The Gaussian Source-and-Data-Streams Problem

Shraga I. Bross, *Senior Member, IEEE*, and Amos Lapidoth, *Fellow, IEEE*

***Abstract*— A Gaussian source and two data streams are to be transmitted over a Gaussian broadcast channel: the first stream, the "common stream," is to be decoded by both receivers, and the second, the "private stream," only by the strong receiver. Both receivers wish to estimate the source sequence, though with possibly different mean squared-errors. The quadruples of achievable rates and estimation errors are characterized, and it is shown that—once the data rates have been fixed—there is no tension between the estimation errors. Only the "equal bandwidth" case is treated, where the rate at which the source emits symbols is also the rate at which the channel is used.**

***Index Terms*— Gaussian broadcast channel, Gaussian source, mean squared-error, source-channel coding.**

I. Introduction

A Memoryless Gaussian source and two independent data streams are to be transmitted over an average-power limited one-to-two Gaussian Broadcast Channel (BC): one data stream, the "common stream," is to be decoded reliably by both receivers, and the second, the "private stream," only by the strong receiver. Both receivers wish to estimate the source sequence, with possibly different maximally-allowed mean squared-error (MSE) distortions. Here we characterize the achievable quadruples of data rates and distortions as a function of the allocated transmit power P and the noise variances N_1 and N_2 experienced by the two receivers.

When the maximally-allowed distortions at both receivers exceed the source's variance, the all-zero estimator is admissible, and our problem reduces to that of finding the capacity region $\mathcal{C}_{\text{G-BC}}$ of the Gaussian BC, a problem which is solved, for example, in [9, Sec. 15.1.3] or [10, Sec. 5.5]. Denoting the rate of the common-stream R_{c} and the rate of the private-stream R_1, this region comprises the rate pairs (R_{c}, R_1) that simultaneously satisfy

$$R_1 \le \frac{1}{2}\log\left(1+\frac{\alpha P}{N_1}\right) \tag{1a}$$

$$R_{\text{c}} \le \frac{1}{2}\log\left(1+\frac{(1-\alpha)P}{\alpha P+N_2}\right) \tag{1b}$$

for some $0 \le \alpha \le 1$. Here $N_1 > 0$ is the variance of the noise experienced by the stronger receiver, and N_2 is the variance of the noise experienced by the weaker receiver, so

$$N_2 \ge N_1 > 0. \tag{2}$$

At the other extreme, if no data are to be transmitted whence both data rates are zero, then the problem reduces to determining the least MSE distortions that can be achieved when sending a Gaussian source over an average-power limited Gaussian BC. These distortions were found by Goblick [11], who showed that the least distortions are achieved by uncoded transmission. Our result can thus be viewed as unifying Goblick's result [11] and the Cover-Bergmans [2], [8] capacity region.

We emphasize that we only treat the "equal bandwidth case," where the rate at which the source emits symbols (in source-symbols per second) is equal to the rate at which the BC is used (in channel-uses per second). The problem under the "bandwidth expansion setting," i.e., when the number of channel uses per source-sample exceeds one, is more complicated and is as yet not fully solved; see [13] and [17].

We also emphasize that we only treat scalar sources. The bivariate version of our problem with zero-rate data streams, i.e., the bivariate version of Goblick's setting, is discussed in [4], [21]. There it is shown that uncoded transmission is optimal only for some values of the power and noise variances. This makes the treatment of our problem in the bivariate setting more elaborate.

Our interest in the source-and-data-streams problem is related to recent explorations into the feasibility of upgrading existing communications systems that currently broadcast analog signals (such as television or radio signals) to allow them to also downstream digital data without significantly degrading the reception of the analog content. Our results can be viewed as information-theoretic limits on the performance of such systems. This setting has previously been studied by Zhao and Chen [23]. In fact, our converse is very similar to theirs, and our plausibility argument of Section IV-A is very similar to their achievability sketch [23, Sec. III-A.]. For the reasons we outline in Section IV ahead, a rigorous proof of achievability is more intricate.

The paper is organized as follows. In Section II we provide a formal statement of our problem and present our main result. Section III proves the converse and Section IV the achievability.

II. Problem Statement and Main Result

We adopt the following convention. Random variables are denoted by upper-case letters and their realizations by the

Manuscript received December 23, 2018; revised April 11, 2019; accepted April 17, 2019. Date of publication May 1, 2019; date of current version August 14, 2019. The work of S. Bross was supported by the Israel Science Foundation under Grant 455/14. The associate editor coordinating the review of this paper and approving it for publication was C. Tian. *(Corresponding author: Shraga I. Bross.)*

S. I. Bross is with the Faculty of Engineering, Bar-Ilan University, Ramat Gan 52900, Israel (e-mail: brosss@biu.ac.il).

A. Lapidoth is with ETH Zurich, 8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

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/TCOMM.2019.2914384

corresponding lower-case letter. A generic realization of the random variable X is hence denoted x. Random vectors are denoted by bold upper-case letters and their realizations by the corresponding bold lower-case letter. Their dimension is usually implicit. Thus, $\mathbf{X}$ denotes the random vector $(X_1, X_2, \ldots, X_n)$, and $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denotes its realization. The set in which a message takes values is denoted using a caligraphic font: the common message W_{c} takes values in the set $\mathcal{W}_{\mathrm{c}}$ and the private message W_1 in $\mathcal{W}_1$. The set of real numbers is denoted $\mathbb{R}$, and its n-fold Cartesian power $\mathbb{R}^n$. If α is in the interval $[0, 1]$, then we sometimes write $\bar{\alpha}$ for $1 - \alpha$:

$$\bar{\alpha} \triangleq 1 - \alpha, \quad \alpha \in [0, 1]. \tag{3}$$

Using a Gaussian broadcast channel n times, we wish to transmit an n-tuple $\mathbf{S}$ of source symbols $S_1, \ldots, S_n$, which are independent and identically distributed (IID) centered Gaussians of variances $\sigma^2 > 0$, as well as a pair of messages $W = (W_{\mathrm{c}}, W_1)$ that is drawn independently of $\mathbf{S}$ uniformly over the set $\mathcal{W}_{\mathrm{c}} \times \mathcal{W}_1$, where

$$\mathcal{W}_{\mathrm{c}} = \{1, \ldots, 2^{nR_{\mathrm{c}}}\} \quad \text{and} \quad \mathcal{W}_1 = \{1, \ldots, 2^{nR_1}\};$$

W_{c} is the "common message;" W_1 is the "private message;" and their corresponding rates are R_{c} and R_1. An encoder for our setting is thus a mapping

$$\varphi^{(n)} \colon \mathbb{R}^n \times \mathcal{W}_{\mathrm{c}} \times \mathcal{W}_1 \to \mathbb{R}^n, \tag{4}$$

where the set of reals $\mathbb{R}$ is the source's alphabet as well as the BC's input and output alphabets. Applying the encoder $\varphi^{(n)}$ to $(\mathbf{S}, W_{\mathrm{c}}, W_1)$ yields the n-tuple $\mathbf{X}$ comprising the n channel inputs $X_1, \ldots, X_n$:

$$\mathbf{X} = \varphi^{(n)}(\mathbf{S}, W_{\mathrm{c}}, W_1). \tag{5}$$

The channel inputs are subjected to an average-power constraint

$$\frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\big[X_k^2\big] \le P, \tag{6}$$

where $\mathsf{E}[\cdot]$ denotes the expectation operator (in this case with respect to W_{c}, W_1, and $\mathbf{S}$). This constraint can be expressed in terms of $\mathbf{X}$'s Euclidean norm $\|\cdot\|$ as

$$\frac{1}{n} \mathsf{E}\big[\|\mathbf{X}\|^2\big] \le P. \tag{7}$$

When $\mathbf{X}$ is transmitted, the strong receiver observes the n-tuple $\mathbf{Y}_1 \in \mathbb{R}^n$ that is given by

$$\mathbf{Y}_1 = \mathbf{X} + \mathbf{Z}_1, \tag{8}$$

where $\mathbf{Z}_1$ is a random n-vector whose components are IID $\mathcal{N}(0, N_1)$, where N_1 is positive and $\mathcal{N}(\mu, \sigma^2)$ denotes the mean-μ variance-σ^2 univariate Gaussian distribution. Based on $\mathbf{Y}_1$, the strong receiver must guess the message pair $W = (W_{\mathrm{c}}, W_1)$ and estimate the source sequence $\mathbf{S}$. It performs the former task by applying some decoding rule

$$\phi_W^{(1)} \colon \mathbb{R}^n \to \mathcal{W}_{\mathrm{c}} \times \mathcal{W}_1 \tag{9}$$

to produce the guess

$$(\hat{W}_{\mathrm{c}}^{(1)}, \hat{W}_1^{(1)}) = \phi_W^{(1)}(\mathbf{Y}_1) \tag{10}$$

with resulting average probability of error

$$P_{\mathrm{e}}^{(1)} = \Pr\Big[(\hat{W}_{\mathrm{c}}^{(1)}, \hat{W}_1^{(1)}) \neq (W_{\mathrm{c}}, W_1)\Big]. \tag{11}$$

In order to estimate the source sequence, it applies some estimation rule

$$\phi_S^{(1)} \colon \mathbb{R}^n \to \mathbb{R}^n \tag{12}$$

to produce the estimate

$$\hat{\mathbf{S}}_1 = \phi_S^{(1)}(\mathbf{Y}_1) \tag{13}$$

with resulting average MSE distortion

$$\frac{1}{n} \mathsf{E}\Big[\|\mathbf{S} - \hat{\mathbf{S}}_1\|^2\Big].$$

The weaker receiver observes the n-tuple

$$\mathbf{Y}_2 = \mathbf{X} + \mathbf{Z}_2, \tag{14}$$

where $\mathbf{Z}_2$ is a random n-vector whose components are IID $\mathcal{N}(0, N_2)$, with

$$N_2 \ge N_1 > 0. \tag{15}$$

It too wishes to estimate $\mathbf{S}$ but, unlike the strong receiver, it only wishes to guess the common message W_{c}. It does so using some decoding rule

$$\phi_{W_{\mathrm{c}}}^{(2)} \colon \mathbb{R}^n \to \mathcal{W}_{\mathrm{c}} \tag{16}$$

to produce the guess $\hat{W}_{\mathrm{c}}^{(2)}$

$$\hat{W}_{\mathrm{c}}^{(2)} = \phi_{W_{\mathrm{c}}}^{(2)}(\mathbf{Y}_2) \tag{17}$$

with resulting average probability of error

$$P_{\mathrm{e}}^{(2)} = \Pr\Big[\hat{W}_{\mathrm{c}}^{(2)} \neq W_{\mathrm{c}}\Big]. \tag{18}$$

Like the strong receiver, it forms its estimate $\hat{\mathbf{S}}_2$ of $\mathbf{S}$ by applying some estimation rule

$$\phi_S^{(2)} \colon \mathbb{R}^n \to \mathbb{R}^n \tag{19}$$

to produce the estimate

$$\hat{\mathbf{S}}_2 = \phi_S^{(2)}(\mathbf{Y}_2) \tag{20}$$

with resulting average MSE distortion

$$\frac{1}{n} \mathsf{E}\Big[\|\mathbf{S} - \hat{\mathbf{S}}_2\|^2\Big].$$

The encoder, the broadcast channel, and the decoders are depicted in Figure 1.

Definition 1: The tuple $(R_{\mathrm{c}}, R_1, D_1, D_2)$ is *achievable* if, for every $\varepsilon > 0$, there exist for all sufficiently-large block-lengths n a power-P-limited encoder $\varphi^{(n)}$ whose rates exceed $(R_{\mathrm{c}} - \varepsilon, R_1 - \varepsilon)$ and decoding/estimation mappings $\big(\phi_W^{(1)}, \phi_S^{(1)}\big)$ and $\big(\phi_{W_{\mathrm{c}}}^{(2)}, \phi_S^{(2)}\big)$ such that

$$\overline{\lim_{n \to \infty}} \; \frac{1}{n} \mathsf{E}\big[\big\|\mathbf{S} - \hat{\mathbf{S}}_\nu\big\|^2\big] \le D_\nu + \varepsilon, \quad \nu = 1, 2 \tag{21a}$$

and

$$\lim_{n\to\infty}\big(P_{\mathrm{e}}^{(1)}+P_{\mathrm{e}}^{(2)}\big)=0. \tag{21b}$$

Remark 1: Since imposing additional source reconstruction constraints cannot help, the achievability of the tuple $(R_{\mathrm{c}},R_1,D_1,D_2)$ implies that the rates (R_{c},R_1) must lie in the capacity region $\mathcal{C}_{\text{G-BC}}$ of the Gaussian BC.

Remark 2: The achievability of a quadruple depends only on the BC's marginals, i.e., on the distributions of $(\mathbf{X},\mathbf{Y}_1)$ and $(\mathbf{X},\mathbf{Y}_2)$. We shall therefore henceforth assume w.l.g. that the BC is physically degraded so

$$\mathbf{Z}_2=\mathbf{Z}_1+\tilde{\mathbf{Z}}_2, \tag{22}$$

where $\tilde{\mathbf{Z}}_2$ is a Gaussian n-vector that is independent of $(W_{\mathrm{c}},W_1,\mathbf{X},\mathbf{Z}_1)$ and whose components are IID $\mathcal{N}(0,N_2-N_1)$.

Define the signal-to-noise ratios

$$\mathsf{SNR}_\nu=\frac{P}{N_\nu},\quad \nu\in\{1,2\} \tag{23}$$

and define

$$g=\frac{N_2}{N_1}. \tag{24}$$

With these definitions we can now state our main result.

Theorem 2: A quadruple $(R_{\mathrm{c}},R_1,D_1,D_2)$ is achievable if, and only if, all three of the following conditions hold:

$$(R_{\mathrm{c}},R_1)\in\mathcal{C}_{\text{G-BC}} \tag{25a}$$
$$D_1\geq\sigma^2\,D_{1,\min}(R_{\mathrm{c}},R_1) \tag{25b}$$
$$D_2\geq\sigma^2 D_{2,\min}(R_{\mathrm{c}},R_1), \tag{25c}$$

where

$$D_{1,\min}(R_{\mathrm{c}},R_1)=\frac{2^{2R_1}}{(\mathsf{SNR}_1+g)2^{-2R_{\mathrm{c}}}-(g-1)} \tag{26}$$

and

$$D_{2,\min}(R_{\mathrm{c}},R_1)=\frac{2^{2(R_{\mathrm{c}}+R_{\mathrm{d}})}}{\mathsf{SNR}_2+1}, \tag{27}$$

where R_{d}, or $R_{\mathrm{d}}(R_1)$, is defined as (*c.f.* [1])

$$R_{\mathrm{d}}=\frac{1}{2}\log\Big(1+(2^{2R_1}-1)\frac{N_1}{N_2}\Big). \tag{28}$$

Remark 3: The theorem would also hold if we replace the limit superior with a limit inferior in our definition of achievability, i.e., in (21a) of Definition 1.

To prove Theorem 2, we need to show that Conditions (25) are necessary and sufficient. Necessity is proved in Section III and sufficiency in Section IV.

Remark 4: Conditions (25) are also necessary if the transmitter and the two receivers have access to a common source of randomness that is independent of the source and messages.

Proof: See Appendix A. ∎

III. Necessity

To prove necessity, fix some $\varepsilon>0$ and assume the existence of a sequence of encoders, decoders, and estimators as in Definition 1. For each blocklength n, denote the average MSE distortions achieved by the two receivers

$$\delta_n^{(\nu)}=\frac{1}{n}\,\mathsf{E}\Big[\big\|\mathbf{S}-\hat{\mathbf{S}}_\nu\big\|^2\Big],\quad \nu\in\{1,2\}, \tag{29}$$

and define

$$\varepsilon_n=\max\big\{P_{\mathrm{e}}^{(1)},P_{\mathrm{e}}^{(2)}\big\}. \tag{30}$$

The achievability of the quadruple $(R_{\mathrm{c}},R_1,D_1,D_2)$ implies that

$$\overline{\lim_{n\to\infty}}\,\delta_n^{(\nu)}\leq D_\nu+\varepsilon,\quad \nu\in\{1,2\} \tag{31}$$

and

$$\lim_{n\to\infty}\varepsilon_n=0. \tag{32}$$

By Fano's inequality we obtain as in [9, Eqs. (7.100)–(7.101)]

$$\begin{aligned} &I(\mathbf{Y}_1;W_1)\\ &\quad\geq n\big(R_1-P_{\mathrm{e}}^{(1)}R_1-n^{-1}\big) &\text{(33a)}\\ &\quad\geq n\big(R_1-\max\{P_{\mathrm{e}}^{(1)},P_{\mathrm{e}}^{(2)}\}\cdot\max\{R_1,R_{\mathrm{c}}\}-n^{-1}\big) &\text{(33b)}\\ &\quad= n\big(R_1-\eta_n\big), &\text{(33c)}\end{aligned}$$

where η_n is defined as

$$\eta_n=\max\{P_{\mathrm{e}}^{(1)},P_{\mathrm{e}}^{(2)}\}\cdot\max\{R_1,R_{\mathrm{c}}\}+n^{-1}$$

and therefore, by (32), tends to zero

$$\lim_{n\to\infty}\eta_n=0. \tag{34a}$$

Rearranging (33c) and repeating the argument in (33) with the substitution of $(W_{\mathrm{c}},R_{\mathrm{c}},\mathbf{Y}_2)$ for $(W_1,R_1,\mathbf{Y}_1)$, we obtain

$$n(R_1-\eta_n)\leq I(\mathbf{Y}_1;W_1) \tag{34b}$$
$$n(R_{\mathrm{c}}-\eta_n)\leq I(\mathbf{Y}_2;W_{\mathrm{c}}). \tag{34c}$$

To relate (31) to mutual informations, recall the rate-distortion function

$$R_{\text{Gau}}(\Delta)=\frac{1}{2}\log^+\left(\frac{\sigma^2}{\Delta}\right),\quad \Delta>0 \tag{35}$$

of a memoryless variance-σ^2 Gaussian source with respect to the MSE criterion [9, Sec. 10.3.2]. Here $\log^+(\xi)\triangleq\max\{\log\xi,0\}$ for all $\xi>0$. The converse to the Rate Distortion theorem, e.g., [9, Eqs. (10.61) and (10.71)], implies that

$$\begin{aligned} nR_{\text{Gau}}\big(\delta_n^{(\nu)}\big)&\leq I(\mathbf{S};\hat{\mathbf{S}}_\nu)\\ &\leq I(\mathbf{S};\mathbf{Y}_\nu),\quad \nu\in\{1,2\}, &\text{(36)}\end{aligned}$$

where the second inequality follows from the Data Processing inequality.

We shall next use (34) and (36) to establish necessity. Since the necessity of (25a) follows from Remark 1, we focus on (25b) and (25c), beginning with the former.

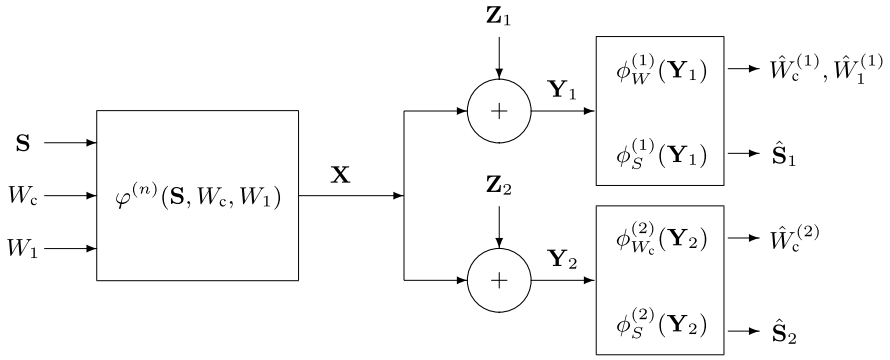

Fig. 1. Broadcasting a Gaussian source sequence $\mathbf{S}$ and a message pair $W = (W_c, W_1)$ over a Gaussian BC.

A. Necessity of (25b)

Starting from (36) and using the independence between $\mathbf{S}$ and (W_c, W_1),

$$\begin{aligned} nR_{\text{Gau}}\big(\delta_n^{(1)}\big) &\le I(\mathbf{S};\mathbf{Y}_1) \\ &\le I(\mathbf{S};\mathbf{Y}_1|W_c,W_1) \\ &= h(\mathbf{Y}_1|W_c,W_1) - h(\mathbf{Y}_1|\mathbf{S},W_c,W_1) \\ &= h(\mathbf{Y}_1|W_c,W_1) - h(\mathbf{Y}_1|\mathbf{X},\mathbf{S},W_c,W_1) \\ &= h(\mathbf{Y}_1|W_c,W_1) - h(\mathbf{Z}_1) \\ &= h(\mathbf{Y}_1|W_c,W_1) - \frac{n}{2}\log(2\pi e N_1), \end{aligned} \tag{37}$$

where the second equality holds because $\mathbf{X}$ is computable from $(\mathbf{S}, W_c, W_1)$.[1]

An upper bound on $h(\mathbf{Y}_1|W_c, W_1)$ will thus provide us with an upper bound on $R_{\text{Gau}}\big(\delta_n^{(1)}\big)$ and hence with a lower bound on $\delta_n^{(1)}$.

To derive such an upper bound, we first use Fano's inequality (34b) and the independence between W_1 and W_c to obtain

$$\begin{aligned} n(R_1 - \eta_n) &\le I(\mathbf{Y}_1; W_1) \\ &\le I(\mathbf{Y}_1; W_1|W_c) \\ &= h(\mathbf{Y}_1|W_c) - h(\mathbf{Y}_1|W_c,W_1) \end{aligned} \tag{38}$$

or

$$h(\mathbf{Y}_1|W_c,W_1) \le h(\mathbf{Y}_1|W_c) - n(R_1 - \eta_n). \tag{39}$$

We next upper-bound $h(\mathbf{Y}_1|W_c)$ in terms of $h(\mathbf{Y}_2|W_c)$ using (22), the independence between $\tilde{\mathbf{Z}}_2$ and $(W_c, W_1, \mathbf{X}, \mathbf{Z}_1)$, and the conditional Entropy-Power Inequality (EPI) [10, Sec. 2.2]:

$$\begin{aligned} 2^{\frac{2}{n}h(\mathbf{Y}_2|W_c)} &= 2^{\frac{2}{n}h(\mathbf{Y}_1+\tilde{\mathbf{Z}}_2|W_c)} \\ &\ge 2^{\frac{2}{n}h(\mathbf{Y}_1|W_c)} + 2^{\frac{2}{n}h(\tilde{\mathbf{Z}}_2)} \\ &= 2^{\frac{2}{n}h(\mathbf{Y}_1|W_c)} + 2\pi e(N_2 - N_1) \end{aligned} \tag{40}$$

or

$$2^{\frac{2}{n}h(\mathbf{Y}_1|W_c)} \le 2^{\frac{2}{n}h(\mathbf{Y}_2|W_c)} - 2\pi e(N_2 - N_1). \tag{41}$$

[1]This does not hold when the transmitter and receivers have access to a common source of randomness. Remark 4 addresses this scenario.

Inequality (41) and the monotonicity of the logarithmic function combine with (39) to yield

$$\begin{aligned} h(\mathbf{Y}_1|W_c,W_1) \le \frac{n}{2}\log\Big(2^{\frac{2}{n}h(\mathbf{Y}_2|W_c)} - 2\pi e(N_2 - N_1)\Big) \\ - n(R_1 - \eta_n). \end{aligned} \tag{42}$$

From this we complete the derivation of the upper bound on $h(\mathbf{Y}_1|W_c, W_1)$ by deriving an upper bound on $h(\mathbf{Y}_2|W_c)$ using Fano's inequality (34c):

$$\begin{aligned} n(R_c - \eta_n) &\le I(\mathbf{Y}_2; W_c) \\ &= h(\mathbf{Y}_2) - h(\mathbf{Y}_2|W_c) \\ &\le \frac{n}{2}\log\big(2\pi e(P + N_2)\big) - h(\mathbf{Y}_2|W_c) \end{aligned} \tag{43}$$

or

$$2^{\frac{2}{n}h(\mathbf{Y}_2|W_c)} \le 2\pi e(P + N_2)\, 2^{-2(R_c - \eta_n)}, \tag{44}$$

where (43) follows from the power constraint on $\mathbf{Y}_2$ that is induced by the power constraint on $\mathbf{X}$ and from the fact that the IID multivariate Gaussian distribution maximizes differential entropy subject to a power constraint [10, Eq. (2.8)].

From (44) and (42) we obtain the desired upper bound on $h(\mathbf{Y}_1|W_c, W_1)$:

$$\begin{aligned} &h(\mathbf{Y}_1|W_c,W_1) \\ &\quad\le \frac{n}{2}\log\Big(2\pi e(P + N_2)2^{-2(R_c - \eta_n)} - 2\pi e(N_2 - N_1)\Big) \\ &\qquad - n(R_1 - \eta_n). \end{aligned} \tag{45}$$

Equipped with this upper bound on $h(\mathbf{Y}_1|W_c, W_1)$, we return to (37) to obtain

$$\begin{aligned} &R_{\text{Gau}}\big(\delta_n^{(1)}\big) \\ &\quad\le \frac{1}{2}\log\Big(2\pi e(P + N_2)2^{-2(R_c - \eta_n)} - 2\pi e(N_2 - N_1)\Big) \\ &\qquad - (R_1 - \eta_n) - \frac{1}{2}\log(2\pi e N_1). \end{aligned} \tag{46}$$

Taking the limit superior over n and using (34a) and the monotonicity of $R_{\text{Gau}}(\cdot)$,

$$\begin{aligned} &R_{\text{Gau}}\Big(\underline{\lim}\,\delta_n^{(1)}\Big) \\ &\quad\le \frac{1}{2}\log\Big(2\pi e(P + N_2)2^{-2R_c} - 2\pi e(N_2 - N_1)\Big) \\ &\qquad - R_1 - \frac{1}{2}\log(2\pi e N_1). \end{aligned} \tag{47}$$

It thus follows from (31) and the monotonicity of $R_{\text{Gau}}(\cdot)$ that

$$\begin{aligned} &R_{\text{Gau}}(D_1+\varepsilon)\\ &\quad\le \frac{1}{2}\log\Big(2\pi e(P+N_2)2^{-2R_{\text{c}}}-2\pi e(N_2-N_1)\Big)\\ &\qquad -R_1-\frac{1}{2}\log(2\pi eN_1). \end{aligned} \tag{48}$$

Letting $\varepsilon \downarrow 0$ and using the continuity of $R_{\text{Gau}}(\cdot)$ on the positive reals,

$$\begin{aligned} R_{\text{Gau}}(D_1)\le \frac{1}{2}\log\Big((P+N_2)2^{-2R_{\text{c}}}-(N_2-N_1)\Big)\\ -R_1-\frac{1}{2}\log N_1. \end{aligned} \tag{49}$$

This[2] and the explicit expression (35) for $R_{\text{Gau}}(D_1)$ concludes the proof of the necessity of (25b).

It would have been possible to infer (48) from (46) also if we had replaced the limit superior in (31) with a limit inferior (*c.f.* Remark 3). In fact, it follows from (46) that for fixed (R_{c}, R_1),

$$\delta_n^{(1)} \ge \sigma^2\, D_{1,\min}(R_{\text{c}}, R_1) - \Psi^{(1)}(\varepsilon_n), \tag{50a}$$

where ε_n is defined in (30), and $\Psi^{(1)}(\cdot)$ is a nonnegative function (that depends only on $R_{\text{c}}, R_1, P, N_1, N_2, \sigma^2$ and not on the codebook) for which

$$\lim_{\varepsilon_n \downarrow 0} \Psi^{(1)}(\varepsilon_n) = 0, \tag{50b}$$

and which can be chosen to be monotonically nonincreasing.

B. Necessity of (25c)

Let $C_{X\to Y_2}$ denote the capacity of the Gaussian channel from X to Y_2

$$C_{X\to Y_2} = \frac{1}{2}\log\left(1+\frac{P}{N_2}\right). \tag{51}$$

Since the capacity is the maximum of mutual information,

$$\begin{aligned} &nC_{X\to Y_2}\\ &\quad\ge I(\mathbf{X};\mathbf{Y}_2)\\ &\quad= I(W_{\text{c}}, W_1, \mathbf{S};\mathbf{Y}_2)\\ &\quad= I(W_{\text{c}};\mathbf{Y}_2)+I(\mathbf{S};\mathbf{Y}_2|W_{\text{c}})+I(W_1;\mathbf{Y}_2|W_{\text{c}},\mathbf{S})\\ &\quad\ge I(W_{\text{c}};\mathbf{Y}_2)+I(\mathbf{S};\mathbf{Y}_2)+I(W_1;\mathbf{Y}_2|W_{\text{c}},\mathbf{S})\\ &\quad\ge n(R_{\text{c}}-\eta_n)+nR_{\text{Gau}}\big(\delta_n^{(2)}\big)+I(W_1;\mathbf{Y}_2|W_{\text{c}},\mathbf{S}), \end{aligned} \tag{52}$$

where the first equality holds because $\mathbf{X}$ is computable from $(W_{\text{c}}, W_1, \mathbf{S})$ (*c.f.* Footnote 1) and because, conditional on $\mathbf{X}$, the pair $\mathbf{Y}_2$ and $(W_{\text{c}}, W_1, \mathbf{S})$ are independent; the second equality follows from the chain rule for mutual information; the following inequality follows from the independence between $\mathbf{S}$ and W_{c}; and the final inequality follows from (34c) and (36). Thus

$$\begin{aligned} &nR_{\text{Gau}}\big(\delta_n^{(2)}\big)\\ &\quad\le nC_{X\to Y_2}-n(R_{\text{c}}-\eta_n)-I(W_1;\mathbf{Y}_2|W_{\text{c}},\mathbf{S})\\ &\quad= nC_{X\to Y_2}-n(R_{\text{c}}-\eta_n)-h(\mathbf{Y}_2|W_{\text{c}},\mathbf{S})\\ &\qquad+h(\mathbf{Y}_2|W_{\text{c}},W_1,\mathbf{S})\\ &\quad= nC_{X\to Y_2}-n(R_{\text{c}}-\eta_n)-h(\mathbf{Y}_2|W_{\text{c}},\mathbf{S})\\ &\qquad+\frac{n}{2}\log(2\pi eN_2). \end{aligned} \tag{53}$$

Using the conditional EPI, we can relate $h(\mathbf{Y}_2|W_{\text{c}},\mathbf{S})$ to $h(\mathbf{Y}_1|W_{\text{c}},\mathbf{S})$ and hence to $I(W_1;\mathbf{Y_1}|\mathbf{W}_{\text{c}},\mathbf{S})$ (because $h(\mathbf{Y}_1|W_{\text{c}},W_1,\mathbf{S})$ is simply $h(\mathbf{Z}_1)$):

$$\begin{aligned} &2^{\frac{2}{n}h(\mathbf{Y}_2|W_{\text{c}},\mathbf{S})}\\ &\quad= 2^{\frac{2}{n}h(\mathbf{Y}_1+\tilde{\mathbf{Z}}_2|W_{\text{c}},\mathbf{S})}\\ &\quad\ge 2^{\frac{2}{n}h(\mathbf{Y}_1|W_{\text{c}},\mathbf{S})}+2^{\frac{2}{n}h(\tilde{\mathbf{Z}}_2)}\\ &\quad= 2^{\frac{2}{n}h(\mathbf{Y}_1|W_{\text{c}},\mathbf{S})}+2\pi e(N_2-N_1)\\ &\quad= 2^{\frac{2}{n}\left[I(W_1;\mathbf{Y}_1|W_{\text{c}},\mathbf{S})+\frac{n}{2}\log(2\pi eN_1)\right]}+2\pi e(N_2-N_1)\\ &\quad= 2^{\frac{2}{n}I(W_1;\mathbf{Y}_1|W_{\text{c}},\mathbf{S})}2\pi eN_1+2\pi e(N_2-N_1). \end{aligned} \tag{54}$$

From (53) and (54) we obtain

$$\begin{aligned} R_{\text{Gau}}\big(\delta_n^{(2)}\big)\le C_{X\to Y_2}-R_{\text{c}}+\eta_n\\ -\frac{1}{2}\log\Big(\frac{2^{\frac{2}{n}I(W_1;\mathbf{Y}_1|W_{\text{c}},\mathbf{S})}N_1+(N_2-N_1)}{N_2}\Big). \end{aligned} \tag{55}$$

The right-hand side (RHS) can be further upper-bounded using the inequality

$$I(W_1;\mathbf{Y}_1|W_{\text{c}},\mathbf{S}) \ge n(R_1-\eta_n) \tag{56}$$

(which holds by Fano's inequality (34b) and the independence between W_1 and $(W_{\text{c}},\mathbf{S})$) to yield

$$\begin{aligned} R_{\text{Gau}}\big(\delta_n^{(2)}\big)\le C_{X\to Y_2}-R_{\text{c}}+\eta_n\\ -\frac{1}{2}\log\Big(\frac{2^{2(R_1-\eta_n)}N_1+(N_2-N_1)}{N_2}\Big). \end{aligned} \tag{57}$$

Since the RHS converges as n tends to infinity, and since η_n converges to zero,

$$\begin{aligned} \varlimsup_{n\to\infty} R_{\text{Gau}}\big(\delta_n^{(2)}\big)\le C_{X\to Y_2}-R_{\text{c}}\\ -\frac{1}{2}\log\Big(\frac{2^{2R_1}N_1+(N_2-N_1)}{N_2}\Big)\\ = C_{X\to Y_2}-R_{\text{c}}-R_{\text{d}}, \end{aligned} \tag{58}$$

where the equality follows from the definition of R_{d} (28). This and the monotonicity of $R_{\text{Gau}}(\cdot)$ implies that

$$R_{\text{Gau}}\Big(\underline{\lim}\,\delta_n^{(2)}\Big) \le C_{X\to Y_2}-R_{\text{c}}-R_{\text{d}}, \tag{59}$$

which, together with (31) and the monotonicity of $R_{\text{Gau}}(\cdot)$, establishes that

$$R_{\text{Gau}}(D_2+\varepsilon) \le C_{X\to Y_2}-R_{\text{c}}-R_{\text{d}}. \tag{60}$$

Since this is true for every $\varepsilon > 0$, we can take the limit as $\varepsilon \downarrow 0$ and use the continuity of $R_{\text{Gau}}(\cdot)$ at D_2 to establish that for all positive D_2

$$R_{\text{Gau}}(D_2) \le C_{X\to Y_2}-R_{\text{c}}-R_{\text{d}}. \tag{61}$$

[2]A necessary condition for the inequality $R_{\text{Gau}}(D_1)\le\xi$ to hold is $D_1 \ge \sigma^2 2^{-2\xi}$. (This is also sufficient if ξ is nonnegative.)

This (*c.f.* Footnote 2) in combination with (51), establishes (25c).

It would have been possible to infer (60) from (57) also if we had replaced the limit superior in (31) with a limit inferior (*c.f.* Remark 3). In fact, it follows from (57) that for fixed (R_{c}, R_1),

$$\delta_n^{(2)} \geq \sigma^2\, D_{2,\min}(R_{\mathrm{c}}, R_1) - \Psi^{(2)}(\varepsilon_n), \tag{62a}$$

where ε_n is defined in (30), and $\Psi^{(2)}(\cdot)$ is a nonnegative function (that depends only on $R_{\mathrm{c}}, R_1, P, N_1, N_2, \sigma^2$ and not on the codebook) for which

$$\lim_{\varepsilon_n \downarrow 0} \Psi^{(2)}(\varepsilon_n) = 0, \tag{62b}$$

and which can be chosen to be monotonically nonincreasing.

IV. Sufficiency

To establish sufficiency, we shall prove the following proposition:

Proposition 1: For any choice of $0 \leq \gamma, \beta \leq 1$, the tuple

$$\Big(R_{\mathrm{c}}(\gamma,\beta),\ R_1(\gamma,\beta),\ \sigma^2\, D_{1,\min}\big(R_{\mathrm{c}}(\gamma,\beta), R_1(\gamma,\beta)\big),\ \sigma^2\, D_{2,\min}\big(R_{\mathrm{c}}(\gamma,\beta), R_1(\gamma,\beta)\big)\Big)$$

is achievable, where

$$R_{\mathrm{c}}(\gamma,\beta) = \frac{1}{2}\log\Big(1 + \frac{\gamma\bar{\beta}P}{\gamma\beta P + \bar{\gamma}P + N_2}\Big) \tag{63a}$$

$$R_1(\gamma,\beta) = \frac{1}{2}\log\Big(1 + \frac{\gamma\beta P}{N_1}\Big) \tag{63b}$$

and, as in (3), $\bar{\beta}$ denotes $1-\beta$ and $\bar{\gamma}$ denotes $1-\gamma$.
Since every pair (R_{c}, R_1) in the capacity region $\mathcal{C}_{\text{G-BC}}$ is equal to $\big(R_{\mathrm{c}}(\gamma,\beta), R_1(\gamma,\beta)\big)$ for some choice of γ and β, this proposition will indeed establish sufficiency.

The proof of the proposition is a bit technical, so we begin with a plausibility argument before proceeding with a rigorous proof. The plausibility argument has a number of shortcomings. The first has to do with the random-coding argument and the code-averaged distortion. The problem arises because we are dealing here with two separate distortion constraints, which must be satisfied simultaneously. When there is but a single distortion constraint, the random coding argument guarantees that if the code-averaged distortion meets the constraint then there must exist a (deterministic) code that also meets the constraint. But when there are two distortions to deal with, the fact that each of the code-averaged distortion constraints is satisfied does not imply that there exists a single code that simultaneously meets the two constraints. To deal with this issue, the rigorous proof—rather than studying the code-averaged distortions—studies the probabilities that a randomly chosen codebook meets the constraints. These probabilities are shown to tend to one, which implies that the probability that both constraints are satisfied also tends to one, and the random coding argument can be invoked.

Another shortcoming is due to the unboundedness of the MSE distortion, which implies that the effect on the distortion of representation failures, even if rare, need not be negligible. To address this issue, the rigorous proof considers ensembles of codes whose codewords are not Gaussian but drawn uniformly on the n-dimensional sphere. This must also be done for the Gelfand-Pinsker/Costa codebook, where it results in an unwieldy distribution for the power in the sum of the state (source) and the codeword. Since this sum is later treated as noise, and since it is not Gaussian, this requires an analysis of the probability of error in non-Gaussian noise. To control the power in this sum, our Gelfand-Pinsker/Costa encoder does not search for the nearest codeword but for the one whose inner product with the state sequence is closest to our target value; see (86).

Finally, our plausibility argument tacitly assumes that the decoded codewords can be stripped-off perfectly; it does not therefore account for the effect of decoding errors on the distortions. The rigorous proof accounts for such errors by introducing $\mathbf{\Delta}_1$ in (105).

A. Sufficiency: A Plausibility Argument

Fix some $0 \leq \gamma, \beta \leq 1$. Given the source sequence $\mathbf{S}$ and the message pair (W_{c}, W_1), the encoder sends the n-tuple $\mathbf{X}(\mathbf{S}, W_{\mathrm{c}}, W_1)$ that is given by

$$\mathbf{X}(\mathbf{S}, W_{\mathrm{c}}, W_1) = \mathbf{X}_{\mathrm{a}}(\mathbf{S}) + \mathbf{X}_{\text{Gau}}(W_{\mathrm{c}}) + \mathbf{X}_{\text{D-P}}(W_1; \mathbf{S}), \tag{64}$$

where $\mathbf{X}_{\mathrm{a}}(\mathbf{S})$ is a power-$\bar{\gamma}P$ scaled version of $\mathbf{S}$,

$$\mathbf{X}_{\mathrm{a}}(\mathbf{S}) = \sqrt{\frac{\bar{\gamma}P}{\sigma^2}}\mathbf{S}, \tag{65}$$

and where the remaining power, namely γP, is used by the remaining terms on the RHS of (64): $\gamma\bar{\beta}P$ by $\mathbf{X}_{\text{Gau}}(W_{\mathrm{c}})$ and $\gamma\beta P$ by $\mathbf{X}_{\text{D-P}}(W_1; \mathbf{S})$. (The terms on the RHS of (64) are orthogonal, so their powers add.) The term $\mathbf{X}_{\text{Gau}}(W_{\mathrm{c}})$ is the codeword indexed by W_{c} in a power-$\gamma\bar{\beta}P$ Gaussian codebook. The term $\mathbf{X}_{\text{D-P}}(W_1; \mathbf{S})$ is the sequence transmitted in order to convey Message W_1 in Costa's scheme of power $\gamma\beta P$ for writing on dirty paper [7] when the "dirt sequence" is $\mathbf{X}_{\mathrm{a}}(\mathbf{S})$ (which is known noncausally to the transmitter) and the noise is the noise corrupting the strong receiver, i.e., $\mathbf{Z}_1$.

The decoders operate as follows. The weak receiver, Receiver 2, uses nearest-neighbor decoding to decode W_{c} treating $\mathbf{X}_{\mathrm{a}}(\mathbf{S}) + \mathbf{X}_{\text{D-P}}(W_1; \mathbf{S})$ as noise that is added on top of the noise $\mathbf{Z}_2$ corrupting its terminal. The total effective noise is thus of power $\bar{\gamma}P + \gamma\beta P + N_2$. And since the desired signal $\mathbf{X}_{\text{Gau}}(W_{\mathrm{c}})$ is of power $\gamma\bar{\beta}P$, it can decode W_{c} whenever R_{c} is smaller than the RHS of (63a) [14]. This condition guarantees that also the strong receiver can decode W_{c}. The weak receiver then subtracts $\mathbf{X}_{\text{Gau}}(W_{\mathrm{c}})$ from its received sequence, and thus obtains

$$\tilde{\mathbf{Y}}_2 \triangleq \mathbf{X}_{\mathrm{a}}(\mathbf{S}) + \mathbf{X}_{\text{D-P}}(W_1; \mathbf{S}) + \mathbf{Z}_2$$

and forms its linear minimum MSE estimate $\hat{\mathbf{S}}_2$ of the source sequence $\mathbf{S}$ based on $\tilde{\mathbf{Y}}_2$

$$\hat{\mathbf{S}}_2 = \frac{\sqrt{\bar{\gamma}P\sigma^2}}{\gamma\beta P + \bar{\gamma}P + N_2}\tilde{\mathbf{Y}}_2 \tag{66}$$

with corresponding distortion

$$D_2(\gamma,\beta) = \sigma^2\frac{\gamma\beta P + N_2}{\gamma\beta P + \bar{\gamma}P + N_2} \tag{67}$$

$$= \sigma^2 D_{2,\min}\big(R_{\mathrm{c}}(\gamma,\beta), R_1(\gamma,\beta)\big), \tag{68}$$

where the second equality follows from a straightforward calculation, which is carried out in Appendix B.

After decoding W_{c}, also the strong receiver subtracts $\mathbf{X}_{\mathrm{Gau}}(W_{\mathrm{c}})$ from its received sequence to form

$$\tilde{\mathbf{Y}}_1 \triangleq \mathbf{X}_{\mathrm{a}}(\mathbf{S}) + \mathbf{X}_{\text{D-P}}(W_1; \mathbf{S}) + \mathbf{Z}_1. \tag{69}$$

It then decodes W_1 using Costa's decoder. Since the dirty-paper coding renders the "dirt" harmless, it can recover W_1 whenever R_1 is lower than the RHS of (63b). It finally forms its estimate of the source as

$$\hat{\mathbf{S}}_1 = \frac{\sqrt{\bar{\gamma} P \sigma^2}}{\gamma\beta P + \bar{\gamma} P + N_1} \tilde{\mathbf{Y}}_1 \tag{70}$$

and the corresponding achievable distortion is

$$D_1(\gamma, \beta) = \sigma^2 \frac{\gamma\beta P + N_1}{\gamma\beta P + \bar{\gamma} P + N_1} \tag{71}$$

$$= \sigma^2 D_{1,\min}\big(R_{\mathrm{c}}(\gamma, \beta), R_1(\gamma, \beta)\big), \tag{72}$$

where the distortion can be read off from the parametric equations in [20, Th. 2, Eq. (5)] or those following [3, Eq. (173)] by replacing γ there with 1; and the second equality follows from a straightforward calculation, which is carried out in Appendix C.

As noted in [20, Sec. II.C, Footnote 2], the decoded Costa's codeword $\mathbf{U}(W_1; \mathbf{S})$ and $\mathbf{S}$ are conditionally independent given $\tilde{\mathbf{Y}}_1$, hence the state estimation error cannot be further reduced using $\mathbf{U}(W_1; \mathbf{S})$ when given $\tilde{\mathbf{Y}}_1$.

B. Sufficiency: A proof

In order to prove Proposition 1, we shall need the following lemma on linear estimation of vectors.

Lemma 1: Let $\mathbf{x}, \mathbf{c} \in \mathbb{R}^n$ and $\mu, \eta \in \mathbb{R}$ be deterministic, and let $\mathbf{Z}$ be a centered random n-vector whose components are of variance σ^2 and uncorrelated. Let

$$\mathbf{Y} = \mathbf{x} + \mathbf{c} + \mathbf{Z}. \tag{73}$$

Then

$$\mathsf{E}\Big[\|\eta\mathbf{Y} - \mu\mathbf{x}\|^2\Big] = (\eta - \mu)^2 \|\mathbf{x}\|^2 + 2\eta(\eta - \mu)\langle \mathbf{c}, \mathbf{x}\rangle + \eta^2 \|\mathbf{c}\|^2 + n\eta^2\sigma^2, \tag{74}$$

where $\langle \mathbf{u}, \mathbf{v}\rangle$ denotes the Euclidean inner product $\sum u_i\, v_i$.

Proof: This calculation can be carried out, for example, by decomposing $\mathbf{c}$ into two parts: one that is co-linear with $\mathbf{x}$ and one that is orthogonal to $\mathbf{x}$. The details are omitted. ∎

We are now ready to prove Proposition 1.

Proof of Proposition 1: Inspired by [5, Remark III.5] and [6], our scheme will not describe the IID Gaussian n-tuple $\mathbf{S}$ directly. Instead, we will describe its scaled version

$$\mathbf{S}' = \sqrt{n\sigma^2} \frac{\mathbf{S}}{\|\mathbf{S}\|}, \tag{75}$$

which is uniformly distributed over the n-sphere:

$$\|\mathbf{S}'\| = \sqrt{n\sigma^2}. \tag{76}$$

Asymptotically, as n tends to infinity, the average MSE incurred when estimating $\mathbf{S}$ using some estimator $\hat{\mathbf{S}}'$ for $\mathbf{S}'$ is no worse than when that estimator is used to estimate $\mathbf{S}'$, because, by the Norm Inequality, for every estimator $\hat{\mathbf{S}}'$ of $\mathbf{S}'$,

$$\mathsf{E}\Big[\big\|\hat{\mathbf{S}}' - \mathbf{S}\big\|^2\Big]^{1/2} = \mathsf{E}\Big[\big\|(\hat{\mathbf{S}}' - \mathbf{S}') + (\mathbf{S}' - \mathbf{S})\big\|^2\Big]^{1/2} \le \mathsf{E}\Big[\big\|\hat{\mathbf{S}}' - \mathbf{S}'\big\|^2\Big]^{1/2} + \mathsf{E}\Big[\big\|\mathbf{S}' - \mathbf{S}\big\|^2\Big]^{1/2}$$

and[3]

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{E}\Big[\|\mathbf{S}' - \mathbf{S}\|^2\Big] = 0. \tag{77}$$

The transmitted signal in our scheme has the form

$$\mathbf{x}(\mathbf{s}', w_{\mathrm{c}}, w_1) = \mathbf{x}_{\mathrm{a}}(\mathbf{s}') + \mathbf{x}_{\text{D-P}}(w_1; \mathbf{s}' | \mathcal{C}_{\text{D-P}}^{(n)}) + \mathbf{x}_{\mathrm{Gau}}(w_{\mathrm{c}} | \mathcal{C}_{\mathrm{Gau}}^{(n)}), \tag{78}$$

where $\mathbf{x}_{\mathrm{a}}(\mathbf{s}')$ is a scaled-to-power-$\bar{\gamma}P$ version of $\mathbf{s}'$

$$\mathbf{x}_{\mathrm{a}}(\mathbf{s}') = \sqrt{\frac{\bar{\gamma} P}{\sigma^2}}\, \mathbf{s}' \tag{79a}$$

$$\|\mathbf{x}_{\mathrm{a}}(\mathbf{s}')\|^2 = n\bar{\gamma}P; \tag{79b}$$

$\mathbf{x}_{\text{D-P}}(w_1; \mathbf{s}' | \mathcal{C}_{\text{D-P}}^{(n)})$ is a variant of Dirty-Paper coding [7] using the codebook $\mathcal{C}_{\text{D-P}}^{(n)}$ when the message is w_1 and the interference is $\mathbf{x}_{\mathrm{a}}(\mathbf{s}')$; and $\mathbf{x}_{\mathrm{Gau}}(w_{\mathrm{c}} | \mathcal{C}_{\mathrm{Gau}}^{(n)})$ is a variant of the encoding of w_{c} using the codebook $\mathcal{C}_{\mathrm{Gau}}^{(n)}$ for the Gaussian channel. Often we shall make the codebooks implicit and write $\mathbf{x}_{\text{D-P}}(w_1; \mathbf{s}')$ and $\mathbf{x}_{\mathrm{Gau}}(w_{\mathrm{c}})$.

We next describe the codebooks in greater detail, starting with $\mathcal{C}_{\text{D-P}}^{(n)}$. It is constructed from an ensemble of codes as in [5], but with a slightly different encoding rule: The ensemble is constructed starting with the positive parameters

$$\tilde{N},\ \tilde{P},\ Q,\ \tilde{R}, \text{ and } \tilde{R}', \tag{80}$$

where

$$Q = \frac{1}{n} \|\mathbf{X}_{\mathrm{a}}(\mathbf{S}')\|^2 = \bar{\gamma} P \tag{81}$$

and

$$\tilde{N} = N_1. \tag{82}$$

Associated with $\tilde{P}$ and $\tilde{N}$ is

$$\tilde{\alpha} \triangleq \frac{\tilde{P}}{\tilde{P} + \tilde{N}}. \tag{83}$$

The codebooks in the ensemble consist of $2^{n\tilde{R}}$ bins, each containing $2^{n\tilde{R}'}$ codewords. The k-th codeword in the m-th bin is denoted $\mathbf{V}_{m,k}$. The $2^{n(\tilde{R}+\tilde{R}')}$ codewords are drawn

[3] To justify (77) use the co-linearity of $\mathbf{S}'$ with $\mathbf{S}$ to conclude that $\mathsf{E}\Big[\|\mathbf{S}' - \mathbf{S}\|^2\Big] = \mathsf{E}\big[(\|\mathbf{S}'\| - \|\mathbf{S}\|)^2\big]$; recall (75) and $\mathsf{E}\Big[\|\mathbf{S}\|^2\Big] = n\sigma^2$; and express $\mathsf{E}[\|\mathbf{S}\|]$ as $\sigma\sqrt{2}\,\Gamma\big((n+1)/2\big)/\Gamma(n/2)$ [15, Eq. (19.42)] (where $\Gamma(\cdot)$ denotes the Gamma function). One can then conclude the proof of (77) by noting [12, 8.328] [19, Eq. (2.36)] that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \sqrt{2}\,\Gamma\big((n+1)/2\big)/\Gamma(n/2) = 1.$$

In fact, using the log-convexity of $\Gamma(\cdot)$ one can show that for $0 < s < 1$,

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s}.$$

(This ratio is often called *Gautchi's Ratio*.)

independently and uniformly over the centered n-sphere of radius $\sqrt{n(\tilde{P}+\tilde{\alpha}^2 Q)}$, so

$$\|\mathbf{V}_{m,k}\|^2 = n(\tilde{P}+\tilde{\alpha}^2 Q). \tag{84}$$

To specify how we pick $\mathcal{C}_{\text{D-P}}^{(n)}$ from this ensemble, we next consider a specific encoder and a specific genie-aided decoder. To describe the encoder, let us define the angle $\sphericalangle(\mathbf{w},\mathbf{v})$ between two nonzero vectors as the angle between 0 and π such that

$$\cos\sphericalangle(\mathbf{w},\mathbf{v}) = \frac{\langle \mathbf{w},\mathbf{v}\rangle}{\|\mathbf{w}\|\,\|\mathbf{v}\|}. \tag{85}$$

To send the message M after observing $\mathbf{S}'$, the encoder searches Bin M for the codeword whose angle with $\mathbf{S}'$ (and hence also with $\mathbf{X}_{\text{a}}(\mathbf{S}')$) is of cosine that is closest to $(\tilde{\alpha}^2 Q/(\tilde{P}+\tilde{\alpha}^2 Q))^{1/2}$. Denoting this codeword $\mathbf{V}_{M,K^\star}$,

$$K^\star = \arg\min_k \left| \left\langle \frac{\mathbf{S}'}{\|\mathbf{S}'\|}, \frac{\mathbf{V}_{M,k}}{\|\mathbf{V}_{M,k}\|} \right\rangle - \sqrt{\frac{\tilde{\alpha}^2 Q}{\tilde{P}+\tilde{\alpha}^2 Q}} \right|. \tag{86}$$

It then sets

$$\mathbf{X}_{\text{D-P}}(M;\mathbf{S}') = \mathbf{V}_{M,K^\star} - \tilde{\alpha}\,\mathbf{X}_{\text{a}}(\mathbf{S}'), \tag{87a}$$

provided that this does not result in the power in $\mathbf{X}_{\text{a}}(\mathbf{S}') + \mathbf{X}_{\text{D-P}}(M;\mathbf{S}')$ being too large, i.e., provided that

$$\frac{1}{n}\left\|\mathbf{V}_{M,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')\right\|^2 \le \bar{\gamma}P + \gamma\beta P, \tag{87b}$$

and otherwise it sets

$$\mathbf{X}_{\text{D-P}}(M;\mathbf{S}') = \mathbf{0}. \tag{87c}$$

The genie-aided decoder bases its guess on the vector

$$\mathbf{X}_{\text{a}}(\mathbf{S}') + \mathbf{X}_{\text{D-P}}(M;\mathbf{S}') + \mathbf{Z}_1 \tag{88}$$

and searches the codewords $\{\mathbf{V}_{m,k}\}$ for the codeword of largest inner product with it; its guess is the bin containing this codeword.

Assume now that $\tilde{R}'$ is sufficiently large so that

$$1 - 2^{-2\tilde{R}'} > \frac{\tilde{\alpha}^2 Q}{\tilde{P}+\tilde{\alpha}^2 Q}. \tag{89}$$

It then follows using standard results on the area of spherical caps [5, Appendix B] that the normalized inner product between the selected codeword and $\mathbf{S}'$ converges in probability:

$$\operatorname*{p-lim}_{n\to\infty} \left\langle \frac{\mathbf{S}'}{\|\mathbf{S}'\|}, \frac{\mathbf{V}_{M,K^\star}}{\|\mathbf{V}_{M,K^\star}\|} \right\rangle = \sqrt{\frac{\tilde{\alpha}^2 Q}{\tilde{P}+\tilde{\alpha}^2 Q}}. \tag{90}$$

This in combination with (79b), (81), and (84) establishes the asymptotic orthogonality

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\left\langle \mathbf{V}_{M,K^\star} - \tilde{\alpha}\,\mathbf{X}_{\text{a}}(\mathbf{S}'), \mathbf{X}_{\text{a}}(\mathbf{S}') \right\rangle = 0. \tag{91}$$

Moreover, (90) in combination with (79b), (81), and (84) implies that

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n}\left\|\mathbf{V}_{M,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')\right\|^2 = \bar{\gamma}P + \tilde{P}. \tag{92}$$

Consequently, if

$$\tilde{P} < \gamma\beta P, \tag{93}$$

then the probability that (87b) is violated tends to zero:

$$\lim_{n\to\infty} \Pr\left[\frac{1}{n}\left\|\mathbf{V}_{M,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')\right\|^2 > \bar{\gamma}P + \gamma\beta P\right] = 0. \tag{94}$$

As in [5], if

$$\tilde{R} + \tilde{R}' < \frac{1}{2}\log\left(1 + \frac{\tilde{P}}{\tilde{N}} + \frac{Q\tilde{P}}{\tilde{N}(\tilde{P}+\tilde{N})}\right), \tag{95}$$

then the probability of a decoding error tends to zero as $n\to\infty$. That is, if $\mathcal{E}_{\text{D-P}}^{\text{dec}}$ denotes the event corresponding to a decoding error, then

$$\lim_{n\to\infty} \Pr\left(\mathcal{E}_{\text{D-P}}^{\text{dec}}\right) = 0. \tag{96}$$

It follows from (90), (91), (94), and (96) that there exists a sequence $\delta_n \downarrow 0$ such that the probability of a decoding error or

$$\left| \left\langle \frac{\mathbf{S}'}{\|\mathbf{S}'\|}, \frac{\mathbf{V}_{M,K^\star}}{\|\mathbf{V}_{M,K^\star}\|} \right\rangle - \sqrt{\frac{\tilde{\alpha}^2 Q}{\tilde{P}+\tilde{\alpha}^2 Q}} \right| > \delta_n \tag{97a}$$

or

$$\left| \frac{1}{n}\left\langle \mathbf{V}_{M,K^\star} - \tilde{\alpha}\,\mathbf{X}_{\text{a}}(\mathbf{S}'), \mathbf{X}_{\text{a}}(\mathbf{S}') \right\rangle \right| > \delta_n \tag{97b}$$

or

$$\frac{1}{n}\left\|\mathbf{V}_{M,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')\right\|^2 > \bar{\gamma}P + \gamma\beta P \tag{97c}$$

tends to zero as $n\to\infty$. Fix such a sequence $\{\delta_n\}$, and let $\mathcal{E}_{\text{D-P}}^{\text{enc}}$ denote the event that at least one of the inequalities in (97) is satisfied.

By the random-coding argument, there exists a sequence of codes $\{\mathcal{C}_{\text{D-P}}^{(n)}\}$ that, for the above $\{\delta_n\}$, satisfies

$$\lim_{n\to\infty} \Pr[\mathcal{E}_{\text{D-P}}^{\text{enc}} | \mathcal{C}_{\text{D-P}}^{(n)}] = 0 \tag{98}$$

and

$$\lim_{n\to\infty} \Pr[\mathcal{E}_{\text{D-P}}^{\text{dec}} | \mathcal{C}_{\text{D-P}}^{(n)}] = 0. \tag{99}$$

If a genie provides the decoder with the vector (88), then—by letting $\tilde{P}\uparrow\gamma\beta P$ and by considering the limit under which $\tilde{R}'$ is decreased until (89) holds with equality—we can achieve (using the genie-aided decoder) rates that approach $R_1(\gamma,\beta)$ of (63b) [5].

Having completed the description of the construction of the sequence $\{\mathcal{C}_{\text{D-P}}^{(n)}\}$, we next turn to the Gaussian codes $\{\mathcal{C}_{\text{Gau}}^{(n)}\}$. It follows from (98) (*c.f.* (92)) that, using the codes $\{\mathcal{C}_{\text{D-P}}^{(n)}\}$, the power in $\mathbf{V}_{W_1,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')$ converges in probability to $\bar{\gamma}P + \tilde{P}$:

$$\lim_{n\to\infty} \Pr\left[\left|\frac{1}{n}\left\|\mathbf{V}_{W_1,K^\star} + (1-\tilde{\alpha})\,\mathbf{X}_{\text{a}}(\mathbf{S}')\right\|^2 - (\bar{\gamma}P + \tilde{P})\right| > \delta \,\middle|\, \mathcal{C}_{\text{D-P}}^{(n)}\right] = 0, \quad \forall \delta > 0. \tag{100}$$

Consequently, with these codes, the power in

$$\mathbf{X}_{\text{a}}(\mathbf{S}') + \mathbf{X}_{\text{D-P}}(W_1;\mathbf{S}'|\mathcal{C}_{\text{D-P}}^{(n)}) + \mathbf{Z}_2 \tag{101}$$

—which for the purpose of guessing W_{c} by the weak receiver we think of as noise—converges in probability to $\bar{\gamma}P+\tilde{P}+N_2$. We now think of (101) as power-$(\bar{\gamma}P+\tilde{P}+N_2)$ additive noise and construct the codebooks $\{\mathcal{C}_{\mathrm{Gau}}^{(n)}\}$ to combat it as in [14]: We consider the performance of the nearest-neighbor decoder on the ensemble of codes whose $2^{nR_{\mathrm{c}}}$ codewords are drawn independently and uniformly over the n-sphere

$$\|\mathbf{X}_{\mathrm{Gau}}(w_{\mathrm{c}})\|^2 = n\gamma\bar{\beta}P. \tag{102}$$

The (ensemble-averaged) probability of decoding error (when the additive noise is as in (101)) then tends to zero, provided that R_{c} is smaller than $R_{\mathrm{c}}(\gamma,\beta)$ of (63a). Subject to this condition, we can thus choose a sequence of deterministic codes $\{\mathcal{C}_{\mathrm{Gau}}^{(n)}\}$ for which the probability of error with nearest-neighbor decoding tends to zero as $n\to\infty$. Since the probability of error tends to zero at the poor receiver, it can also be made to tend to zero at the better receiver (e.g., by injecting Gaussian noise of variance N_2-N_1). And using a genie argument [22, p. 419], [18], it follows from (99) that W_1 can be recovered at the better receiver too.

It remains to propose estimation rules and to study their performance. Define (*c.f.* (66))

$$\eta_2 = \frac{\sqrt{\bar{\gamma}P\sigma^2}}{\gamma\beta P+\bar{\gamma}P+N_2} \tag{103}$$

and (*c.f.* (70))

$$\eta_1 = \frac{\sqrt{\bar{\gamma}P\sigma^2}}{\gamma\beta P+\bar{\gamma}P+N_1}. \tag{104}$$

Receiver 1 estimates $\mathbf{S}'$ (and hence also $\mathbf{S}$) as follows: First it decodes the common message by forming $\hat{W}_{\mathrm{c}}^{(1)}$, and it then subtracts the corresponding codeword $\mathbf{X}_{\mathrm{Gau}}(\hat{W}_{\mathrm{c}}^{(1)}|\mathcal{C}_{\mathrm{Gau}}^{(n)})$ from its received sequence $\mathbf{Y}_1$ to obtain $\tilde{\mathbf{Y}}_1+\mathbf{\Delta}_1$, where $\tilde{\mathbf{Y}}_1$ is defined in (69), and $\mathbf{\Delta}_1$ is defined as

$$\mathbf{\Delta}_1 = \mathbf{X}_{\mathrm{Gau}}(W_{\mathrm{c}}|\mathcal{C}_{\mathrm{Gau}}^{(n)}) - \mathbf{X}_{\mathrm{Gau}}(\hat{W}_{\mathrm{c}}^{(1)}|\mathcal{C}_{\mathrm{Gau}}^{(n)}). \tag{105}$$

It then forms its estimate

$$\hat{\mathbf{S}}_1 = \eta_1\Big(\tilde{\mathbf{Y}}_1+\mathbf{\Delta}_1\Big), \tag{106}$$

where η_1 is given in (104); *c.f.* (70). Expressing the estimation error as

$$\big(\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big)+\eta_1\mathbf{\Delta}_1, \tag{107}$$

we obtain from the Norm Inequality

$$\frac{1}{\sqrt{n}}\mathsf{E}\Big[\big\|\hat{\mathbf{S}}_1-\mathbf{S}'\big\|^2\Big]^{1/2} \le \frac{1}{\sqrt{n}}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\Big]^{1/2} + \frac{1}{\sqrt{n}}\mathsf{E}\Big[\big\|\eta_1\mathbf{\Delta}_1\big\|^2\Big]^{1/2}. \tag{108}$$

Since we have chosen our codewords on the sphere, $\|\mathbf{X}_{\mathrm{Gau}}(W_{\mathrm{c}}|\mathcal{C}_{\mathrm{Gau}}^{(n)})\|^2$ equals $n\gamma\bar{\beta}P$ deterministically, and consequently

$$\mathsf{E}\big[\|\mathbf{\Delta}_1\|^2\big] \le n\,4\gamma\bar{\beta}P\,P_{\mathrm{e}}^{(1)}, \tag{109}$$

so the second term on the RHS of (108) tends to zero in the limit as n tends to infinity, because in this limit $P_{\mathrm{e}}^{(1)}$ tends to zero. Thus

$$\varlimsup_{n\to\infty}\frac{1}{n}\mathsf{E}\Big[\big\|\hat{\mathbf{S}}_1-\mathbf{S}'\big\|^2\Big] = \varlimsup_{n\to\infty}\frac{1}{n}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\Big]. \tag{110}$$

To study the RHS of (110), we condition on $\mathbf{S}'=\mathbf{s}'$ and $W_1=w_1$. Under this conditioning, $\mathbf{X}_{\mathrm{a}}(\mathbf{S}')$ and $\mathbf{X}_{\text{D-P}}(W_1;\mathbf{S}'|\mathcal{C}_{\text{D-P}}^{(n)})$ are deterministic. We first argue that, irrespective of $\mathbf{s}'$ (on the sphere) and of w_1,

$$\frac{1}{n}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\,\Big|\,W_1=w_1,\mathbf{S}'=\mathbf{s}'\Big] \le K_1(\eta_1,\gamma,\beta,P,N_1), \tag{111}$$

where the constant $K_1(\cdot)$ depends on its arguments but not on n. Indeed, by (87),

$$\frac{1}{n}\big\|\mathbf{x}_{\mathrm{a}}(\mathbf{s}')+\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)})\big\|^2 \le \bar{\gamma}P+\gamma\beta P, \tag{112}$$

so, by the Norm Inequality,

$$\frac{1}{n}\Big\|\eta_1\big(\mathbf{x}_{\mathrm{a}}(\mathbf{s}')+\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)})\big)-\mathbf{s}'\Big\|^2 \le \Big(|\eta_1|\sqrt{\bar{\gamma}P+\gamma\beta P}+\sigma\Big)^2 \tag{113}$$

and thus

$$\frac{1}{n}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\,\Big|\,W_1=w_1,\mathbf{S}'=\mathbf{s}'\Big] \le \Big(|\eta_1|\sqrt{\bar{\gamma}P+\gamma\beta P}+\sigma\Big)^2+\eta_1^2N_1. \tag{114}$$

We can therefore choose $K_1(\cdot)$ as the RHS of the above.

Inequality (111) holds for all pairs $(w_1,\mathbf{s}')$. For pairs that do not result in the event $\mathcal{E}_{\text{D-P}}^{\mathrm{enc}}$, we can do better. For such pairs we obtain from Lemma 1 upon substituting $\mathbf{x}_{\mathrm{a}}(\mathbf{s}')$ for $\mathbf{x}$; $\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)})$ for $\mathbf{c}$; $\mathbf{Z}_1$ for $\mathbf{Z}$; and (*c.f.* (79a))

$$\mu = \sqrt{\frac{\sigma^2}{\bar{\gamma}P}} \tag{115}$$

$$\begin{aligned}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\,\Big|\,W_1=w_1,\mathbf{S}'=\mathbf{s}'\Big] &= (\eta_1-\mu)^2\,\|\mathbf{x}_{\mathrm{a}}(\mathbf{s}')\|^2\\ &\quad+2\eta_1(\eta_1-\mu)\Big\langle\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)}),\mathbf{x}_{\mathrm{a}}(\mathbf{s}')\Big\rangle\\ &\quad+\eta_1^2\Big\|\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)})\Big\|^2+n\eta_1^2N_1.\end{aligned} \tag{116}$$

Using (79b), the inequality

$$\Big|\Big\langle\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)}),\mathbf{x}_{\mathrm{a}}(\mathbf{s}')\Big\rangle\Big| \le n\delta_n \tag{117}$$

(which holds by the negation of (97b)), and the inequality

$$\Big\|\mathbf{x}_{\text{D-P}}(w_1;\mathbf{s}'|\mathcal{C}_{\text{D-P}}^{(n)})\Big\|^2 \le n\gamma\beta P+2n\delta_n \tag{118}$$

(which holds by the negation of (97b) and the negation of (97c)) we obtain that for pairs $(w_1,\mathbf{s}')$ that do not result in the event $\mathcal{E}_{\text{D-P}}^{\mathrm{enc}}$

$$\frac{1}{n}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\,\Big|\,W_1=w_1,\mathbf{S}'=\mathbf{s}'\Big] = (\eta_1-\mu)^2\bar{\gamma}P+\eta_1^2\gamma\beta P+\eta_1^2N_1+o(1), \tag{119}$$

where the $o(1)$ term depends on η_1, μ, and δ_n and tends to zero as n tends to infinity.

In the average of

$$\frac{1}{n}\mathsf{E}\Big[\big\|\eta_1\tilde{\mathbf{Y}}_1-\mathbf{S}'\big\|^2\,\Big|\,W_1=w_1,\mathbf{S}'=\mathbf{s}'\Big]$$

over the pairs $(w_1, \mathbf{s}')$, the contribution of the pairs for which the event $\mathcal{E}_{\text{D-P}}^{\text{enc}}$ occurs is at most

$$\Pr\left(\mathcal{E}_{\text{D-P}}^{\text{enc}}\right) K_1(\eta_1, \gamma, \beta, P, N_1),$$

and the contribution of the pairs for which this event does not occur is at most

$$(\eta_1 - \mu)^2 \bar{\gamma} P + \eta_1^2 \gamma \beta P + \eta_1^2 N_1 + o(1),$$

(where we have upper bounded $1 - \Pr\left(\mathcal{E}_{\text{D-P}}^{\text{enc}}\right)$ by 1). Recalling that $\Pr\left(\mathcal{E}_{\text{D-P}}^{\text{enc}}\right)$ tends to zero, we thus obtain

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \mathsf{E}\left[\left\|\eta_1 \tilde{\mathbf{Y}}_1 - \mathbf{S}'\right\|^2\right] \le (\eta_1 - \mu)^2 \bar{\gamma} P + \eta_1^2 \gamma \beta P + \eta_1^2 N_1, \tag{120}$$

which evaluates (using (115), and (104)) to the RHS of (71).

The analysis of the estimation error at the weak terminal is nearly identical and is omitted. ■

APPENDIX

A. Common Randomness

We next prove Remark 4 by showing that allowing the transmitter and the receivers access to a common source of randomness (that is independent of the messages and of the source) does not enlarge the set of achievable quadruples. *A fortiori*, nor does allowing for stochastic encoders. We thus consider the case where the transmitted sequence, in addition to depending on the source and messages, also depends on the realization θ_n of a random variable Θ_n that is drawn according to $P_\Theta^{(n)}$ from $\mathcal{O}_n$ and which is revealed to both receivers. As in (30), we define

$$\varepsilon_n(\theta_n) = \max\left\{P_{\text{e}}^{(1)}(\theta_n), P_{\text{e}}^{(2)}(\theta_n)\right\},$$

where the argument θ_n indicates the dependence on the realization of the common randomness. Thus, $P_{\text{e}}^{(1)}(\theta_n)$ can be viewed as the probability of error at Terminal 1 conditional on $\Theta_n = \theta_n$, and we use similar notation for the other quantities that depend on θ_n. Averaging over Θ_n,

$$\varepsilon_n(P_\Theta^{(n)}) = \int_{\mathcal{O}_n} \varepsilon_n(\theta_n)\, \mathrm{d}P_\Theta^{(n)}, \tag{121}$$

where the argument $P_\Theta^{(n)}$ on the LHS indicates that the quantity is being averaged over $P_\Theta^{(n)}$, with similar notation for other such averages. Thus, for example, for $\nu \in \{1, 2\}$ we define $\delta_n^{(\nu)}(\theta_n)$ analogously to (29), and we define

$$\delta_n^{(\nu)}(P_\Theta^{(n)}) = \int_{\mathcal{O}_n} \delta_n^{(\nu)}(\theta_n)\, \mathrm{d}P_\Theta^{(n)}. \tag{122}$$

Consider now a sequence of codes with common randomness that are specified by the measures $\{P_\Theta^{(n)}\}$ and that have vanishing probability of error. Since $\varepsilon_n(P_\Theta^{(n)})$ tends to zero, we can pick a subsequence $\{n_k\}$ of blocklengths for which

$$\varepsilon_{n_k}(P_\Theta^{(n_k)}) < \frac{1}{k}. \tag{123}$$

Define now

$$\mathcal{G}_{n_k} = \left\{\theta_{n_k} \in \mathcal{O}_{n_k} : \varepsilon_{n_k}(\theta_{n_k}) \le \frac{\log k}{k}\right\}. \tag{124}$$

It then follows from (121) and (123) using Markov's inequality that

$$P_\Theta^{(n_k)}\left(\mathcal{G}_{n_k}\right) \ge 1 - \frac{1}{\log k} \tag{125}$$

and, consequently,

$$\lim_{k\to\infty} P_\Theta^{(n_k)}\left(\mathcal{G}_{n_k}\right) = 1. \tag{126}$$

Since the distortion is nonnegative, we can lower-bound the expected distortion in (122) by limiting the integration to $\mathcal{G}_{n_k}$:

$$\delta_{n_k}^{(\nu)}(P_\Theta^{(n_k)}) \ge \int_{\mathcal{G}_{n_k}} \delta_{n_k}^{(\nu)}(\theta_{n_k})\, \mathrm{d}P_\Theta^{(n_k)}. \tag{127}$$

For θ_{n_k} in $\mathcal{G}_{n_k}$ we can lower bound the distortion using (50a) and (62a):

$$\delta_{n_k}^{(\nu)}(\theta_{n_k}) \ge \sigma^2\, D_{\nu,\min}(R_{\text{c}}, R_1) - \Psi^{(\nu)}\Big(\frac{\log k}{k}\Big), \quad \theta_{n_k} \in \mathcal{G}_{n_k}. \tag{128}$$

This and (127) implies

$$\delta_{n_k}^{(\nu)}(P_\Theta^{(n_k)}) \ge P_\Theta^{(n_k)}\left(\mathcal{G}_{n_k}\right)\Big(\sigma^2 D_{\nu,\min}(R_{\text{c}}, R_1) - \Psi^{(\nu)}\Big(\frac{\log k}{k}\Big)\Big). \tag{129}$$

Using (126) we can thus infer that

$$\underline{\lim_{k\to\infty}}\, \delta_{n_k}^{(\nu)}(P_\Theta^{(n_k)}) \ge \sigma^2\, D_{\nu,\min}(R_{\text{c}}, R_1), \quad \nu \in \{1, 2\}, \tag{130}$$

thus establishing that the necessity of (25b) and (25c) also when the transmitter and the receivers share common randomness.

B. Justifying (68)

To justify (68), we begin with the expression for $R_{\text{c}}(\gamma, \beta)$ (63a) and obtain

$$2^{2R_{\text{c}}(\gamma,\beta)} = \frac{P + N_2}{\gamma\beta P + \bar{\gamma} P + N_2}. \tag{131}$$

And starting from the expression for $R_1(\gamma, \beta)$ (63b) and using the definition of R_{d} (28) we obtain

$$R_{\text{d}}\big(R_1(\gamma, \beta)\big) = \frac{1}{2}\log\Big(1 + \frac{\gamma\beta P}{N_2}\Big), \tag{132}$$

so

$$2^{2R_{\text{d}}\left(R_1(\gamma,\beta)\right)} = \frac{N_2 + \gamma\beta P}{N_2}. \tag{133}$$

Consequently,

$$\begin{aligned} &D_{2,\min}\big(R_{\text{c}}(\gamma, \beta), R_1(\gamma, \beta)\big) \\ &\quad = \frac{N_2}{P + N_2} 2^{2(R_{\text{c}}(\gamma,\beta) + R_{\text{d}})} \\ &\quad = \frac{\gamma\beta P + N_2}{\gamma\beta P + \bar{\gamma} P + N_2}, \end{aligned} \tag{134}$$

where the first equality follows from (27), and the second from (131) and (133).

C. Justifying (72)

To justify (72), we begin with the definition of $R_{\mathrm{c}}(\gamma,\beta)$ (63a) (*c.f.* (131))

$$2^{-2R_{\mathrm{c}}(\gamma,\beta)} = \frac{\gamma\beta P + \bar{\gamma}P + N_2}{P + N_2}, \tag{135}$$

which implies that $\gamma\beta P$ can be represented as

$$\gamma\beta P = (P + N_2)2^{-2R_{\mathrm{c}}(\gamma,\beta)} - N_2 - \bar{\gamma}P. \tag{136}$$

Substituting this expression for $\gamma\beta P$ in the expression for $R_1(\gamma,\beta)$ (63b), we obtain

$$R_1(\gamma,\beta) = \frac{1}{2}\log\frac{(P + N_2)2^{-2R_{\mathrm{c}}(\gamma,\beta)} - \bar{\gamma}P - (N_2 - N_1)}{N_1}. \tag{137}$$

Hence,

$$\begin{aligned} &D_{1,\min}\big(R_{\mathrm{c}}(\gamma,\beta), R_1(\gamma,\beta)\big) \\ &\quad= \frac{N_1 2^{2R_1(\gamma,\beta)}}{(P + N_2)2^{-2R_{\mathrm{c}}(\gamma,\beta)} - (N_2 - N_1)} \\ &\quad= \frac{(P + N_2)2^{-2R_{\mathrm{c}}(\gamma,\beta)} - \bar{\gamma}P - N_2 + N_1}{(P + N_2)2^{-2R_{\mathrm{c}}(\gamma,\beta)} - N_2 + N_1} \\ &\quad= \frac{\gamma\beta P + N_1}{\gamma\beta P + \bar{\gamma}P + N_1}, \end{aligned} \tag{138}$$

where the first equality follows from (26), the second from (137), and the third from (136).

ACKNOWLEDGMENT

The authors thank the anonymous referees and the Associate Editor for their helpful comments.

REFERENCES

[1] B. Bandemer and A. El Gamal, "Communication with disturbance constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4488–4502, Aug. 2014.
[2] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.
[3] S. I. Bross and A. Lapidoth, "The rate-and-state capacity with feedback," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1893–1918, Mar. 2018.
[4] S. I. Bross, A. Lapidoth, and S. Tinguely, "Broadcasting correlated Gaussians," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3057–3068, Jul. 2010.
[5] S. I. Bross, A. Lapidoth, and M. Wigger, "Dirty-paper coding for the Gaussian multiaccess channel with conferencing," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5640–5668, Sep. 2012.
[6] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.
[7] M. Costa, "Writing on dirty paper (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[8] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[10] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[11] T. J. Goblick, Jr., "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct. 1965.
[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. San Diego, CA, USA: Academic, 2000.
[13] K. Khezeli and J. Chen, "A source-channel separation theorem with application to the source broadcast problem," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1764–1781, Apr. 2016.
[14] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.
[15] A. Lapidoth, *A Foundation in Digital Communication*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2017.
[16] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2714–2752, Jun. 2010.
[17] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3778–3788, Aug. 2006.
[18] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.
[19] W. Szpankowski, *Average Case Analysis of Algorithms on Sequences*. New York, NY, USA: Wiley, 2001.
[20] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1486–1495, Apr. 2005.
[21] C. Tian, S. Diggavi, and S. Shamai (Shitz), "The achievable distortion region of sending a bivariate Gaussian source on the Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6419–6427, Oct. 2011.
[22] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. Prospect Heights, IL, USA: Waveland Press, 1990.
[23] Y. Zhao and B. Chen, "Capacity theorems for multi-functioning radios," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 2406–2410.

Shraga I. Bross (S'89–M'92–SM'09) received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology, Haifa, in 1978 and 1983, respectively, and the Ph.D. degree from the University of Maryland at College Park, College Park, in 1991, all in electrical engineering. From 1991 to 1992, he was a Post-Doctoral Fellow with the ECE Department, University of Waterloo, Canada. From 1992 to 1998, he was with Orckit Communications Ltd., Tel Aviv, Israel, as a Senior Scientist. From 1998 to 2006, he was a Senior Research Fellow with the EE Department, Technion. Since 2007, he has been with the Faculty of Engineering, Bar-Ilan University, Israel, where he is currently an Associate Professor. His research interests are in digital communications and information theory.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree *(summa cum laude)* in mathematics, the B.Sc. degree *(summa cum laude)* in electrical engineering, and the M.Sc. degree in electrical engineering from the Technion—Israel Institute of Technology in 1986, 1986, and 1990, respectively, and the Ph.D. degree in electrical engineering from Stanford University in 1995.

From 1995 to 1999, he was an Assistant and an Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), and was the KDD Career Development Associate Professor in communications and technology. He is currently a Professor of information theory with ETH Zurich, Zurich, Switzerland.

He has authored the textbook *A Foundation in Digital Communication*, (Cambridge University Press, Second Ed., 2017). His research interests are in digital communications and information theory.