Decoder-Assisted Communications Over Additive Noise Channels

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

***Abstract*— A number of additive noise networks are studied in the presence of a helper that observes the noise and assists the decoder by providing it with a rate-limited description of said noise. It is shown that "flash helping"—where noise descriptions are provided infrequently but with great precision—is often optimal and typically increases capacity by the maximal allowed description rate. It requires no binning. The discrete setting of the modulo-additive noise channel is also discussed.**

***Index Terms*— Broadcast channel, capacity, dirty-paper coding, helper, multi-access channel, state.**

I. Introduction

MOTIVATED by the promise of cooperative communications, we study scenarios where an altruistic helper—wishing to communicate no message of its own—observes the noise disturbing a communication network and wishes to help the receiver(s) to combat it. To this end, the helper can provide the decoder(s) with a rate-limited description of the noise. We quantify the capacity gain afforded by such help for a number of multi-terminal networks.

For the single-user channel, it has been recently shown by Bross and Lapidoth [1] that the capacity gain is equal to the helper's description rate. This result also follows from the work of Cover and Kim [2] and Kim [3] on relay channels, but the coding technique proposed in [1], "flash helping," is very different: it does not require binning. Moreover, as we show here, it generalizes to a number of multi-terminal networks.

The idea behind flash helping is to describe the noise infrequently but with great precision. This technique has recently found applications also in settings where, rather than the decoder, it is the encoder that receives the help [4].

Viewing the noise as "state," our scenario falls under the heading of communication with rate-limited side-information at the decoder. This general problem has been studied extensively, starting with the work of Ahlswede and Han [5] and Heegard and El Gamal [6]. In general, computing the capacity of general state-dependent channels with rate-R_e side-information at the encoder and rate-R_d side-information at the decoder is still open. Only special cases have been solved [7].

The case addressed in [1] corresponds to an additive noise channel with $R_\mathrm{e} = 0$, and the one in [4] to an additive noise channel with $R_\mathrm{d} = 0$. Extending these results to the additive Gaussian noise channel with general R_e and R_d is straightforward.

Since we only consider side-information at the decoder here, we shall adopt the notation of [1] and denote the helper's rate R_h. Except in dealing with modulo-additive noise, we shall focus on additive noise networks under average-power constraints. As noted in [1], flash helping can achieve capacity also under more general cost constraints.

The rest of this paper is organized as follows: In the next section we review the additive noise single-user channel. Section III discusses the additive noise multiple-access channel (MAC) and Section IV the dirty-paper channel [8]. The additive Gaussian-noise broadcast channel is analyzed in Section V. Finally, Section VI shows that a discrete-alphabet analog of flash-helping achieves the capacity of the discrete modulo-additive noise channel, where the noise is described almost-losslessly some of the time and not at all the rest of the time. This highlights the structural similarity of the problems for finite and infinite alphabets.

In all the cases we study, the outer bounds imposed on the capacity region by the Cut-Set Theorem [9, Thm. 15.10.1] are achievable. The capacity region is thus "as large as one could reasonably hope for." It would not even be larger if the helper were cognizant also of the transmitted messages. For more on the Cut-Set Bound for our setting see [1, Fig. 1].

II. The Additive Noise Single-User Channel

A single-user additive noise channel with a helper is depicted in Figure 1. Its time-k output Y_k is

$$Y_k = x_k + Z_k, \tag{1}$$

where $x_k \in \mathbb{R}$ denotes its time-k input, and the noise sequence $\{Z_k\}$ comprises independent identically distributed (IID) random variables of finite second moment N. A rate-R blocklength-n encoder ϕ_{enc} maps the message M, taking values in the set $\mathcal{M} = \{1, \dots, 2^{nR}\}$, to the codeword $\mathbf{x}(m) = (x_1(m), \dots, x_n(m)) \in \mathbb{R}^n$:

$$\phi_{\text{enc}} \colon \mathcal{M} \to \mathcal{X}^n, \; m \mapsto \mathbf{x}(m). \tag{2}$$

An average-power constraint is imposed on the encoder requiring that, for every message $m \in \mathcal{M}$, the corresponding

Manuscript received September 18, 2019; revised February 17, 2020; accepted March 19, 2020. Date of publication March 30, 2020; date of current version July 15, 2020. This article was presented in part at the 2019 IEEE Information Theory Workshop. The associate editor coordinating the review of this article and approving it for publication was R. F. Schaefer. *(Corresponding author: Shraga I. Bross.)*

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

Amos Lapidoth and Gian Marti are with ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch; marti@isi.ee.ethz.ch).

Digital Object Identifier 10.1109/TCOMM.2020.2984215

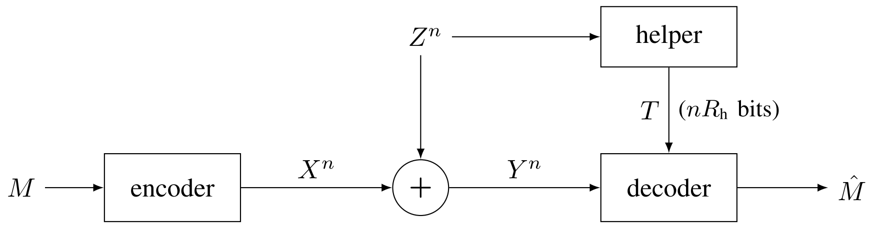


Fig. 1. The additive noise channel with a helper.

codeword $\mathbf{x}(m)$ satisfy $\|\mathbf{x}(m)\|^2 \leq n\mathsf{P}$, where P is the maximal allowed power. When P is zero, only the all-zero codeword can be used, and no data can be transmitted reliably. We shall therefore assume throughout that P is (strictly) positive

$$\mathsf{P} > 0. \tag{3}$$

The helper observes the noise sequence $Z_1, \ldots, Z_n$ and describes it to the decoder using nR_h bits. It does so using the function

$$\phi_\text{help}\colon \mathbb{R}^n \to \{0,1\}^{nR_\text{h}},\ \mathbf{z} \mapsto t \tag{4}$$

where $T = \phi_\text{help}(Z_1, \ldots, Z_n)$ represents the helper's nR_h-bit description of the noise. Based on this description and the channel outputs $Y_1, \ldots, Y_n$, the decoder guesses the message m using the function

$$\psi_\text{dec}\colon \mathbb{R}^n \times \{0,1\}^{nR_\text{h}} \to \mathcal{M},\ (\mathbf{y}, t) \mapsto \hat{m}. \tag{5}$$

A rate R is achievable if for every $\epsilon > 0$ we can find for every sufficiently large blocklength n a triple $(\phi_\text{enc}, \phi_\text{help}, \psi_\text{dec})$ for which the probability of error, uniformly averaged over the message set, is upper-bounded by ϵ. The supremum of the achievable rates is denoted $C(R_\text{h})$.

The following result is due to Cover and Kim [2] and can be obtained from [3], which deals with discrete-memoryless relay channels, using a limiting argument. Nevertheless, we shall present an alternative proof because it introduces a coding technique that we shall later need to treat multi-user settings. Unlike Kim's approach, our scheme requires no binning. As in [3], the converse follows from the Cut-Set Bound, so we provide only a brief sketch.

Theorem 1 ([2], [3]): The capacity of the single-user additive noise channel with decoder assistance is

$$C(R_\text{h}) = C(0) + R_\text{h}. \tag{6}$$

Our proof is based on a lemma on quantization [1, Thm. 3] which we restate here for completeness.

Lemma 1: Let $Z_1, Z_2, \ldots$ be a sequence of random variables whose average second moment converges to σ^2

$$\lim_{\ell\to\infty} \frac{1}{\ell} \sum_{k=1}^{\ell} \mathbb{E}\big[Z_k^2\big] = \sigma^2, \tag{7}$$

and assume that the empirical average of their squares converges in probability to σ^2

$$\frac{1}{\ell} \sum_{k=1}^{\ell} Z_k^2 \xrightarrow{\text{p}} \sigma^2. \tag{8}$$

Then, given any rate R_h and any $\tilde{\epsilon} > 0$, there exists for all sufficiently large blocklengths ℓ a rate-R_h blocklength-ℓ rate-distortion codebook with normalized average MSE distortion

$$\frac{1}{\ell}\, \mathbb{E}\Big[\big\|Z^\ell - \hat{Z}^\ell\big\|^2\Big] \leq \sigma^2\, 2^{-2R_\text{h}} + \tilde{\epsilon}, \tag{9}$$

where $\hat{Z}^\ell$ is the reconstruction ℓ-sequence of $Z^\ell = (Z_1, \ldots, Z_\ell)$ based on its (ℓR_h)-bit description.

Proof: We now prove the achievability part of Theorem 1.

Achievability via Flash Helping: In the following $A^\ell = (A_1, \ldots, A_\ell)$, and $[1:\ell]$ denotes the set $\{1, \ldots, \ell\}$.

Before describing our proposed scheme, which is based on time-sharing and resource-allocation, we begin with a calculation. Suppose that the helper describes the noise sequence Z^ℓ using ℓR_h bits, and the decoder, based on this description, produces the estimate $\hat{Z}^\ell$ of Z^ℓ with corresponding error $\tilde{Z}^\ell$, where

$$\tilde{Z}_k = Z_k - \hat{Z}_k, \qquad k \in [1:\ell]. \tag{10}$$

It then subtracts this estimate from the received sequence Y^ℓ and obtains the sequence $\tilde{Y}^\ell$

$$\tilde{Y}_k = X_k + \tilde{Z}_k, \quad k \in [1:\ell]. \tag{11}$$

We now study $I\big(X^\ell; \tilde{Y}^\ell\big)$ when $X_1, \ldots, X_\ell$ are IID according to some input distribution P_X of second moment smaller than P and having finite differential entropy $h(P_X) > -\infty$. (For example, since P is positive, P_X could correspond to a centered Gaussian of variance $\mathsf{P}/2$.) Since the noise Z^ℓ and its description are independent of X^ℓ,

$$\begin{aligned} I\big(X^\ell; \tilde{Y}^\ell\big) &= h\big(\tilde{Y}^\ell\big) - h\big(\tilde{Z}^\ell\big) && (12)\\ &\geq h\big(X^\ell\big) - h\big(\tilde{Z}^\ell\big) && (13)\\ &= \ell\, h(P_X) - h\big(\tilde{Z}^\ell\big), && (14) \end{aligned}$$

where the inequality follows by conditioning on $\tilde{Z}^\ell$. To further lower-bound $I\big(X^\ell; \tilde{Y}^\ell\big)$, we shall upper-bound $h\big(\tilde{Z}^\ell\big)$ in terms of the estimation error. Of all multivariate distributions of a given second moment matrix, the centered multivariate Gaussian maximizes differential entropy [9, Thm. 8.6.5]. Furthermore, under a constraint on the trace of the second moment matrix, the IID Gaussian distribution maximizes differential entropy. Thus, with $\tilde{\mathbf{Z}}$ denoting $\tilde{Z}^\ell$,

$$\begin{aligned} h(\tilde{\mathbf{Z}}) &\leq \frac{\ell}{2} \log\left(2\pi e \frac{\operatorname{tr}(\mathbb{E}[\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\intercal])}{\ell}\right) && (15)\\ &= \frac{\ell}{2} \log\left(2\pi e \frac{1}{\ell}\, \mathbb{E}\big[\|\tilde{\mathbf{Z}}\|^2\big]\right). && (16) \end{aligned}$$

From (16) and (14) we obtain upon dividing by ℓ,

$$\frac{1}{\ell} I\big(X^\ell; \tilde{Y}^\ell\big) \geq h(P_X) - \frac{1}{2}\log(2\pi e) - \frac{1}{2}\log\left(\frac{1}{\ell}\,\mathbb{E}\big[\|\tilde{\mathbf{Z}}\|^2\big]\right). \tag{17}$$

Lemma 1 guarantees that, given any $\tilde{\delta} > 0$, we can find some sufficiently large ℓ for which there exists a rate-R_h blocklength-ℓ rate-distortion codebook that allows the helper to describe the noise sequence with average distortion that is bounded by

$$\frac{1}{\ell}\,\mathbb{E}\big[\|\tilde{\mathbf{Z}}\|^2\big] \leq \mathsf{N}\,2^{-2R_\mathrm{h}}\,(1+\tilde{\delta}). \tag{18}$$

For such ℓ, it follows from (18) and (17) that

$$\frac{1}{\ell} I\big(X^\ell; \tilde{Y}^\ell\big) \geq h(P_X) - \frac{1}{2}\log\big(2\pi e\,\mathsf{N}(1+\tilde{\delta})\big) + R_\mathrm{h}, \tag{19}$$

which implies by the channel coding theorem [9, Thm. 7.7.1] that, by coding over supersymbols of length ℓ, the right-hand side (RHS) of (19) (or, more precisely, the maximum between it and zero) is achievable.

We now come to "flash helping" and consider time sharing, where for some (small) $0 < \tau < 1$ the channel is used without help in $(1-\tau)$ of the time (i.e., during $(1-\tau)n$ channel uses), and with rate-(R_h/τ) help in τ of the time (i.e., in the remaining τn channel uses).[1] In the absence of help, the rate $C(0)$ is achievable, and in its presence the RHS of (19) is achievable. The overall achievable rate is thus lower-bounded by

$$(1-\tau)C(0) + \tau\left(h(P_X) - \frac{1}{2}\log\big(2\pi e\,\mathsf{N}(1+\tilde{\delta})\big) + \frac{R_\mathrm{h}}{\tau}\right). \tag{20}$$

By studying the limit of (20) as $\tau \downarrow 0$—a limit corresponding to very rare (τ of the time) but high precision (at rate R_h/τ) noise description—we obtain the achievability of $C(0)+R_\mathrm{h}$ and hence conclude the proof. Since the helper is used only τ of the time, and since τ approaches zero, we refer to this form of help as "flash helping."

Sketch of the Converse: By Fano's inequality [9, Thm. 2.10.1], the capacity is bounded from above by the per-symbol mutual information $\frac{1}{n}I(M;\mathbf{Y},T)$ between the message M and the decoder's input $(\mathbf{Y},T)$. We can decompose this as $\frac{1}{n}I(M;\mathbf{Y},T) = \frac{1}{n}I(M;\mathbf{Y}) + \frac{1}{n}I(M;T|\mathbf{Y})$, where the first of the terms is bounded by $C(0)$ and the second by $\frac{1}{n}I(M;T|\mathbf{Y}) \leq \frac{1}{n}H(T) \leq R_\mathrm{h}$, and the result follows. Since we do not assume independence between T and M, the converse holds even if the helper is cognizant of the transmitted message. ∎

Remark 1: Theorem 1 is extended in [1] to noise with memory and to more general peak- and average-power constraints.

Our description in (4) of the helper may seem too expansive, as it allows the helper to observe the noise sequence noncausally and to transmit its description in bursts. However, these issues can be easily overcome, because the description is provided to the decoder and not to the encoder. We can address them by introducing some decoding delay: we consider transmission in blocks, where the helper transmits its description of the noise corrupting Block i during Block $i+1$, and the decoder decodes the message corresponding to Block i only after it has received this description, i.e., after Block $i+1$. This approach resolves the causality issue at the cost of some delay and also alleviates the need for bursty transmissions by the helper.

[1] We ignore the fact, τn need not be an integer. This could be remedied e.g. by choosing $\tau = \kappa/n$ for some integer κ (so that $\tau n = \kappa$). The limit $\tau \downarrow 0$ should then be understood as jointly letting $\kappa, n \to \infty$ at relative velocities such that $\tau = \kappa/n \to 0$.

III. The Additive Noise MAC

A two-to-one additive noise multiple-access channel (MAC) with a helper is depicted in Figure 2. Its time-k output is

$$Y_k = x_{1,k} + x_{2,k} + Z_k, \tag{21}$$

where $x_{1,k}$ and $x_{2,k}$ are the time-k inputs produced by the two users, and the noise sequence $\{Z_k\}$ comprises, as in the single-user case, IID random variables of second-moment N. In analogy to the single-user case, we assume that the inputs are subject to an average-power constraint

$$\frac{1}{n}\sum_{k=1}^{n} x_{i,k}(m_i)^2 \leq \mathsf{P}_i, \quad i \in \{1,2\}. \tag{22}$$

Here $x_{i,k}(m_i)$ denotes the time-k input produced by Transmitter i to convey its message M_i, where M_1 and M_2 are independent and of rates R_1 and R_2 respectively, and

$$\mathsf{P}_1, \mathsf{P}_2 > 0. \tag{23}$$

The helper observes the noise sequence and describes it to the receiver using nR_h bits.

Let $C(R_\mathrm{h})$ denote the capacity region of this MAC under the average probability of error criterion. The following theorem shows that $C(R_\mathrm{h})$ is the Minkowski sum of the capacity in the absence of a helper $C(0)$ and the set $\{(R_1,R_2) \in \mathbb{R}_+^2 \colon R_1 + R_2 \leq R_\mathrm{h}\}$.

Theorem 2: The capacity region of the additive noise MAC with a helper is

$$C(R_\mathrm{h}) = C(0) + \big\{(R_1,R_2) \in \mathbb{R}_+^2 \colon R_1 + R_2 \leq R_\mathrm{h}\big\} \tag{24}$$

where $\mathbb{R}_+$ denotes the nonnegative reals.

Proof: The helping strategy we propose is similar to the one we employed in the single-user case with Z^ℓ estimated as $\hat{Z}^\ell$ based on the helper's ℓR_h-bit description and with the corresponding estimation error $\tilde{Z}^\ell = Z^\ell - \hat{Z}^\ell$. For independent ℓ-vectors X_1^ℓ, X_2^ℓ, and $\tilde{Z}^\ell$, and with $\tilde{Y}^\ell = X_1^\ell + X_2^\ell + \tilde{Z}^\ell$ we have

$$\begin{aligned} I\big(X_1^\ell; \tilde{Y}^\ell \big| X_2^\ell\big) &= h\big(\tilde{Y}^\ell \big| X_2^\ell\big) - h\big(\tilde{Y}^\ell \big| X_1^\ell, X_2^\ell\big) \\ &= h\big(\tilde{Y}^\ell \big| X_2^\ell\big) - h\big(\tilde{Z}^\ell\big) \\ &\geq h\big(X_1^\ell\big) - h\big(\tilde{Z}^\ell\big). \end{aligned} \tag{25a}$$

Likewise,

$$I\big(X_2^\ell; \tilde{Y}^\ell \big| X_1^\ell\big) \geq h\big(X_2^\ell\big) - h\big(\tilde{Z}^\ell\big), \tag{25b}$$

and

$$I\big(X_1^\ell, X_2^\ell; \tilde{Y}^\ell\big) = h\big(\tilde{Y}^\ell\big) - h\big(\tilde{Z}^\ell\big) \geq h\big(X_1^\ell\big) - h\big(\tilde{Z}^\ell\big). \tag{25c}$$

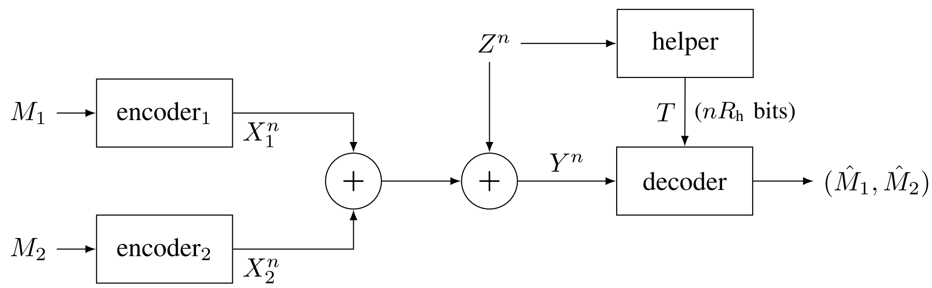


Fig. 2. The additive noise MAC with a helper.

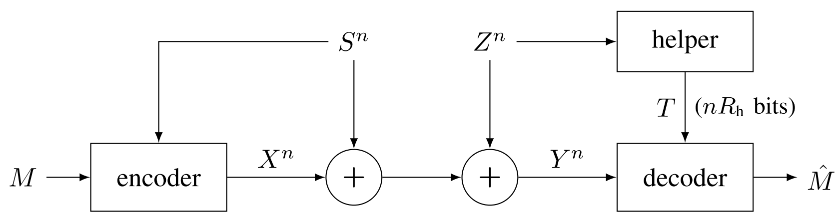


Fig. 3. The dirty-paper channel with a helper.

Suppose now that, in addition to being independent, the vectors X_1^ℓ and X_2^ℓ have components that are drawn IID according to some distribution P_X of second moment smaller than $\min\{\mathsf{P}_1, \mathsf{P}_2\}$ and having finite differential entropy. It then follows from (25) using (16) and (18) that (for sufficiently large ℓ)

$$\frac{1}{\ell} I\big(X_1^\ell; \tilde{Y}^\ell \big| X_2^\ell\big) \geq h(P_X) - \frac{1}{2}\log\big(2\pi e\,\mathsf{N}(1+\tilde{\delta})\big) + R_\mathrm{h}, \tag{26a}$$

$$\frac{1}{\ell} I\big(X_2^\ell; \tilde{Y}^\ell \big| X_1^\ell\big) \geq h(P_X) - \frac{1}{2}\log\big(2\pi e\,\mathsf{N}(1+\tilde{\delta})\big) + R_\mathrm{h}, \tag{26b}$$

$$\frac{1}{\ell} I\big(X_1^\ell, X_2^\ell; \tilde{Y}^\ell\big) \geq h(P_X) - \frac{1}{2}\log\big(2\pi e\,\mathsf{N}(1+\tilde{\delta})\big) + R_\mathrm{h}. \tag{26c}$$

The result now follows by coding over supersymbols of length ℓ and by considering flash helping, where, for $\tau \downarrow 0$, the channel is used with rate-R_h/τ help during τ of the time and without help during the remaining $(1-\tau)$ of the time.

The converse can be proved by augmenting the noise description to the channel output sequence and applying the standard steps of the MAC's converse to the resulting channel while noting that the entropy of the noise description cannot exceed nR_h. ■

Remark 2 ([10]): If the noise is Gaussian, then (24) continues to hold in the presence of feedback provided that $C(R_\mathrm{h})$ is interpreted as the feedback capacity and that $C(0)$ is therefore as given by Ozarow [11].

IV. Writing on Dirty Paper

Flash helping can also be applied to Costa's "writing on dirty paper" setting [8] when a helper observes the noise sequence. Such a scenario is depicted in Figure 3. The time-k channel output Y_k is

$$Y_k = X_k + S_k + Z_k, \tag{27}$$

where X_k, S_k, and Z_k are the time-k channel input, state, and noise, respectively. The encoder observes the state sequence S^n noncausally, and the symbol $X_k(m, S^n)$ it produces at time k is thus a function of both the message m and the observed state sequence S^n. The average-power constraint we consider requires that, for each message $m \in \mathcal{M}$,

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[X_k(m, S^n)^2\right] \leq \mathsf{P}, \tag{28}$$

where the expectation is over the state sequence S^n, and we assume that $\mathsf{P} > 0$.

The state sequence $\{S_k\}$ and the noise sequence $\{Z_k\}$ are independent, with the former being IID $\sim \mathcal{N}(0, \sigma_S^2)$ and the latter IID $\sim \mathcal{N}(0, \sigma_Z^2)$. Their joint law does not depend on the message m.

The noise sequence $\{Z_k\}$ is observed by the helper who describes it to the decoder at the rate R_h. The capacity we seek is denoted $C(R_\mathrm{h})$. It was computed in the absence of a helper by Costa [8] who showed that

$$C(0) = \frac{1}{2}\log\left(1 + \frac{\mathsf{P}}{\sigma_Z^2}\right). \tag{29}$$

Theorem 3: The capacity $C(R_\mathrm{h})$ of the dirty-paper channel with a helper is

$$C(R_\mathrm{h}) = C(0) + R_\mathrm{h}. \tag{30}$$

Proof: The converse follows from the Cut-Set Bound [9, Thm. 15.10.1], so we focus on achievability. Once again, we use flash helping by considering a time-sharing scheme with duty-cycle $0 < \tau < 1$, which we later drive to zero.

During $1-\tau$ of the time, we use Costa's dirty-paper coding without any help. The achievable rate during this period is the RHS of (29). In the remaining time, the helper describes the noise sequence at rate R_h/τ, and the encoder uses the single-user scheme of Section II with a twist: it subtracts the state sequence. This subtraction comes at a power cost.

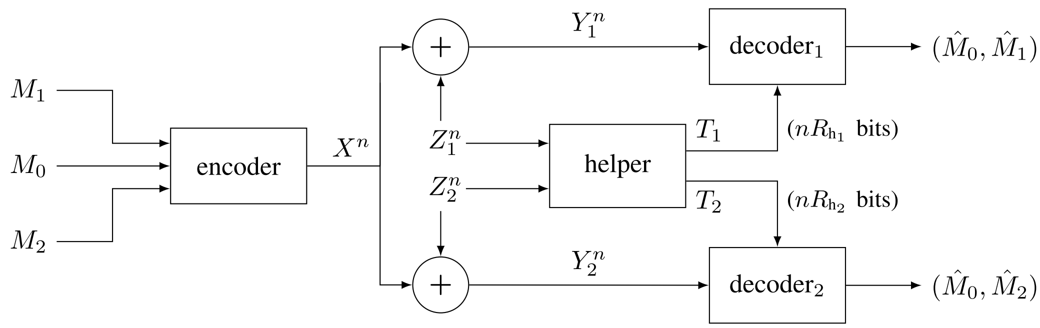


Fig. 4. The broadcast channel with a helper.

Indeed, since the variance of the state is σ_S^2, subtracting the sequence during τn channel uses increases the transmitted power by $\tau\sigma_S^2$. This additional power, however, will wash out when we later consider[2] $\tau \downarrow 0$. (The continuity of $C(R_{\mathrm{h}})$ in P for $\mathsf{P} > 0$ can be established by showing that $C(R_{\mathrm{h}})$ is concave in P using a time-sharing argument.) The achievable rate during this period is lower-bounded by the RHS of (19). The rate achieved by the time-sharing scheme can be thus lower-bounded as in (20). As $\tau \downarrow 0$ it converges to the RHS of (30), and the excess power that results from subtracting the state sequence tends to zero. ■

The proof extends verbatim to the case where the noise and state are not necessarily Gaussian, provided that they have a finite second moment:

Remark 3: As long as the noise and state are of finite second moment, the relation $C(R_{\mathrm{h}}) = C(0) + R_{\mathrm{h}}$ continues to hold provided that we interpret $C(0)$ as the capacity without a helper for the given (not necessarily Gaussian) noise and state distributions.

V. The Gaussian Additive Noise Broadcast Channel

Flash helping can also be used to communicate over the Gaussian broadcast channel with a helper, a network which is depicted in Figure 4. However, some additional work is needed because, in the presence of a helper, the notion of degradedness is more involved: with help, the more-noisy receiver may actually be the better receiver.

The time-k channel outputs observed by the respective receivers are

$$Y_{1,k} = x_k + Z_{1,k}, \tag{31a}$$

$$Y_{2,k} = x_k + Z_{2,k}, \tag{31b}$$

where x_k is the time-k input, and where the noise tuples $\{(Z_{1,k}, Z_{2,k})\}$ are IID centered bivariate Gaussians of covariance matrix

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{32}$$

and we assume without loss of generality that

$$\sigma_2 \geq \sigma_1 > 0. \tag{33}$$

The noise tuples are observed by a helper who can describe them to the two decoders at respective rates R_{h_1} and R_{h_2}.

[2]This loss can be avoided if one uses Costa's scheme.

The encoder wishes to transmit three independent messages M_0, M_1, and M_2 at rates R_0, R_1, and R_2. Both decoders must recover the common message M_0 reliably, while only Decoder 1 must recover M_1 and only Decoder 2 must recover M_2.

The encoder ϕ_{enc} is thus a mapping

$$\begin{aligned} \phi_{\mathrm{enc}} \colon & \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \to \mathbb{R}^n, \\ & (m_0, m_1, m_2) \mapsto \mathbf{x}(m_0, m_1, m_2) = (x_1, \ldots, x_n), \end{aligned} \tag{34}$$

where $\mathcal{M}_i = \{1, \ldots, 2^{nR_i}\}$, $i \in \{0, 1, 2\}$, are the message sets, and where we impose the average-power constraint that for all message tuples (m_0, m_1, m_2)

$$\|\mathbf{x}(m_0, m_1, m_2)\|^2 \leq n\mathsf{P}. \tag{35}$$

A helper is a mapping

$$\begin{aligned} \phi_{\mathrm{help}} \colon & \mathbb{R}^n \times \mathbb{R}^n \to \{0,1\}^{nR_{\mathrm{h}_1}} \times \{0,1\}^{nR_{\mathrm{h}_2}}, \\ & (\mathbf{z}_1, \mathbf{z}_2) \mapsto \bigl(t_1(\mathbf{z}_1, \mathbf{z}_2), t_2(\mathbf{z}_1, \mathbf{z}_2)\bigr), \end{aligned} \tag{36}$$

with the understanding that $T_1(\mathbf{Z}_1, \mathbf{Z}_2)$ is presented to Decoder 1 and $T_2(\mathbf{Z}_1, \mathbf{Z}_2)$ to Decoder 2. The decoders are thus mappings

$$\psi_1 \colon \mathbb{R}^n \times \{0,1\}^{nR_{\mathrm{h}_1}} \to \mathcal{M}_0 \times \mathcal{M}_1, \ (\mathbf{y}_1, t_1) \mapsto (\hat{m}_0, \hat{m}_1), \tag{37a}$$

$$\psi_2 \colon \mathbb{R}^n \times \{0,1\}^{nR_{\mathrm{h}_2}} \to \mathcal{M}_0 \times \mathcal{M}_2, \ (\mathbf{y}_2, t_2) \mapsto (\hat{m}_0, \hat{m}_2). \tag{37b}$$

The capacity with respect to the average-probability-of-error-criterion is denoted $C(R_{\mathrm{h}_1}, R_{\mathrm{h}_2})$ and is characterized in the following theorem.

Theorem 4: The capacity region of the Gaussian broadcast channel with separate helping bit-pipes to each of the decoders is (with $\bar{\alpha} \triangleq 1 - \alpha$)

$$\begin{aligned} C(R_{\mathrm{h}_1}, R_{\mathrm{h}_2}) = \bigcup_{\alpha \in [0,1]} \biggl\{ & (R_0, R_1, R_2) \in \mathbb{R}_+^3 \colon \\ & R_0 + R_2 \leq \frac{1}{2} \log\left(1 + \frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P} + \sigma_2^2}\right) + R_{\mathrm{h}_2} \\ & R_1 \leq \frac{1}{2} \log\left(1 + \frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1} \\ & R_0 + R_1 \leq \frac{1}{2} \log\left(1 + \frac{\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1} \biggr\}. \end{aligned} \tag{38}$$

Proof: The proof is deferred to the appendix. ■

Remark 4: The capacity region does not depend on the noise correlation ρ. Moreover, it can be achieved by a helper of the form

$$(\mathbf{z}_1, \mathbf{z}_2) \mapsto \big(t_1(\mathbf{z}_1), t_2(\mathbf{z}_2)\big), \tag{39}$$

i.e., a helper that—rather than describing the noise sequences jointly—describes them separately, each to the appropriate decoder.

Remark 5: Achievability holds also if the noises are not Gaussian or not memoryless: it suffices that their empirical second moment converge in probability to finite σ_1^2 and σ_2^2.

When only a common message is to be sent (i.e., $R_1 = R_2 = 0$), Theorem 4 implies that the largest achievable rate is

$$\min_{i\in\{1,2\}} \left\{ \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_i^2}\right) + R_{\mathrm{h}_i} \right\}. \tag{40}$$

This can be generalized to non-Gaussian noises and to more receivers:

Remark 6 ([10]): The largest achievable common-message rate $C(R_{\mathrm{h}_1}, \dots, R_{\mathrm{h}_\mathsf{L}})$ to L receivers is

$$\max_{P_X} \min_{i\in[1:\mathsf{L}]} \big\{ I(X;Y_i) + R_{\mathrm{h}_i} \big\}, \tag{41}$$

where the maximum is over all input distributions P_X under which $\mathbb{E}_P[X^2] \le \mathsf{P}$.

VI. The Modulo-Additive Noise Channel

This section demonstrates that a discrete form of flash-helping can sometimes even achieve the capacity of finite-alphabet channels. The model we treat is the single-user modulo-additive noise channel. The extension to the MAC is straightforward.

Consider the modulo-additive noise channel whose time-k output Y_k is

$$Y_k = x_k + Z_k, \tag{42}$$

where x_k, Z_k, and Y_k all take values in the modulo-additive group $\mathcal{X} = \{0, 1, \dots, |\mathcal{X}| - 1\}$, and "+" denotes addition modulo $|\mathcal{X}|$. Irrespective of the input sequence $\{x_k\}$, the noise sequence $\{Z_k\}$ is IID $\sim P_Z$, with corresponding entropy $H(Z)$. In the absence of a helper, the capacity is [9, Thm. 7.2.1]

$$C(0) = \log|\mathcal{X}| - H(Z). \tag{43}$$

Suppose now that the noise sequence is observed by a helper who then describes it to the decoder at rate R_h.

As can be inferred from [2], the capacity is $\min\{C(0) + R_\mathrm{h}, \log|\mathcal{X}|\}$. Unlike Kim's proof, which uses binning, here we will show achievability using flash helping. The converse follows from the Cut-Set Bound and is omitted.

Theorem 5: The capacity of the modulo-additive noise channel with a rate-R_h helper is

$$C(R_\mathrm{h}) = \min\big\{C(0) + R_\mathrm{h}, \log|\mathcal{X}|\big\}. \tag{44}$$

Proof: If $R_\mathrm{h} > H(Z)$, then the helper can describe the noise sequence (almost) losslessly. The receiver can then subtract it and thus obtain a noise-free channel of capacity $\log|\mathcal{X}|$. We therefore focus on the case where $R_\mathrm{h} \le H(Z)$.

In this case the helper can use flash helping by using nR_h bits to describe the first $n(R_\mathrm{h}/H(Z))$ noise samples and by not describing the remaining noise samples at all. The receiver then subtracts the noise samples for which it has a description.

When the noise is subtracted, the rate $\log|\mathcal{X}|$ is achievable, and when not, $\log|\mathcal{X}| - H(Z)$. The aggregate rate is thus

$$\frac{R_\mathrm{h}}{H(Z)}\log|\mathcal{X}| + \left(1 - \frac{R_\mathrm{h}}{H(Z)}\right)\big(\log|\mathcal{X}| - H(Z)\big), \tag{45}$$

which simplifies to $\log|\mathcal{X}| - H(Z) + R_\mathrm{h}$, i.e., to $C(0) + R_\mathrm{h}$. ∎

Appendix
Proof of Theorem 4

A. Achievability

The achievability proof is based on the classical random coding argument. We generate codebooks at random and show that the different probabilities of error (and hence also their sum) vanish as the blocklength tends to infinity. We then conclude that there must exist some choice of the codebooks for which the sum of (and hence all) the probabilities of error vanish.

Given some (small) $0 < \tau < 1$, we split the blocklength n into three intervals of lengths[3] $(1-2\tau)n$, τn, and τn. We refer to the first interval as the "common interval," to the second as the "Decoder-1 interval," and to the last as the "Decoder-2 interval." In our analysis we shall first drive n to infinity and, afterwards, drive τ to zero.

Unless otherwise specified, we shall assume that all the vectors/tuples in this proof are row-vectors. For a generic n-vector $\mathbf{w}$, we use $\mathbf{w}_{(1-2\tau)}$ to denote its first $n(1-2\tau)$ components stacked into an $n(1-2\tau)$-dimensional row-vector; we use $\mathbf{w}_{(\tau_1)}$ to denote the next $n\tau$ components stacked into an $n\tau$-dimensional row-vector, and $\mathbf{w}_{(\tau_2)}$ to denote the last $n\tau$ components of $\mathbf{w}$ stacked into an $n\tau$-dimensional row-vector. Because all the vectors are row-vectors, we can express $\mathbf{w}$ in terms of $\mathbf{w}_{(1-2\tau)}$, $\mathbf{w}_{(\tau_1)}$, and $\mathbf{w}_{(\tau_2)}$ as

$$\mathbf{w} = \mathbf{w}_{(1-2\tau)}{}^\frown\mathbf{w}_{(\tau_1)}{}^\frown\mathbf{w}_{(\tau_2)}, \tag{46}$$

where "$\frown$" indicates concatenation. For example, the noise sequence $\mathbf{z}_1$ experienced by Receiver 1 is an n-tuple that can be written as $\mathbf{z}_{1,(1-2\tau)}{}^\frown\mathbf{z}_{1,(\tau_1)}{}^\frown\mathbf{z}_{1,(\tau_2)}$.

In our coding scheme, the helper provides Decoder 1 with an nR_{h_1}-bit description of $\mathbf{z}_{1,(\tau_1)}$ and Decoder 2 with an nR_{h_2}-bit description of $\mathbf{z}_{2,(\tau_2)}$. Decoder 1 ignores $\mathbf{y}_{1,(\tau_2)}$ and Decoder 2 ignores $\mathbf{y}_{2,(\tau_1)}$. This explains our naming of the intervals.

[3] We ignore the fact that $n\tau$ need not be an integer. This can be remedied in a way that does not affect the rates, see Footnote 1.

1) Encoding: We use *rate-splitting* for the private messages. We decompose M_1 of rate R_1 into $M_1 = \big(M_1^{(0)}, M_1^{(1)}\big)$ of rates $R_1^{(0)}$, $R_1^{(1)}$, and M_2 of rate R_2 into $M_2 = \big(M_2^{(0)}, M_2^{(1)}\big)$ of rates $R_2^{(0)}$, $R_2^{(1)}$. The general form of the transmitted n-tuple $\mathbf{x}\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}, m_2^{(1)}\big)$ that is used to convey the messages $\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}, m_2^{(1)}\big)$ is

$$\begin{aligned}&\mathbf{x}\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}, m_2^{(1)}\big)\\ &\quad= \Big(\mathbf{u}\big(m_0, m_2^{(0)}\big) + \mathbf{v}_{(1-2\tau)}\big(m_1^{(0)}\big)\Big)\\ &\qquad\frown \mathbf{v}_{(\tau_1)}\big(m_1^{(1)}\big|m_0, m_2^{(0)}\big)\frown \mathbf{v}_{(\tau_2)}\big(m_2^{(1)}\big|m_0, m_2^{(0)}\big), \quad (47)\end{aligned}$$

where the terms appearing on the RHS will be described shortly. Given some $0 \leq \alpha \leq 1$, we draw the $2^{n(R_0+R_2^{(0)})}$ length-$n(1-2\tau)$ codewords $\big\{\mathbf{u}\big(m_0, m_2^{(0)}\big)\big\}$ independently, with the components of each being drawn IID $\sim \mathcal{N}(0, \alpha\mathsf{P})$. Independently of these codewords, we draw the $2^{nR_1^{(0)}}$ length-$n(1-2\tau)$ codewords $\big\{\mathbf{v}_{(1-2\tau)}\big(m_1^{(0)}\big)\big\}$ independently, with the components being IID $\sim \mathcal{N}(0, \bar{\alpha}\mathsf{P})$. Here and throughout, $\bar{\alpha} \triangleq 1-\alpha$.

For each codeword $\mathbf{u}\big(m_0, m_2^{(0)}\big)$, we independently draw $2^{nR_1^{(1)}}$ length-$n\tau$ codewords $\big\{\mathbf{v}_{(\tau_1)}\big(m_1^{(1)}\big|m_0, m_2^{(0)}\big)\big\}$ with the components of each being drawn IID $\sim \mathcal{N}(0, \mathsf{P})$. Similarly, and independently of those, for each $\mathbf{u}\big(m_0, m_2^{(0)}\big)$, we independently draw $2^{nR_2^{(1)}}$ length-$n\tau$ codewords $\big\{\mathbf{v}_{(\tau_2)}\big(m_2^{(1)}\big|m_0, m_2^{(0)}\big)\big\}$ with the components of each being drawn IID $\sim \mathcal{N}(0, \mathsf{P})$. The transmitted n-tuple is now as given in (47).[4]

Decoder 1 ignores $\mathbf{y}_{1,(\tau_2)}$, and Decoder 2 ignores $\mathbf{y}_{2,(\tau_1)}$, so the relevant sections of $\mathbf{x}$ for Decoder 1 and 2 are

$$\begin{aligned}&\mathbf{x}^{(1)}\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}\big)\\ &= \Big(\mathbf{u}\big(m_0, m_2^{(0)}\big) + \mathbf{v}_{(1-2\tau)}\big(m_1^{(0)}\big)\Big)\frown \mathbf{v}_{(\tau_1)}\big(m_1^{(1)}\big|m_0, m_2^{(0)}\big), \quad (48)\\ &\mathbf{x}^{(2)}\big(m_0, m_1^{(0)}, m_2^{(0)}, m_2^{(1)}\big)\\ &= \Big(\mathbf{u}\big(m_0, m_2^{(0)}\big) + \mathbf{v}_{(1-2\tau)}\big(m_1^{(0)}\big)\Big)\frown \mathbf{v}_{(\tau_2)}\big(m_2^{(1)}\big|m_0, m_2^{(0)}\big). \quad (49)\end{aligned}$$

2) Helping: The helper provides Decoder 1 with an nR_{h_1}-bit description of $\mathbf{z}_{1,(\tau_1)}$ and Decoder 2 with an nR_{h_2}-bit description of $\mathbf{z}_{2,(\tau_2)}$. To describe $\mathbf{z}_{1,(\tau_1)}$, a random codebook is generated containing $2^{n\tau\tilde{R}_{\mathsf{h}_1}/\tau} = 2^{n\tilde{R}_{\mathsf{h}_1}}$ codewords $\{\hat{\mathbf{z}}_{1,(\tau_1)}\}$ that are drawn independently and uniformly over the $n\tau$-dimensional sphere of radius

$$r = \sqrt{n\tau\sigma_1^2\left(1 - 2^{-2R_{\mathsf{h}_1}/\tau}\right)}. \quad (50)$$

The noise sequence $\mathbf{z}_{1,(\tau_1)}$ is then described by the codeword $\hat{\mathbf{z}}^*_{1,(\tau_1)}$ that has an "almost orthogonal" error vector

$$\hat{\mathbf{z}}^*_{1,(\tau_1)} = \underset{\hat{\mathbf{z}}_{1,(\tau_1)}}{\operatorname{argmin}}\, \big\langle \hat{\mathbf{z}}_{1,(\tau_1)}, \mathbf{z}_{1,(\tau_1)} - \hat{\mathbf{z}}_{1,(\tau_1)} \big\rangle. \quad (51)$$

The helper provides Decoder 1 with the index of $\hat{\mathbf{z}}^*_{1,(\tau_1)}$.

[4]Drawing the codewords according to these distributions will typically result in minor violations of the power constraint. This can be remedied in the standard way of backing-off in the power by ϵ and then expurgating the codewords that violate the power constraint [9], [12].

Using standard results on the covering of the unit n-sphere [13], [14], it can be shown that, for any $\delta > 0$, for some choice of $\tilde{R}_{\mathsf{h}_1} < R_{\mathsf{h}_1} + \delta$ asymptotic orthogonality holds in probability:

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n\tau} \big\langle \hat{\mathbf{z}}^*_{1,(\tau_1)}, \mathbf{z}_{1,(\tau_1)} - \hat{\mathbf{z}}^*_{1,(\tau_1)} \big\rangle = 0. \quad (52)$$

Using this asymptotic orthogonality and by opening the square,

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n\tau} \big\| \mathbf{z}_{1,(\tau_1)} - \hat{\mathbf{z}}^*_{1,(\tau_1)} \big\|^2 = \sigma_1^2\, 2^{-2R_{\mathsf{h}_1}/\tau}. \quad (53)$$

The procedure for describing $\mathbf{z}_{2,(\tau_2)}$ to Decoder 2 is analogous: $2^{n\tilde{R}_{\mathsf{h}_2}}$ codewords $\{\hat{\mathbf{z}}_{2,(\tau_2)}\}$ are drawn independently uniformly over the $n\tau$-dimensional sphere of radius

$$r = \sqrt{n\tau\sigma_2^2\left(1 - 2^{-2R_{\mathsf{h}_2}/\tau}\right)}. \quad (54)$$

The codeword $\hat{\mathbf{z}}^*_{2,(\tau_2)}$ describing $\mathbf{z}_{2,(\tau_2)}$ is the one leading to the error that is most orthogonal, and, for any $\delta > 0$ and for some $\tilde{R}_{\mathsf{h}_2} < R_{\mathsf{h}_2} + \delta$,

$$\operatorname*{p-lim}_{n\to\infty} \frac{1}{n\tau} \big\| \mathbf{z}_{2,(\tau_2)} - \hat{\mathbf{z}}^*_{2,(\tau_2)} \big\|^2 = \sigma_2^2\, 2^{-2R_{\mathsf{h}_2}/\tau}. \quad (55)$$

3) Decoding: Decoder 1 ignores the sequence $\mathbf{y}_{1,(\tau_2)}$ that it receives during the Decoder-2 interval and bases its decision on $\mathbf{y}_{1,(1-2\tau)}\frown\mathbf{y}_{1,(\tau_1)}$ only. From $\mathbf{y}_{1,(\tau_1)}$ it subtracts the description $\hat{\mathbf{z}}^*_{1,(\tau_1)}$ of $\mathbf{z}_{1,(\tau_1)}$ to form

$$\hat{\mathbf{y}}_{1,(\tau_1)} = \mathbf{y}_{1,(\tau_1)} - \hat{\mathbf{z}}^*_{1,(\tau_1)}. \quad (56)$$

For consistency we also define $\hat{\mathbf{y}}_{1,(1-2\tau)} \triangleq \mathbf{y}_{1,(1-2\tau)}$ and $\hat{\mathbf{y}}_1 \triangleq \hat{\mathbf{y}}_{1,(1-2\tau)}\frown\hat{\mathbf{y}}_{1,(\tau_1)}$.

Decoder 1 decodes the messages intended for it (as well as $\hat{m}_2^{(0)}$, which it discards) based on $\hat{\mathbf{y}}_1$ by assuming that in the common interval the noise is Gaussian with variance σ_1^2, that in the Decoder-1 interval it is Gaussian with variance $\sigma_1^2 2^{-2R_{\mathsf{h}_1}/\tau}$, and that in both intervals the signal power is P. It thus employs scaled nearest-neighbor decoding and produces

$$\begin{aligned}&\big(\hat{m}_0, \hat{m}_1^{(0)}, \hat{m}_1^{(1)}, \hat{m}_2^{(0)}\big)\\ &\quad= \underset{(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)})}{\operatorname{argmin}} D_1\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}\big), \quad (57)\end{aligned}$$

where

$$\begin{aligned}&D_1\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}\big)\\ &\triangleq \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \Big(\hat{y}_{1,k} - x_k^{(1)}\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}\big)\Big)^2\\ &\quad+ \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} 2^{2R_{\mathsf{h}_1}/\tau}\Big(\hat{y}_{1,k} - x_k^{(1)}\big(m_0, m_1^{(0)}, m_1^{(1)}, m_2^{(0)}\big)\Big)^2 \quad (58)\\ &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \Big(\hat{y}_{1,k} - u_k(m_0, m_2^{(0)}) - v_{(1-2\tau),k}(m_1^{(0)})\Big)^2\\ &\quad+ \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} 2^{2R_{\mathsf{h}_1}/\tau}\Big(\hat{y}_{1,k} - v_{(\tau_1),k}\big(m_1^{(1)}\big|m_0, m_2^{(0)}\big)\Big)^2 \quad (59)\end{aligned}$$

is the decoding metric.

Decoder 2 ignores the sequence $\mathbf{y}_{2,(\tau_1)}$ and bases its decision on $\hat{\mathbf{y}}_2 \triangleq \hat{\mathbf{y}}_{2,(1-2\tau)} {}^\frown \hat{\mathbf{y}}_{2,(\tau_2)}$, where $\hat{\mathbf{y}}_{2,(1-2\tau)} = \mathbf{y}_{2,(1-2\tau)}$ and where $\hat{\mathbf{y}}_{2,(\tau_2)} = \mathbf{y}_{2,(\tau_2)} - \hat{\mathbf{z}}^*_{2,(\tau_2)}$. It does not attempt to decode $m_1^{(0)}$ and hence treats the superpositioned codeword $\mathbf{v}_{(1-2\tau)}$ of the common interval as noise. It assumes therefore that the noise is Gaussian with variance $\sigma_2^2 + \bar{\alpha}\mathsf{P}$ during the common interval and with variance $\sigma_2^2\, 2^{-2R_{\mathrm{h}_1}/\tau}$ during the Decoder-2 interval, and that the signal power is $\alpha\mathsf{P}$ during the common interval and P during the Decoder-2 interval. It therefore produces the guess

$$\left(\hat{m}_0, \hat{m}_2^{(0)}, \hat{m}_2^{(1)}\right) = \operatorname*{argmin}_{(m_0, m_2^{(0)}, m_2^{(1)})} D_2\left(m_0, m_2^{(0)}, m_2^{(1)}\right), \quad (60)$$

where

$$\begin{aligned} &D_2\left(m_0, m_2^{(0)}, m_2^{(1)}\right) \\ &\triangleq \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P} + \sigma_2^2}\left(\hat{y}_{2,k} - u_k\left(m_0, m_2^{(0)}\right)\right)^2 \\ &\quad + \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} 2^{2R_{\mathrm{h}_2}/\tau}\left(\hat{y}_{2,k} - v_{(\tau_2),k}\left(m_2^{(1)} \middle| m_0, m_2^{(0)}\right)\right)^2 \end{aligned} \quad (61)$$

is the decoding metric. Note that the second sum in (61) ranges over the indices from $n(1-2\tau)+1$ to $n(1-\tau)$ because, by construction, $\hat{\mathbf{y}}_2$ is a vector of length $n(1-\tau)$.

4) Error Analysis: Since the codebooks are Gaussian, and since the decoding is scaled nearest-neighbor, we can employ the technique of [15] and [16]. The analysis is, however, a bit more involved because we are in a multiple-user setting.

Since our codewords are chosen independently with identical distribution, we may assume without loss of generality that the transmitted message is $\left(m_0,\, m_1^{(0)},\, m_1^{(1)},\, m_2^{(0)},\, m_2^{(1)}\right) = (1,1,1,1,1)$ and denote the corresponding codeword $\mathbf{x}(1,1,1,1,1)$ or $\mathbf{x}(\mathbf{1})$. Note that by independence, by the weak law of large numbers, and by (53) and (55),

$$\operatorname*{p\text{-}lim}_{n\to\infty} \frac{1}{n}\left\|\hat{\mathbf{y}}_{1,(1-2\tau)}\right\|^2 = (1-2\tau)\left(\mathsf{P} + \sigma_1^2\right), \quad (62)$$

$$\operatorname*{p\text{-}lim}_{n\to\infty} \frac{1}{n}\left\|\hat{\mathbf{y}}_{1,(\tau_1)}\right\|^2 = \tau\left(\mathsf{P} + 2^{-2R_{\mathrm{h}_1}/\tau}\,\sigma_1^2\right), \quad (63)$$

$$\operatorname*{p\text{-}lim}_{n\to\infty} \frac{1}{n}\left\|\hat{\mathbf{y}}_{2,(1-2\tau)}\right\|^2 = (1-2\tau)\left(\mathsf{P} + \sigma_2^2\right), \quad (64)$$

$$\operatorname*{p\text{-}lim}_{n\to\infty} \frac{1}{n}\left\|\hat{\mathbf{y}}_{2,(\tau_2)}\right\|^2 = \tau\left(\mathsf{P} + 2^{-2R_{\mathrm{h}_2}/\tau}\,\sigma_2^2\right). \quad (65)$$

Consider now first the chance of a decoding error at Decoder 1: From the decoding metric definition (59) it follows that, as $n \to \infty$, the correct codeword will accumulate the metric $\sigma_1^2(1-\tau)$ with probability tending to one. The probability of a decoding error is therefore upper-bounded asymptotically by the probability that some other codeword accumulates a metric below σ_1^2. We next analyze the rate constraints that guarantee that this probability vanish as $n \to \infty$. We have to consider four different error cases:

- Case 1: $\left(\hat{m}_0, \hat{m}_2^{(0)}\right) \neq (1,1)$, $\hat{m}_1^{(0)} \neq 1$
  In this case, a candidate codeword $\mathbf{X}$ is independent of $\hat{\mathbf{y}}_1$, regardless of whether or not $\hat{m}_1^{(1)} = 1$ (since the codebooks for $M_1^{(1)}$ are independent of each other for different $\left(m_0, m_2^{(0)}\right)$). An arbitrary codeword $\mathbf{X}$ drawn IID $\sim \mathcal{N}(0, \mathsf{P})$ induces a random metric D_1,

$$\begin{aligned} D_1 &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \left(\hat{y}_{1,k} - X_k\right)^2 \\ &\quad + \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} 2^{2R_{\mathrm{h}_1}/\tau}\left(\hat{y}_{1,k} - X_k\right)^2. \end{aligned} \quad (66)$$

Conditioned on $\hat{\mathbf{y}}_1$, the logarithmic moment generating function (MGF) of D_1 is given for any $\theta < 0$ as

$$\begin{aligned} \Lambda_n(\theta) &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \frac{\hat{y}_{1,k}^2\theta}{1-2\theta\mathsf{P}/n} - \frac{1}{2}\sum_{k=1}^{n(1-2\tau)} \log(1-2\theta\mathsf{P}/n) \\ &\quad + \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{1,k}^2\theta}{1 - 2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}/n} \\ &\quad - \frac{1}{2}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\left(1 - 2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}/n\right). \end{aligned} \quad (67)$$

Therefore, by (62) and (63), it follows that

$$\begin{aligned} &\frac{1}{n}\Lambda_n(n\theta) \\ &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \frac{\hat{y}_{1,k}^2\theta}{1-2\theta\mathsf{P}} - \frac{1}{2n}\sum_{k=1}^{n(1-2\tau)} \log(1-2\theta\mathsf{P}) \\ &\quad + \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{1,k}^2\theta}{1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}} \\ &\quad - \frac{1}{2n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\left(1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}\right) \quad (68) \\ &= \frac{\theta}{1-2\theta\mathsf{P}} \cdot \frac{1}{n}\sum_{k=1}^{n(1-2\tau)} \hat{y}_{1,k}^2 - \frac{1-2\tau}{2}\log(1-2\theta\mathsf{P}) \\ &\quad + \frac{\theta}{1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}} \cdot \frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \hat{y}_{1,k}^2 \\ &\quad - \frac{\tau}{2}\log\left(1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}\right) \quad (69) \\ &\overset{n\to\infty}{\to} (1-2\tau)\,\frac{(\mathsf{P}+\sigma_1^2)\theta}{1-2\theta\mathsf{P}} - \frac{1-2\tau}{2}\log(1-2\theta\mathsf{P}) \\ &\quad + \tau\,\frac{\left(2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P} + \sigma_1^2\right)\theta}{1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}} \\ &\quad - \frac{\tau}{2}\log\left(1-2\theta\, 2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}\right) \quad (70) \\ &\triangleq \Lambda(\theta) \quad (71) \end{aligned}$$

with probability one. By the Gärtner–Ellis Theorem [17, Thm. 2.3.6],

$$\lim_{n\to\infty} \frac{1}{n}\log\Pr\left(D_1 < \sigma_1^2\right) = -\Lambda^*\left(\sigma_1^2\right) \quad (72)$$

where

$$\Lambda^*\big(\sigma_1^2\big) = \sup_{\theta<0}\big\{\theta\sigma_1^2 - \Lambda(\theta)\big\} \tag{73}$$

is the Fenchel–Legendre transform of $\Lambda(\cdot)$. In other words, for the case at hand, the probability of error for a single random codeword $\mathbf{X}$ decays exponentially with error exponent $\Lambda^*(\sigma_1^2)$. Hence, by the union bound, as long as the number of candidate codewords $\mathbf{X}$ grows exponentially at an exponent below $\Lambda^*(\sigma_1^2)$, the probability of this error event decays to zero as $n \to \infty$. Furthermore, for any $\theta < 0$, $\theta\sigma_1^2 - \Lambda(\theta)$ represents a lower bound on $\Lambda^*(\sigma_1^2)$. The choice of $\theta = -\frac{1}{2\sigma_1^2}$ establishes

$$\begin{aligned}\Lambda^*\big(\sigma_1^2\big) &\geq -\frac{1}{2\sigma_1^2}\cdot\sigma_1^2 - \Lambda\left(-\frac{1}{2\sigma_1^2}\right) &(74)\\ &= \frac{1-2\tau}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right)\\ &\quad+\frac{\tau}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\cdot 2^{2R_{\mathrm{h}_1}/\tau}\right)-\frac{\tau}{2}. &(75)\end{aligned}$$

Therefore, the probability of the current error case vanishes as long as

$$\begin{aligned}&R_0+\underbrace{R_1^{(0)}+R_1^{(1)}}_{R_1}+R_2^{(0)}\\ &<\frac{1-2\tau}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right)+\frac{\tau}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\cdot 2^{2R_{\mathrm{h}_1}/\tau}\right)\\ &\quad-\frac{\tau}{2}. \end{aligned}\tag{76}$$

Upon letting $\tau \downarrow 0$, we see that, as long as the rates satisfy

$$R_0+R_1+R_2^{(0)}<\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right)+R_{\mathrm{h}_1}, \tag{77}$$

the probability of an error of the kind at hand vanishes.

- Case 2: $\big(\hat{m}_0,\hat{m}_2^{(0)}\big)\neq(1,1)$, $\hat{m}_1^{(0)}=1$
In this case, the metric D_1 reduces to

$$\begin{aligned}D_1 &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)}\big(u_k(1,1)+z_{1,k}-U_k\big)^2\\ &\quad+\frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)}2^{2R_{\mathrm{h}_1}/\tau}\big(\hat{y}_{1,k}-X_k\big)^2, \end{aligned}\tag{78}$$

where $\mathbf{u}(1,1)$ and $\mathbf{U}$ are independent. Conditioned on $\mathbf{u}(1,1)+\mathbf{z}_1$ and on $\hat{\mathbf{y}}_{1,(\tau_1)}$, the logarithmic MGF of D_1 is given for $\theta<0$ as

$$\begin{aligned}\Lambda_n(\theta) &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)}\frac{\big(u_k(1,1)+z_{1,k}\big)^2\theta}{1-2\theta\alpha\mathsf{P}/n}\\ &\quad-\frac{1}{2}\sum_{k=1}^{n(1-2\tau)}\log(1-2\theta\alpha\mathsf{P}/n)\\ &\quad+\frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)}\frac{\hat{y}_{1,k}^2\theta}{1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}/n}\\ &\quad-\frac{1}{2}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)}\log\big(1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}/n\big). \end{aligned}\tag{79}$$

By the law of large numbers and by (63), we get

$$\begin{aligned}&\frac{1}{n}\Lambda_n(n\theta)\\ &= \frac{1}{n}\sum_{k=1}^{n(1-2\tau)}\frac{(u_k(1,1)+z_{1,k})^2\theta}{1-2\theta\alpha\mathsf{P}}\\ &\quad-\frac{1}{2n}\sum_{k=1}^{n(1-2\tau)}\log(1-2\theta\alpha\mathsf{P})\\ &\quad+\frac{1}{n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)}\frac{\hat{y}_{1,k}^2\theta}{1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}}\\ &\quad-\frac{1}{2n}\sum_{k=n(1-2\tau)+1}^{n(1-\tau)}\log\big(1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}\big) &(80)\\ &\overset{n\to\infty}{\to}(1-2\tau)\frac{(\alpha\mathsf{P}+\sigma_1^2)\theta}{1-2\theta\alpha\mathsf{P}}-\frac{1-2\tau}{2}\log(1-2\theta\alpha\mathsf{P})\\ &\quad+\tau\frac{\big(2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}+\sigma_1^2\big)\theta}{1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}}\\ &\quad-\frac{\tau}{2}\log\big(1-2\theta\,2^{2R_{\mathrm{h}_1}/\tau}\,\mathsf{P}\big) &(81)\\ &\triangleq\Lambda(\theta) &(82)\end{aligned}$$

with probability one. We invoke the Gärtner–Ellis Theorem as in Case 1 and bound the Fenchel–Legendre transform $\Lambda^*(\sigma_1^2)$ from below by $-\frac{1}{2\sigma_1^2}\cdot\sigma_1^2-\Lambda\big(-\frac{1}{2\sigma_1^2}\big)$ to conclude that the probability of this error case decays to zero as $n\to\infty$ whenever

$$\begin{aligned}R_0+R_1^{(1)}+R_2^{(0)} &< \frac{1-2\tau}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\sigma_1^2}\right)\\ &\quad+\frac{\tau}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\cdot 2^{2R_{\mathrm{h}_1}/\tau}\right)-\frac{\tau}{2}\end{aligned}\tag{83}$$

which holds for $\tau\downarrow 0$ whenever

$$R_0+R_1^{(1)}+R_2^{(0)}<\frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\sigma_1^2}\right)+R_{\mathrm{h}_1}. \tag{84}$$

- Case 3: $\big(\hat{m}_0,\hat{m}_2^{(0)}\big)=(1,1)$ $\hat{m}_1^{(0)}\neq 1$
In this case, too, it is immaterial whether or not $\hat{m}_1^{(1)}=1$. Indeed, when $\big(\hat{m}_0,\hat{m}_2^{(0)}\big)=(1,1)$, the guess $\hat{m}_1^{(0)}$ is determined by the first sum on the RHS of (59), which is unaffected by $m_1^{(1)}$. We therefore focus on this sum. When $\big(\hat{m}_0,\hat{m}_2^{(0)}\big)=(1,1)$, the sum reduces to

$$\frac{1}{n}\sum_{k=1}^{n(1-2\tau)}\big(v_{(1-2\tau),k}(1)+z_{1,k}-V_{(1-2\tau),k}\big)^2. \tag{85}$$

This sum is similar to the one encountered in the single-user additive Gaussian noise channel when the noise is of variance σ_1^2 and the codebook is Gaussian with power $\bar{\alpha}\mathsf{P}$. The probability of an error of the type considered in this case will thus vanish as $n\to\infty$ whenever

$$R_1^{(0)}<\frac{1-2\tau}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) \tag{86}$$

which, when $\tau\downarrow 0$, reduces to

$$R_1^{(0)}<\frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right). \tag{87}$$

- Case 4: $\left(\hat{m}_0, \hat{m}_2^{(0)}\right) = (1,1)$, $\hat{m}_1^{(1)} \neq 1$

  When $\left(\hat{m}_0, \hat{m}_2^{(0)}\right) = (1,1)$, the guess $\hat{m}_1^{(1)}$ is determined by the second sum on the RHS of (59), which is unaffected by $m_1^{(0)}$. We therefore focus on this sum, i.e., on

$$D_1' \triangleq \frac{1}{n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} 2^{2R_{\mathsf{h}_1}/\tau} \big(\hat{y}_{1,k} - X_k\big)^2. \quad (88)$$

By (53), the correct codeword will accumulate the metric $\tau\sigma_1^2$ on this segment with probability tending to one. For a wrong codeword, conditioned on $\hat{\mathbf{y}}_1$, the logarithmic MGF of D_1' is

$$\begin{aligned}\Lambda_n(\theta) = &\frac{1}{n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{1,k}^2 \theta}{1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}/n} \\ &- \frac{1}{2} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}/n\big). \quad (89)\end{aligned}$$

Therefore, by (63),

$$\begin{aligned}&\frac{1}{n}\Lambda_n(n\theta) \\ &= \frac{1}{n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{1,k}^2 \theta}{1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}} \\ &\quad - \frac{1}{2n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}\big) \quad (90) \\ &\overset{n\to\infty}{\to} \tau \frac{(2^{2R_{\mathsf{h}_1}/\tau}\mathsf{P} + \sigma_1^2)\theta}{1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}} - \frac{\tau}{2}\log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_1}/\tau}\, \mathsf{P}\big) \quad (91) \\ &\triangleq \Lambda(\theta) \quad (92)\end{aligned}$$

with probability one. By the Gärtner–Ellis Theorem, it follows as in Cases 1 and 2 that the exponent of the probability $\Pr(D_1' < \tau\sigma_1^2)$ is upper-bounded by $-\frac{\tau}{2}\log\big(1 + \frac{\mathsf{P}}{\sigma_1^2} \cdot 2^{2R_{\mathsf{h}_1}/\tau}\big)$. Therefore, by the union bound, the probability of this error case decays to zero as $n \to \infty$ whenever

$$R_1^{(1)} < \frac{\tau}{2}\log\left(1 + \frac{\mathsf{P}}{\sigma_1^2} \cdot 2^{2R_{\mathsf{h}_1}/\tau}\right). \quad (93)$$

When $\tau \downarrow 0$ it suffices that

$$R_1^{(1)} < R_{\mathsf{h}_1}. \quad (94)$$

We next study the performance of Decoder 2. Remember that we treat $\mathbf{v}_{(1-2\tau)}$ as (Gaussian) noise here. By (61), the true codeword will accumulate the metric $\sigma_2^2(1-\tau)$ with probability tending to one. Since we do not try to decode $\mathbf{v}_{(1-2\tau)}$, we must only consider two error events at this decoder.

- Case 1: $\left(\hat{m}_0, \hat{m}_2^{(0)}\right) \neq (1,1)$

  As in Case 1 in the analysis of Decoder 1, a candidate codeword $\mathbf{X}$ is independent of $\hat{\mathbf{y}}_2$ in this case, regardless of whether or not $\hat{m}_0^{(1)} = 1$. Conditioned on $\hat{\mathbf{y}}_2$, the logarithmic MGF of the metric D_2 is given for any $\theta < 0$ as

$$\begin{aligned}\Lambda_n(\theta) = &\frac{1}{n} \sum_{k=1}^{n(1-2\tau)} \frac{\hat{y}_{2,k}^2 \theta}{1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}/n} \\ &- \frac{1}{2} \sum_{k=1}^{n(1-2\tau)} \log\left(1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}/n\right) \\ &+ \frac{1}{n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{2,k}^2 \theta}{1 - 2\theta\, 2^{2R_{\mathsf{h}_2}/\tau}\, \mathsf{P}/n} \\ &- \frac{1}{2} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_2}/\tau}\, \mathsf{P}/n\big). \quad (95)\end{aligned}$$

Therefore, using (64) and (65),

$$\begin{aligned}\frac{1}{n}\Lambda_n(n\theta) = &\frac{1}{n} \sum_{k=1}^{n(1-2\tau)} \frac{\hat{y}_{2,k}^2 \theta}{1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}} \\ &- \frac{1}{2n} \sum_{k=1}^{n(1-2\tau)} \log\left(1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}\right) \\ &+ \frac{1}{n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \frac{\hat{y}_{2,k}^2 \theta}{1 - 2\theta 2^{2R_{\mathsf{h}_2}/\tau}\mathsf{P}} \\ &- \frac{1}{2n} \sum_{k=n(1-2\tau)+1}^{n(1-\tau)} \log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_2}/\tau}\, \mathsf{P}\big) \quad (96) \\ \overset{n\to\infty}{\to} &(1-2\tau) \frac{\frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\big(\mathsf{P}+\sigma_2^2\big)\theta}{1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}} \\ &- \frac{1-2\tau}{2}\log\left(1 - 2\theta \frac{\sigma_2^2}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\alpha\mathsf{P}\right) \\ &+ \tau \frac{(2^{2R_{\mathsf{h}_2}/\tau}\,\mathsf{P} + \sigma_2^2)\theta}{1 - 2\theta 2^{2R_{\mathsf{h}_2}/\tau}\mathsf{P}} \\ &- \frac{\tau}{2}\log\big(1 - 2\theta\, 2^{2R_{\mathsf{h}_2}/\tau}\, \mathsf{P}\big) \quad (97) \\ \triangleq &\Lambda(\theta) \quad (98)\end{aligned}$$

with probability one. Again, by the Gärtner–Ellis Theorem, the error exponent is given as

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\log\Pr\big(D_2 < \sigma_2^2\big) &= -\Lambda^*\big(\sigma_2^2\big) \quad (99) \\ &= -\sup_{\theta<0}\big\{\theta\sigma_2^2 - \Lambda(\theta)\big\}. \quad (100)\end{aligned}$$

We obtain a lower bound on this error exponent by evaluating the expression on the RHS for $\theta = -\frac{1}{2\sigma_2^2}$, which yields

$$\begin{aligned}&-\frac{\tau}{2} + \frac{1-2\tau}{2}\log\left(1 + \frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) \\ &\quad + \frac{\tau}{2}\log\left(1 + \frac{\mathsf{P}}{\sigma_2^2} \cdot 2^{2R_{\mathsf{h}_2}/\tau}\right). \quad (101)\end{aligned}$$

Consequently, this type of error will have vanishing probability as $n \to \infty$ whenever

$$\begin{aligned}R_0 + R_2^{(0)} + R_2^{(1)} < &-\frac{\tau}{2} + \frac{1-2\tau}{2}\log\left(1 + \frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) \\ &+ \frac{\tau}{2}\log\left(1 + \frac{\mathsf{P}}{\sigma_2^2} \cdot 2^{2R_{\mathsf{h}_2}/\tau}\right) \quad (102)\end{aligned}$$

or, upon letting $\tau \downarrow 0$, whenever

$$R_0 + R_2 < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) + R_{\mathrm{h}_2}, \quad (103)$$

where in the last step we recalled that $R_2^{(0)} + R_2^{(1)} = R_2$.

- Case 2: $\left(\hat{m}_0, \hat{m}_2^{(0)}\right) = (1,1)$, $\hat{m}_2^{(1)} \neq 1$
  This case is analogous to Case 4 in the analysis of Decoder 1. Upon letting $\tau \downarrow 0$, this type of error will vanish whenever

$$R_2^{(1)} < R_{\mathrm{h}_2}. \quad (104)$$

We summarize that, as $\tau \downarrow 0$, both decoders will decode the intended messages reliably whenever there exists some $0 < \alpha < 1$ such that

$$R_0 + R_1 + R_2^{(0)} < \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (105a)$$

$$R_0 + R_1^{(1)} + R_2^{(0)} < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (105b)$$

$$R_1^{(0)} < \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right), \quad (105c)$$

$$R_1^{(1)} < R_{\mathrm{h}_1}, \quad (105d)$$

$$R_0 + R_2 < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) + R_{\mathrm{h}_2}, \quad (105e)$$

$$R_2^{(1)} < R_{\mathrm{h}_2}. \quad (105f)$$

We next argue that (105b) is redundant. To this end, we first note that, except for (105c), the RHS of each of the inequalities in (105) is increasing in α. Consequently, when checking whether a rate tuple is achievable, it suffices to choose α to be so small that (105c) holds with (almost) equality and to then check whether the other conditions are satisfied. We now show that (105b) follows from this almost-with-equality form of (105c) and (105a). To this end, we rewrite (105a) as

$$\begin{aligned} &R_0 + R_1^{(0)} + R_1^{(1)} + R_2^{(0)} \\ &\quad < \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) + \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_1^2}\right) + R_{\mathrm{h}_1} \end{aligned} \quad (106)$$

and note that, when we subtract the almost-with-equality version of (105c) from this, we obtain

$$R_0 + R_1^{(1)} + R_2^{(0)} < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_1^2}\right) + R_{\mathrm{h}_1}. \quad (107)$$

Since this condition is more restrictive than (105b), the latter is redundant.

Once we have eliminated (105b), we can simplify the conditions in an additional way by replacing (105c) and (105d) with their sum, because, with rate-splitting, we can express R_1 as $R_1^{(0)} + R_1^{(1)}$ in whichever way we choose and the split does not influence any of the remaining inequalities.

Striking out the redundant condition (105b) and performing the above replacement, we obtain achievability whenever the following set of inequalities holds:

$$R_0 + R_1 + R_2^{(0)} < \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (108a)$$

$$R_1 < \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (108b)$$

$$R_0 + R_2 < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) + R_{\mathrm{h}_2}, \quad (108c)$$

$$R_2^{(1)} < R_{\mathrm{h}_2}. \quad (108d)$$

By writing $R_2^{(1)} = R_2 - R_2^{(0)}$ and applying the Fourier–Motzkin procedure we obtain the sufficiency of

$$R_0 + R_1 < \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (109a)$$

$$R_1 < \frac{1}{2}\log\left(1+\frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1}, \quad (109b)$$

$$R_0 + R_2 < \frac{1}{2}\log\left(1+\frac{\alpha\mathsf{P}}{\bar{\alpha}\mathsf{P}+\sigma_2^2}\right) + R_{\mathrm{h}_2}, \quad (109c)$$

$$R_0 + R_1 + R_2 < \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\sigma_1^2}\right) + R_{\mathrm{h}_1} + R_{\mathrm{h}_2}. \quad (109d)$$

The last condition is redundant: it can be obtained by replacing σ_2^2 in (109c) with σ_1^2 (recall (33)) and adding the result to (109b). This concludes the proof of achievability.

B. Converse

Fix an encoder, a helper, and two decoders, and consider the result of transmitting random messages M_0, M_1, M_2 that are drawn independently and uniformly over their corresponding support set. Using Fano's inequality and the fact that T_2 takes values in a set of cardinality of $2^{nR_{\mathrm{h}_2}}$ and therefore has entropy that is upper-bounded by nR_{h_2},

$$\begin{aligned} &R_0 + R_2 \\ &\quad = \frac{1}{n}I(M_0, M_2; \mathbf{Y_2}, T_2) + \frac{1}{n}H(M_0, M_2|\mathbf{Y_2}, T_2) && (110)\\ &\quad \leq \frac{1}{n}I(M_0, M_2; \mathbf{Y_2}, T_2) + \delta_n && (111)\\ &\quad = \frac{1}{n}I(M_0, M_2; \mathbf{Y_2}) + \frac{1}{n}I(M_0, M_2; T_2|\mathbf{Y_2}) + \delta_n && (112)\\ &\quad \leq \frac{1}{n}I(M_0, M_2; \mathbf{Y_2}) + R_{\mathrm{h}_2} + \delta_n, && (113)\end{aligned}$$

where $\lim_{n\to\infty}\delta_n = 0$. Similarly,

$$\begin{aligned} R_1 &= \frac{1}{n}I(M_1; \mathbf{Y_1}, T_1) + \frac{1}{n}H(M_1|\mathbf{Y_1}, T_1) && (114)\\ &\leq \frac{1}{n}I(M_1; \mathbf{Y_1}, T_1) + \delta_n && (115)\\ &= \frac{1}{n}I(M_1; \mathbf{Y_1}) + \frac{1}{n}I(M_1; T_1|\mathbf{Y_1}) + \delta_n && (116)\\ &\leq \frac{1}{n}I(M_1; \mathbf{Y_1}) + R_{\mathrm{h}_1} + \delta_n, && (117)\end{aligned}$$

and

$$\begin{aligned} &R_0 + R_1 \\ &\quad = \frac{1}{n}I(M_0, M_1; \mathbf{Y_1}, T_1) + \frac{1}{n}H(M_0, M_1|\mathbf{Y_1}, T_1) && (118)\\ &\quad \leq \frac{1}{n}I(M_0, M_1; \mathbf{Y_1}, T_1) + \delta_n && (119)\\ &\quad = \frac{1}{n}I(M_0, M_1; \mathbf{Y_1}) + \frac{1}{n}I(M_0, M_1; T_1|\mathbf{Y_1}) + \delta_n && (120)\\ &\quad \leq \frac{1}{n}I(M_0, M_1; \mathbf{Y_1}) + R_{\mathrm{h}_1} + \delta_n. && (121)\end{aligned}$$

We now have upper bounds on the achievable rates in terms of the mutual informations $I(M_0, M_2; \mathbf{Y_2})$, $I(M_1; \mathbf{Y_1})$, and $I(M_0, M_1; \mathbf{Y_1})$. We next argue that, with the codebook being fixed, these mutual informations depend on the channel law $f_{\mathbf{Y_1},\mathbf{Y_2}|\mathbf{X}}$ only via its marginals $f_{\mathbf{Y_1}|\mathbf{X}}$, $f_{\mathbf{Y_2}|\mathbf{X}}$ and that, consequently, the correlation between $\mathbf{Z_1}$ and $\mathbf{Z_2}$ is immaterial.

Consider first

$$I(M_0, M_2; \mathbf{Y_2}) = H(\mathbf{Y_2}) - H(\mathbf{Y_2}|M_0, M_2). \quad (122)$$

Since the distribution of (M_0, M_2) is fixed (uniform), both quantities on the RHS are determined by the conditional density $f_{\mathbf{Y_2}|M_0,M_2}$. The latter is determined by $f_{\mathbf{Y_2}|\mathbf{X}}$ because, for a fixed codebook $\mathcal{C}$,

$$\begin{aligned} f_{\mathbf{Y_2}|M_0,M_2}(\mathbf{y_2}|m_0, m_2) &= \sum_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{Y_2}|\mathbf{X},M_0,M_2}(\mathbf{y_2}|\mathbf{x}, m_0, m_2) \cdot p_{\mathbf{X}|M_0,M_2}(\mathbf{x}|m_0, m_2) \quad (123)\\ &= \sum_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{Y_2}|\mathbf{X}}(\mathbf{y_2}|\mathbf{x}) \cdot p_{\mathbf{X}|M_0,M_2}(\mathbf{x}|m_0, m_2), \quad (124) \end{aligned}$$

where the second equality holds because $(M_0, M_2) \multimap \mathbf{X} \multimap \mathbf{Y_2}$ forms a Markov chain. Since $p_{\mathbf{X}|M_0,M_2}$ is determined by the encoding rule (which is fixed), the mutual information $I(M_0, M_2; \mathbf{Y_2})$ is determined by the marginal conditional density $f_{\mathbf{Y_2}|\mathbf{X}}$.

An analogous argument shows that $I(M_1; \mathbf{Y_1})$ and $I(M_0, M_1; \mathbf{Y_1})$ are determined by $f_{\mathbf{Y_1}|\mathbf{X}}$. The correlation between $\mathbf{Z_1}$ and $\mathbf{Z_2}$ is thus immaterial and we may assume without loss of generality that—as in the converse of the standard Gaussian BC—the channel is physically degraded. It then follows from said converse that [18, Thm. 5.3]

$$\frac{1}{n} I(M_0, M_2; \mathbf{Y_2}) \leq \frac{1}{2} \log\left(1 + \frac{\alpha \mathsf{P}}{\bar{\alpha}\mathsf{P} + \sigma_2^2}\right), \quad (125a)$$

$$\frac{1}{n} I(M_1; \mathbf{Y_1}) \leq \frac{1}{2} \log\left(1 + \frac{\bar{\alpha}\mathsf{P}}{\sigma_1^2}\right) \quad (125b)$$

for some $\alpha \in [0, 1]$. Furthermore,

$$\begin{aligned} \frac{1}{n} I(M_0, M_1; \mathbf{Y_1}) &\leq \frac{1}{n} I(M_0, M_1, M_2; \mathbf{Y_1}) \quad (126)\\ &= \frac{1}{n} I(\mathbf{X}; \mathbf{Y_1}) \quad (127)\\ &\leq \frac{1}{2} \log\left(1 + \frac{\mathsf{P}}{\sigma_1^2}\right). \quad (128) \end{aligned}$$

These bounds together with (113), (117), and (121) establish the converse upon taking n to infinity. ■

References

[1] S. I. Bross and A. Lapidoth, "The additive noise channel with a helper," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Aug. 2019, pp. 1–5.

[2] T. M. Cover and Y.-H. Kim, "Capacity of a class of deterministic relay channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Nice, France, Jun. 2007, pp. 591–595.

[3] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.

[4] A. Lapidoth and G. Marti, "Encoder-assisted communications over additive noise channels," submitted for publication, 2020.

[5] R. Ahlswede and T. Han, "On source coding with side information via a multiple-access channel and related problems in multi-user information theory," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 396–412, May 1983.

[6] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, Sep. 1983.

[7] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, 2008.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[10] G. Marti, "Channels with a helper," M.S. thesis, Signal Inf. Process. Lab., ETH Zürich, Zurich, Switzerland, Sep. 2019.

[11] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 623–629, Jul. 1984.

[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[13] A. D. Wyner, "Random packings and coverings of the unit n-Sphere," *Bell Syst. Tech. J.*, vol. 46, no. 9, pp. 2111–2118, Nov. 1967.

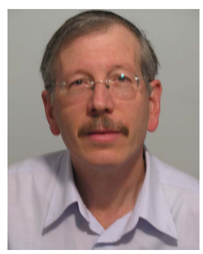
[14] S. I. Bross, A. Lapidoth, and M. Wigger, "Dirty-paper coding for the Gaussian multiaccess channel with conferencing," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5640–5668, Sep. 2012.

[15] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[16] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[17] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 1998.

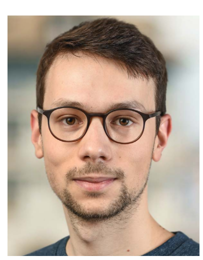
[18] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, Dec. 2011.



Shraga I. Bross (Senior Member, IEEE) 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, 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, in the capacity of 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 digital communications and information theory.

Amos Lapidoth (Fellow, IEEE) 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). He was the KDD Career Development Associate Professor in communications and technology. He is currently a Professor of information theory with the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He is also the author of the textbook *A Foundation in Digital Communication*, (Second Edition: Cambridge University Press, 2017). His research interests are digital communications and information theory.



Gian Marti (Student Member, IEEE) received the B.Sc. degree and the M.Sc. degree (with distinction) in electrical engineering from ETH Zurich in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Signal and Information Processing Laboratory (ISI), ETH Zurich. His research interests are information theory, signal processing, and machine learning.