & \overset{a)}{\le} 1 + \frac{16(n+2)\ c\ \mathsf{snr}}{n} \bar{Q}\left(\frac{n}{2}; \frac{\mathsf{snr}\ d_{\min}^2(\mathbf{X}_D)}{8}\right) \\ & \overset{b)}{\le} 1 + \frac{16(n+2)\ c\ \mathsf{snr}}{n} \bar{Q}\left(\frac{n}{2}; \frac{n}{8c}\right), \end{aligned}$$

where the inequalities follow from: a) using the fact that $d_{\max}^2(\mathbf{X}_D) = 4n$ and the lower bound on $d_{\min}(\mathbf{X}_D)$ in Proposition 15; and b) using the facts that $\bar{Q}(x; y)$ is a decreasing function of y and that $\frac{\mathsf{snr}\ d_{\min}^2(\mathbf{X}_D)}{8} \ge \frac{n}{8c}$ by Proposition 15.

This concludes the proof. ■

A plot of gap_1 in Proposition 16 is given in Fig. 8. It is interesting to note that Proposition 16 recovers only a weaker version (i.e., the gap is not zero) of Proposition 13.

Corollary 1: As $n \to \infty$ for $1 \le \mathsf{snr} \le \mathsf{snr}_0$

$$\lim_{n\to\infty} \mathsf{gap}_{n,1} = \frac{1}{2}\log(8) \approx 1.039. \quad (58)$$

Proof: The proof follows by taking $c = \frac{1}{4}$ in Proposition 16 and using the limits in (52), from which we have that $\lim_{n\to\infty} \mathsf{gap}_{n,1} = \frac{1}{2}\log(8)$. ■

Remark 2: The reason why the result in Corollary 1 does not match that of Proposition 13 exactly but only to within a gap is because the MMSE bound in (50) *is not tight enough. With a more careful bounding of the MMSE one can improve the bound in* (50) *to*

$$P_e^{(n)}(\mathsf{snr}) \le \min_r \bar{Q}\left(\frac{n}{2}; \frac{r^2}{2}\right) + \mathbb{P}\left[N_{\mathcal{B}\left(\frac{\mathbf{Z}}{\mathsf{snr}}, r\right)} > 1\right], \quad (59)$$

where $N_{\mathcal{B}(\mathbf{Z},r)}$ is the number of points in the support of $\mathbf{X}_D$ that fall into the ball $\mathcal{B}\left(\frac{\mathbf{Z}}{\mathsf{snr}}, r\right)$ centered at $\frac{\mathbf{Z}}{\mathsf{snr}}$ and with radius r. However, for any given constellation $\mathbf{X}_D$ the second term $\mathbb{P}\left[N_{\mathcal{B}\left(\frac{\mathbf{Z}}{\mathsf{snr}}, r\right)} > 1\right]$ can be quite difficult to analyze. To avoid this complication we chose a sub-optimal value of $r = \frac{d_{\min}(\mathbf{X}_D)}{2}$ so that $\mathbb{P}\left[N_{\mathcal{B}\left(\frac{\mathbf{Z}}{\mathsf{snr}}, r\right)} > 1\right] = 0$.

III. 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

$$\begin{aligned} -\frac{d\mathrm{mmse}(\mathbf{X}, \mathsf{snr})}{d\mathsf{snr}} &= \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]\right) \\ &:= \frac{1}{n}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}, \mathsf{snr})\right]\right), \quad (60) \end{aligned}$$

which was given in [22, Proposition 9] for $n = 1$ and in [23, Lemma 3] for $n \ge 1$. The first derivative in (60) turns out to be instrumental in proving Theorem 1 as well.

For ease of presentation, in the rest of this 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 [22, Proposition 5] that

$$\mathbb{E}\left[\mathrm{Cov}^2(X|Y)\right] \le \frac{k_1}{\mathsf{snr}^2}, \text{ where } k_1 \le 3 \cdot 2^4. \quad (61)$$

Before using (60) in the proof of Theorem 1, we will need to sharpen the existing constant for $n = 1$ in (61) (given by $k_1 \le 3 \cdot 2^4$) and generalize the bound to any $n \ge 1$, which to the best of our knowledge has not been considered before.

Proposition 17: For any $\mathbf{X}$ and $\mathsf{snr} > 0$ we have

$$\frac{1}{n}\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})]\right) \le \frac{k_n}{\mathsf{snr}^2}, \quad (62a)$$

where

$$k_n \le \frac{n(n+2) - n\ \mathrm{mmse}(\mathbf{Z}\mathbf{Z}^T|\mathbf{Y}) - \mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right)}{n} \le n + 2. \quad (62b)$$

Proof: See Appendix E. ■

In Proposition 17 the bound on k_1 in (61) has been tightened from $k_1 \le 3 \cdot 2^4$ in (61) 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 18: 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

$$\mathrm{Tr}(\mathbf{J}^2(\mathbf{Y})) \ge \frac{n}{(1+\mathsf{snr})^2}. \quad (63)$$

Equality in (63) *is achieved when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.*

Proof: See Appendix F. ■

Observe that, by using the bound in (62) from Proposition 17 together with the lower bound on the Fisher information in Proposition 18, the bound on the constant k_n in (62b) can be tightened to

$$k_n \le \frac{n(n+2) - \frac{n}{(1+\mathsf{snr})^2}}{n} = n + 2 - \frac{1}{(1+\mathsf{snr})^2}. \quad (64)$$

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 [22, 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) (60), and b) the bound in Proposition 17 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 (64), to obtain

$$0 \leq \Delta = \Delta_{(23c)} = \int_{\mathsf{snr}}^{\mathsf{snr}_0} \frac{1}{\gamma^2(1+\gamma)^2} d\gamma .$$

This concludes the proof of Theorem 1.

IV. 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. We also demonstrated how our result can be generalized to vector inputs. 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.

We would also like to mention that the bound on the phase transition region in Proposition 6 has been recently improved in [32] from $O(\frac{1}{n})$ to $O(\frac{1}{\sqrt{n}})$ by using a notion of minimum mean p-th error that generalizes the notion of MMSE. An interesting future direction would be to relate $\mathcal{C}_n(\mathsf{snr}_0, \mathsf{snr}, \beta)$, which does not have an operational meaning of capacity, to non-asymptotic information theory results [39].

Appendix A
Proof of Proposition 4

For the case of $n = 1$ consider an input distribution given by

$$X_a = [-a, 0, a], \quad P_{X_a} = \left[\frac{1}{2a^2}, 1 - \frac{1}{a^2}, \frac{1}{2a^2}\right], \tag{65}$$

for any $a \geq 1$. Note that for the input distribution in (65) $\mathbb{E}[X_a^2] = 1$ for any a. The MMSE of X_a can be upper bounded by

$$\mathrm{mmse}(X_a, \mathsf{snr}) \leq \min\left(1, 4(a^2+1)e^{-\frac{a^2\mathsf{snr}}{8}}\right), \tag{66}$$

where the upper bound in (66) follows by applying the upper bound in Proposition 9 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.

The case of $n > 1$ is straightforward generalization using the bound in (50). This concludes the proof.

Appendix B
Proof of Proposition 6

In order to find the point of intersection snr_L between (12a) and (23a) 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

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

and the width of the phase transition is given by

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

as claimed in (25b). This concludes the proof.

Appendix C
Proof of Proposition 8

We first show the decomposition for mutual information with mixed inputs in (32)

$$\begin{aligned}
I(X_{\mathrm{mix}}, \mathsf{snr}) &= I(X_{\mathrm{mix}}; Y) = I(X_G, X_D; Y) \\
&= I(X_D; Y) + I(X_G; Y|X_D) \\
&= I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) + I(X_G, \mathsf{snr}\delta). \quad (67)
\end{aligned}$$

Next we take the derivative of both sides of (67) with respect to snr. On the left side we get $\frac{d}{d\mathsf{snr}} I(X_{\mathrm{mix}}, \mathsf{snr}) = \frac{1}{2}\mathrm{mmse}(X_{\mathrm{mix}}, \mathsf{snr})$ and on the right we get

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

as claimed in (34). This concludes the proof.

Appendix D
Proof of Proposition 10

By letting $X_D \sim \mathsf{PAM}(N)$, given the bound in Proposition 9 and the requirement in (35) we further constrain the MMSE of X_D to satisfy

$$\begin{aligned} \mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right) &\le d_{\max}^2 \mathrm{e}^{-\frac{\frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}}{8} d_{\min}^2} \\ &\le \frac{(1+\mathsf{snr}_0\delta)(\beta-\delta)}{(1-\delta)(1+\beta\mathsf{snr}_0)}, \end{aligned} \quad (68)$$

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

$$N \le \left\lfloor \sqrt{1 + c_1 \frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}} \right\rfloor, \quad (69a)$$

$$\begin{aligned} c_1 &= \frac{3}{2\log^+\left(\frac{d_{\max}^2(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)} \\ &\le \frac{3}{2\log^+\left(\frac{12(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)}, \end{aligned} \quad (69b)$$

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

$$d_{\max}^2 = (N-1)^2 d_{\min}^2 = 12\frac{(N-1)^2}{N^2-1} = 12\frac{N-1}{N+1} \le 12. \quad (70)$$

For the case of $\mathsf{snr}_0 \le \mathsf{snr}$ we choose the number of points to satisfy (69) 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 7 with the achievable mutual information of a mixed input in Proposition 8, where $I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)$ is lower bounded by Proposition 9.

We obtain

$$\begin{aligned} &\mathsf{gap}_1 + \Delta_{(26a)} \\ &= \mathcal{C}_\infty - I\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right) - I(X_G, \mathsf{snr}\ \delta) \\ &= C_\infty - \left(\log(N) - \frac{1}{2}\log\left(\frac{\pi}{6}\right)\right. \\ &\quad \left. - \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right)\right) \\ &\quad - \frac{1}{2}\log(1+\delta\mathsf{snr}) \\ &\overset{a)}{\le} C_\infty - \left(\frac{1}{2}\log\left(1 + c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}\right) - \log(2)\right. \\ &\quad - \frac{1}{2}\log\left(\frac{\pi}{6}\right) \\ &\quad - \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right) \\ &\quad \left. + \frac{1}{2}\log(1+\delta\mathsf{snr})\right) \\ &= \frac{1}{2}\log\left(\frac{1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}}{1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}}\right) + \frac{1}{2}\log\left(\frac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}}\right) \\ &\quad + \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right) \\ &\quad + \frac{1}{2}\log\left(\frac{4\pi}{6}\right), \end{aligned} \quad (71)$$

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

We next bound each term in (71) individually. The first term in (71) can be bounded as follows:

$$\begin{aligned} &\frac{1}{2}\log\left(\frac{1+\frac{\mathsf{snr}_0(1-\beta)}{1+\beta\mathsf{snr}_0}}{1+c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}}\right) \\ &= \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+1c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0-\beta c_1 c_2\mathsf{snr}_0)}\right) \\ &\overset{b)}{\le} \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0-\beta c_1\mathsf{snr}_0)}\right) \\ &= \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)(1+c_2\beta\mathsf{snr}_0)}{(1+\beta\mathsf{snr}_0)(1+(1-\beta)c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0)}\right) \\ &\overset{c)}{\le} \frac{1}{2}\log\left(\frac{(1+\mathsf{snr}_0)}{(1+(1-\beta)c_1\mathsf{snr}_0+\beta c_2\mathsf{snr}_0)}\right) \\ &\overset{d)}{\le} \frac{1}{2}\log\left(\max\left(\frac{(1+\mathsf{snr}_0)}{(1+c_1\mathsf{snr}_0)}, \frac{(1+\mathsf{snr}_0)}{(1+c_2\mathsf{snr}_0)}\right)\right) \\ &\overset{e)}{\le} \frac{1}{2}\log\left(\max\left(\frac{1}{c_1}, 2\right)\right), \end{aligned} \quad (72)$$

where the inequalities follow from the facts: b) $c_2 = \frac{\mathsf{snr}_0}{1+\mathsf{snr}_0} \le 1$; c) $\frac{1+c_2\beta\mathsf{snr}_0}{1+\beta\mathsf{snr}_0} \le 1$ since $c_2 \le 1$; 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) $\frac{(1+\mathsf{snr}_0)}{(1+c_2\mathsf{snr}_0)} \le \frac{1}{c_2} = \frac{1+\mathsf{snr}_0}{\mathsf{snr}_0} \le 2$ for $\mathsf{snr}_0 \ge 1$.

The second term in (71) can be bounded as follows:

$$\tfrac{1}{2}\log\left(\tfrac{1+\beta\mathsf{snr}}{1+\delta\mathsf{snr}}\right) \le \tfrac{1}{2}\log\left(\tfrac{1+\mathsf{snr}_0}{\mathsf{snr}_0}\right) \le \tfrac{1}{2}\log(2), \quad (73)$$

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 (71) can be bounded as follows:

$$\begin{aligned} &\frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}(1-\delta)}{1+\delta\mathsf{snr}}\right)\right) \\ &\overset{f)}{\le} \frac{1}{2}\log\left(1 + \frac{12}{d_{\min}^2}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right)\right) \\ &\overset{g)}{\le} \frac{1}{2}\log\left(1 + c_1\frac{(1-\delta)\mathsf{snr}_0}{1+\delta\mathsf{snr}_0}\mathrm{mmse}\left(X_D, \frac{\mathsf{snr}_0(1-\delta)}{1+\delta\mathsf{snr}_0}\right)\right) \\ &\overset{h)}{\le} \frac{1}{2}\log\left(1 + c_1\frac{(\beta-\delta)\mathsf{snr}_0}{1+\beta\mathsf{snr}_0}\right) \\ &\overset{i)}{\le} \frac{1}{2}\log\left(1 + c_1\frac{\beta}{1+\beta\mathsf{snr}_0}\right), \end{aligned} \quad (74)$$

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 (69);

h) using the bound in (68); 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 (72), (73), and (74) we get

$$\begin{aligned}
&2(\mathsf{gap}_2 + \Delta_{(26a)}) \\
&\le \log\left(\max\left(\frac{1}{c_1},2\right)\right) + \log\left(\frac{4\pi}{3}\right) \\
&\quad + \log\left(1 + c_1 \frac{\beta}{1+\beta\mathsf{snr}_0}\right) \\
&= \log\left(\max\left(\frac{1}{c_1},2\right) + 2\max(1,2c_1)\frac{\beta}{1+\beta\mathsf{snr}_0}\right) \\
&\quad + \log\left(\frac{4\pi}{3}\right) \\
&\overset{j)}{\le} \log\left(\max\left(\frac{1}{c_1},2\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) + \frac{1}{2}\log\left(\frac{4\pi}{3}\right) \\
&\overset{k)}{=} \log\left(\max\left(\frac{2\log\left(\frac{12(1-\delta)(1+\beta\mathsf{snr}_0)}{(1+\mathsf{snr}_0\delta)(\beta-\delta)}\right)}{3},2\right) + \frac{6\beta}{1+\beta\mathsf{snr}_0}\right) \\
&\quad + \log\left(\frac{4\pi}{3}\right) \\
&\overset{l)}{\le} \log\left(\max\left(\frac{2\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0)}{\beta}\right)}{3},2\right) + \frac{6\beta}{1+\beta\mathsf{snr}_0}\right) \\
&\quad + \log\left(\frac{4\pi}{3}\right) \\
&\overset{m)}{=} \log\left(\frac{2}{3}\log\left(\frac{24(1+(1-\beta)\mathsf{snr}_0)}{\beta}\right) + 6\frac{\beta}{1+\beta\mathsf{snr}_0}\right) \\
&\quad + \log\left(\frac{4\pi}{3}\right),
\end{aligned}$$

where the inequalities follow from: j) the fact that $c_1 \le \frac{3}{2}$; k) using the value of c_1 in (69); 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 $\mathsf{snr} \ge \mathsf{snr}_0$ regime.

We next focus on the $1 \le \mathsf{snr} \le \mathsf{snr}_0$ regime. We use only the discrete part of the mixed input and set $\delta = 0$. From (69) we have that the input parameters must satisfy

$$N \le \left\lfloor\sqrt{1+c_3\mathsf{snr}_0}\right\rfloor, \tag{75a}$$

$$c_3 \le \frac{3}{2\log\left(\frac{12(1+\beta\mathsf{snr}_0)}{\beta}\right)}, \tag{75b}$$

in order to comply with the MMSE constraint in (5c). However, instead of choosing the number of points as in (75) we choose it to be

$$N = \left\lfloor\sqrt{1+c_3\mathsf{snr}}\right\rfloor \le \left\lfloor\sqrt{1+c_3\mathsf{snr}_0}\right\rfloor. \tag{76}$$

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 7 and the achievable mutual information of the mixed input in Proposition 8, where $I(X_D,\mathsf{snr})$ is lower bounded using Proposition 9. We have,

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

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 (76) and bound $\mathrm{mmse}(X,\mathsf{snr}) \le \frac{1}{1+\mathsf{snr}}$; p) using that $\frac{1+(1+c_3)\mathsf{snr}}{1+c_3\mathsf{snr}} \le \frac{1+c_3}{c_3} = 1+\frac{1}{c_3}$; and r) using the value of c_3 from (75).

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

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

$$\begin{aligned}
\mathsf{gap}_3 &\le \mathcal{C}(\beta,\mathsf{snr},\mathsf{snr}_0) - I(X_{\mathrm{mix}},\mathsf{snr}) \\
&\le \mathcal{C}(\beta,\mathsf{snr},\mathsf{snr}_0) \le \frac{1}{2}\log(1+\mathsf{snr}) \\
&\le \frac{1}{2}\log(2).
\end{aligned} \tag{77}$$

This concludes the proof.

Appendix E
Proof of Proposition 17

We will need the following identities for the proof:

$$\mathsf{snr}\cdot\mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbf{Cov}(\mathbf{Z}|\mathbf{Y})], \tag{78a}$$

$$\mathsf{snr}^2\cdot\mathbb{E}[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})] = \mathbb{E}[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})], \tag{78b}$$

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

$$\begin{aligned}
&\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y}) \\
&= \left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\right)^2 \\
&= (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - \mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] \\
&\quad - \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}} + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{a)}{=} (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - 2\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}] \\
&\quad + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{b)}{\preceq} (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - 2\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}} \\
&\quad + (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&= (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2 - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2 \\
&\overset{c)}{=} \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}] - \mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}) - (\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2,
\end{aligned} \tag{79}$$

where the order operations follow from: a) the fact that $\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}$ and $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}]$ are symmetric matrices; b) using $\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}} \preceq \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}]$ (from the positive semi-definite property of the conditional covariance matrix); and c) the fact that, since $\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}) = \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}] - \mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}](\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^{\mathrm{T}}$ and by symmetry of $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}]$, we have that $\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}](\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^{\mathrm{T}} = (\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}])^2$. By using the monotonicity of the trace, properties of the expected value, and the inequality in (79), we have that

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbb{E}[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})]\right) \\
&\quad \le \mathrm{Tr}\Big(\mathbb{E}\Big[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}] - \mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}) \\
&\qquad -(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2\Big]\Big) \\
&\quad = \mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}]\right]\right) - \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y})\right]\right) \\
&\qquad - \mathrm{Tr}\left(\mathbb{E}\left[(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2\right]\right).
\end{aligned} \tag{80}$$

We next focus on each term of the right hand side of (80) individually. The first term can be computed as follows:

$$\begin{aligned}
\mathrm{Tr}\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}})^{\mathrm{T}}|\mathbf{Y}]\right]\right) &\overset{d)}{=} \mathrm{Tr}\left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}]\right) \\
&\overset{e)}{=} \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\right)\right] \\
&= \mathbb{E}\left[\mathrm{Tr}\left(\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}\right)\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{n} Z_i^2\right)^2\right] \\
&\overset{f)}{=} n(n+2),
\end{aligned} \tag{81}$$

where the 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 (80), by definition of the MMSE, we have

$$\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{Cov}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y})\right]\right) = n\mathrm{mmse}(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|\mathbf{Y}). \tag{82}$$

The third term in (80) satisfies

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbb{E}\left[(\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}})^2\right]\right) \\
&\overset{g)}{\ge} \mathrm{Tr}\left(\left(\mathbb{E}\left[\mathbb{E}[\mathbf{Z}|\mathbf{Y}]\mathbb{E}[\mathbf{Z}|\mathbf{Y}]^{\mathrm{T}}\right]\right)^2\right) \\
&= \mathrm{Tr}\left(\left(\mathbb{E}[\mathbf{Z}\mathbf{Z}^{\mathrm{T}}] - \mathbb{E}[\mathbf{Cov}(\mathbf{Z}|\mathbf{Y})]\right)^2\right) \\
&\overset{h)}{=} \mathrm{Tr}\left((\mathbf{I} - \mathsf{snr}\ \mathbb{E}[\mathbf{Cov}(\mathbf{X}|\mathbf{Y})])^2\right) \\
&\overset{i)}{=} \mathrm{Tr}\left(\mathbf{J}^2(\mathbf{Y})\right),
\end{aligned} \tag{83}$$

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 (78); and i) using the identity [22]

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

By putting (81), (82), and (83) together, we have that

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

Finally, using the identity $\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{Z}|\mathbf{Y})\right] = \mathsf{snr}^2\mathbb{E}\left[\mathbf{Cov}^2(\mathbf{X}|\mathbf{Y})\right]$ in (78) concludes the proof.

Appendix F
Proof of Proposition 18

Using the Cramér-Rao lower bound [40, Th. 20] we have that

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

where Λ 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 the matrix $\mathbb{E}[\mathbf{X}\mathbf{X}^T]$. Therefore,

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

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.

REFERENCES

[1] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, H. V. Poor, and S. Shamai (Shitz), "On communications through a Gaussian noise channel with an MMSE disturbance constraint," in *Proc. Workshop Inf. Theory Appl.*, Jan. 2016, pp. 1–8.

[2] R. Bustin and S. Shamai (Shitz), "MMSE of 'bad' codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 733–743, Feb. 2013.

[3] S. Shamai (Shitz), "From constrained signaling to network interference alignment via an information-estimation perspective," *IEEE Inf. Theory Soc. Newslett.*, vol. 62, no. 7, pp. 6–24, Sep. 2012.

[4] R. H. Etkin, D. N. C. 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.

[5] 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, Feb. 2009.

[6] 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, Feb. 2009.

[7] 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, Jul. 2009.

[8] 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.

[9] 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.

[10] A. Bennatan, S. Shamai (Shitz), 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, Apr. 2014.

[11] K. Moshksar, A. Ghasemi, and A. K. 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.

[12] 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, Mar. 2015.

[13] A. Dytso, D. Tuninetti, and N. Devroye, "Interference as noise: Friend or foe?" *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3561–3596, Jun. 2016.

[14] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Mar. 1973, pp. 23–52.

[15] S. Gherekhloo, A. Chaaban, and A. Sezgin, "Expanded GDOF-optimality regime of treating interference as noise in the $m \times 2$ X-channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 355–376, Jan. 2017.

[16] R. S. Cheng and S. Verdú, "On limiting characterizations of memoryless multiuser capacity regions," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 609–612, Mar. 1993.

[17] E. Abbe and L. Zheng, "A coordinate system for Gaussian networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 721–733, Feb. 2012.

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

[19] R. Blasco-Serrano, R. Thobaben, and M. Skoglund, "Communication and interference coordination," in *Proc. Workshop Inf. Theory Appl.*, Feb. 2014, pp. 1–8.

[20] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[21] D. Guo, S. Shamai (Shitz), and S. Verdú, *The Interplay Between Information and Estimation Measures*. Breda, The Netherlands: Now Publishers, 2013.

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

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

[24] R. Bustin, R. F. Schaefer, H. V. Poor, and S. Shamai (Shitz), "On MMSE properties of optimal codes for the Gaussian wiretap channel," in *Proc. IEEE Inf. Theory Workshop*, Apr. 2015, pp. 1–5.

[25] N. Merhav, D. Guo, and S. Shamai (Shitz), "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, Mar. 2010.

[26] 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, Apr. 2014.

[27] 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, Mar. 2012.

[28] 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.

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

[30] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[31] 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, Jul. 2006.

[32] A. Dytso, R. Bustin, D. Tuninetti, N. Devroye, S. Shamai (Shitz), and H. V. Poor, "On the minimum mean p-th error in Gaussian noise channels and its applications," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: https://arxiv.org/abs/1607.01461

[33] L. H. Ozarow and A. D. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1426–1428, Nov. 1990.

[34] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "A generalized Ozarow-Wyner capacity bound with applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1058–1062.

[35] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai (Shitz), "On additive channels with generalized Gaussian noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 426–430.

[36] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf.*, Singapore, Dec. 2017.

[37] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[38] W. Gautschi, "The incomplete gamma functions since Tricomi," in *Tricomi's Ideas and Contemporary Applied Mathematics, Atti dei ConvegniLincei*, vol. 147. Rome, Italy: Accademia dei Lincei, 1998.

[39] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

Alex Dytso is currently a Postdoctoral Researcher in the Department of Electrical Engineering at Princeton University. In 2016, he received a Ph.D. degree from the Department of Electrical and Computer Engineering at the University of Illinois, Chicago. He received his B.S. degree in 2011 from the University of Illinois, Chicago, where he also received the International Engineering Consortium's William L. Everitt Student Award of Excellence for outstanding seniors. His current research interest are in the areas of multi-user information theory and estimation theory, and their applications in wireless networks.

Ronit Bustin (S'09–M'14) received a B.Sc. degree in Computer Science and Electrical Engineering and a M.Sc. degree in Electrical Engineering in 2004 and 2006, respectively, from Tel-Aviv University, Israel. She received a Ph.D. in Electrical Engineering in 2013 from the Technion - Israel Institute of Technology, Haifa. From 2013 to 2015 she was a postdoctoral research associate in the Department of Electrical Engineering at Princeton University. During 2015-2016 she was a post-doctoral research associate in the Department of Electrical Engineering at Tel Aviv University. During 2016-2017 she was a post-doctoral research associate in the Department of Electrical Engineering at the Technion. She currently holds a research position in General Motors - Advanced Technical Center in Israel. Her research interests include multi-user information theory, secrecy constraints, Gaussian MIMO channels, estimation theory, channel coding, interactive communication and communication complexity. Dr. Bustin received the Irwin and Joan Jacobs scholarship for excellence in graduate studies and research, in January 2010. She is a recipient of the Adams fellowship from the Israel Academy of Sciences and Humanities, April 2010, and an Andrew and Erna Finci Viterbi graduate fellow in the faculty of electrical engineering at the Technion for the fall semester 2010-2011. She received the Rothschild fellowship in 2013 and the women postdoctoral scholarship of Israel's Council for Higher Education (VATAT) for her postdoctoral studies at Princeton.

Daniela Tuninetti is currently a Professor within the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in 2005. Dr. Tuninetti received her Ph.D. in Electrical Engineering in 2002 from ENST/Telecom ParisTech (Paris, France, with work done at the Eurecom Institute in Sophia Antipolis, France), and she was a postdoctoral research associate at the School of Communication and Computer Science at the Swiss Federal Institute of Technology in Lausanne (EPFL, Lausanne, Switzerland) from 2002 to 2004. Dr. Tuninetti is a recipient of a best paper award at the European Wireless Conference in 2002, of an NSF CAREER award in 2007, and named UIC University Scholar in 2015. Dr. Tuninetti was the editor-in-chief of the IEEE Information Theory Society Newsletter from 2006 to 2008, an editor for IEEE COMMUNICATION LETTERS from 2006 to 2009, and for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2011 to 2014; she is currently an associate editor for the IEEE TRANSACTIONS ON INFORMATION THEORY. Dr. Tuninetti's research interests are in the ultimate performance limits of wireless interference networks (with special emphasis on cognition and user cooperation), coexistence between radar and communication systems, multirelay networks, content-type coding, and caching systems.

Natasha Devroye is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in January 2009. From July 2007 until July 2008 she was a Lecturer at Harvard University. Dr. Devroye obtained her Ph.D in Engineering Sciences from the School of Engineering and Applied Sciences at Harvard University in 2007, and a Honors B. Eng in Electrical Engineering from McGill University in 2001. Dr. Devroye was a recipient of an NSF CAREER award in 2011 and was named UIC's Researcher of the Year in the "Rising Star" category in 2012. She has been an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and is currently an Associate Editor for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING and the IEEE TRANSACTIONS ON INFORMATION THEORY. Her research focuses on multi-user information theory and applications to cognitive and software-defined radio, radar, relay, zero-error and two-way communication networks.

H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in electrical engineering and computer science from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering. During 2006 to 2016, he served as Dean of Princeton's School of Engineering and Applied Science. He has also held visiting appointments at several other institutions, most recently at Berkeley and Cambridge. His research interests are in the areas of information theory and signal processing, and their applications in wireless networks, energy systems and related fields. Among his publications in these areas is the recent book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Royal Society. In 1990, he served as President of the IEEE Information Theory Society, in 2004-07 as the Editor-in-Chief of these TRANSACTIONS, and in 2009 as General Co-chair of the IEEE International Symposium on Information Theory, held in Seoul, South Korea. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes the 2016 John Fritz Medal, the 2017 IEEE Alexander Graham Bell Medal, Honorary Professorships from Peking University and Tsinghua University, both conferred in 2017, and a D.Sc. *honoris causa* from Syracuse University awarded in 2017.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow, an URSI Fellow, a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award, the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering and the 2017 IEEE Richard W. Hamming Medal.

He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 and 2015 European Commission FP7, Network of Excellence in Wireless COMmunications Best Paper Awards, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, the 2014 EURASIP Best Paper Award (for the *EURASIP Journal on Wireless Communications and Networking*), and the 2015 IEEE Communications Society Best Tutorial Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has also served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY and on the IEEE Information Theory Society Nominations and Appointments Committee.