Encoder-Assisted Communications Over Additive Noise Channels

Amos Lapidoth, *Fellow, IEEE,* and Gian Marti, *Student Member, IEEE*

***Abstract*— A coding technique that is based on flash helping is proposed for communicating over additive noise channels where a helper observes the noise and can describe it to the encoder over a noise-free rate-limited bit pipe. The technique is applicable irrespective of whether the helper observes the noise causally or noncausally. On the single-user channel of general noise, the rate it achieves is the sum of the channel's capacity without a helper and the rate of the bit pipe. For Gaussian noise and under an average-power constraint, it is optimal. Analogous results are derived for the additive noise multiple-access channel and the single-user Exponential channel. The approach is applicable also in some (noncausal) discrete settings, as demonstrated on the discrete modulo-additive noise channel.**

***Index Terms*— Additive noise, exponential channel, gaussian noise, helper, modulo-additive noise, multiple-access channel.**

I. Introduction

FLASH helping has recently been proposed as a capacity-achieving technique that allows a helper observing the noise to describe it over a rate-limited bit-pipe to the receiver [1]. This technique was extended in [2] to some multi-terminal scenarios, but with the help still being provided to the receiver(s). Here we consider help to the transmitter. Since the help now affects the transmitted signal, care must be exercised in accounting for its effect on the transmitted power.

Causality also becomes an important issue. We distinguish between two cases: the *noncausal case*, where the transmitter obtains the helper's description of the entire noise sequence before it begins to transmit, and the *causal case* where the transmitter's time-k symbol is only allowed to depend on help related to the noise sequence up to time-k. As we shall see, flash helping is applicable to both cases. In fact, in all but one of the scenarios we consider, the two cases lead to identical capacities.[1] In those scenarios in which causality is immaterial, we prove achievability for the causal case and the converse for the noncausal case.

The converse for encoder assistance is typically trickier than for decoder assistance. The latter is often proved using the Cut-Set bound [3, Theorem 15.10.1], which shows that rate-R_{h} decoder assistance cannot increase capacity by more than R_{h}. For encoder assistance no such cut-set bound exists. As the example in Appendix A shows, on a general state-dependent discrete memoryless channels (whose output is not necessarily the sum of the input and state), a rate-R_{h} description of the state sequence to the encoder could potentially increase capacity by more than R_{h}. In fact, even on a two-state channel, a rate-1 helper could potentially increase capacity by an arbitrarily large amount (provided that the input and output alphabets are sufficiently large).

To introduce the technique, we first consider the single-user additive Gaussian noise channel under an average-power constraint. The helper observes the noise and provides a rate-R_{h} description of it to the encoder. The symbols produced by the encoder thus depend not only on the transmitted message but also on the helper's description of the noise.

In the noncausal case, the description of the noise is provided to the encoder before transmission begins. Such a scenario may arise if the encoder is located next to an interfering transmitter. The interference, comprising the interferer's codeword, can be viewed as noise, which is known to the interfering transmitter. If a rate-limited channel exists between the two, then the interferer—which is cognizant of the codeword it is about to transmit—can use it to describe the interference to the encoder noncausally.

In the causal case the time-k symbol produced by the encoder can only depend on the message and the help it has received until time-k, with the latter depending only on the noise samples up to that time. Such a scenario may arise in our example if the two transmitters are not synchronized. In this case the interfering transmitter—which may only be cognizant of the message it is currently transmitting and not of the one succeeding it—may only be able to describe the remaining symbols in its frame (block). In the extreme case where its frame is just ending, it can only describe the present noise symbol.

The key idea behind flash helping is to satisfy the rate constraint on the bit pipe by providing the help with great precision but infrequently. To see why this can outperform schemes that provide help with moderate precision continuously, consider the single-user Gaussian noise channel with noise variance $\mathsf{N} > 0$, maximal-allowed average power $\mathsf{P} > 0$, and helper rate $R_{\mathrm{h}} > 0$ bits/channel-use. (Throughout this paper all logarithms are to base two, and all rates are in bits per channel-use.) Further assume noncausal helping. The moderate-but-steady approach would describe the n-length noise sequence using nR_{h} bits and thus result in per-noise-symbol mean squared-error (MSE) $\mathsf{N}2^{-2R_{\mathrm{h}}}$ (assuming an ideal Gaussian rate distortion codebook [3]). The estimate, which is known to the encoder prior to transmission, could be viewed as

Manuscript received October 7, 2019; revised March 25, 2020; accepted July 4, 2020. Date of publication July 29, 2020; date of current version October 21, 2020. *(Corresponding author: Amos Lapidoth.)*

The authors are with ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch; marti@isi.ee.ethz.ch).

Communicated by M. Costa, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2020.3012629

[1]The exception is the discrete modulo-additive noise channel.

"dirt" in Costa's writing-on-dirty-paper setting. Using Costa's coding technique, this dirt could be effectively subtracted off without any power penalty [4], and the remaining effective noise would thus comprise the estimation error. This moderate-but-steady approach thus leads to an achievable rate of

$$\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}2^{-2R_{\text{h}}}}\right). \tag{1}$$

With flash helping we can do better. The idea is to describe the different noise samples with different rates: one could describe the k-th noise symbol using $r_k \geq 0$ bits with corresponding MSE $\mathsf{N}2^{-2r_k}$, as long as the total description rate averaged over the block length n satisfies

$$\frac{1}{n}\sum_{k=1}^{n} r_k \leq R_{\text{h}}, \qquad r_k \geq 0. \tag{2}$$

By allocating the k-th symbol the power $\mathsf{P}_k \geq 0$ with

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{P}_k \leq \mathsf{P}, \qquad \mathsf{P}_k \geq 0 \tag{3}$$

one could obtain the average rate

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}2^{-2r_k}}\right). \tag{4}$$

The next proposition addresses the maximization of (4) subject to (2) and (3). It shows that, as n tends to infinity, the maximum is not achieved by choosing r_k and P_k constant (leading to (1)) but by flash helping, where r_k is zero for all k's other than some ℓ for which it equals nR_{h}. Thus, $r_k = nR_{\text{h}} \cdot \mathbf{1}\{k=\ell\}$, where ℓ is in $[1:n]$ (the set $\{1,\dots,n\}$), and $\mathbf{1}\{\text{statement}\}$ equals 1 if the statement is true and 0 otherwise.

Proposition 1 (The Flash-Helping Inequality): Let N, P, and R_{h} be positive and $\{r_k\}$ and $\{\mathsf{P}_k\}$ satisfy (2) and (3). Then

$$\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}2^{-2\,r_k}}\right) \leq \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\text{h}}. \tag{5}$$

Equality is achieved as n tends to infinity if $\mathsf{P}_k \equiv \mathsf{P}$ and $r_k \equiv nR_{\text{h}} \cdot \mathbf{1}\{k=\ell\}$ for some $\ell \in [1:n]$.

Proof: See Appendix B. ■

The rest of the paper is organized as follows. Section II treats the single-user additive noise channel and Section III its multiple-access counterpart. Section IV treats the Exponential channel. In doing so, it demonstrates how to assist the encoder when the input alphabet is restricted to the nonnegative reals, a case that also occurs, e.g., in the free-space optical channel [5]–[9], where the noise is Gaussian but the input nonnegative. Finally, in Section V it is shown that a discrete-alphabet variant of noncausal flash helping is capacity achieving on the modulo-additive noise channel.

II. The Single-User Channel

Consider the channel depicted in Figure 1, whose time-k output Y_k is

$$Y_k = x_k + Z_k, \tag{6}$$

where $x_k \in \mathbb{R}$ is its time-k input, and the noise samples $\{Z_k\}$ are IID $\sim \mathcal{N}(0,\mathsf{N})$, i.e., independent and identically distributed centered Gaussians of variance $\mathsf{N} > 0$.

A rate-R message set $\mathcal{M}$ for a blocklength-n transmission is a set with 2^{nR} elements. For concreteness we assume that $\mathcal{M} = \{1,\dots,2^{nR}\}$. Since the decoder receives no help, it guesses the message based on the output sequence $\mathbf{y}$ alone. It is thus a mapping $\psi_{\text{dec}}\colon \mathbb{R}^n \to \mathcal{M}$ that maps $\mathbf{y} \in \mathbb{R}^n$ to the decoder's guess $\hat{m}$.

The operation of the encoder and the helper depends on whether the help is provided noncausally or causally. A noncausal helper observes the entire noise sequence before describing it to the encoder. Only after obtaining this description does the encoder begin to transmit. More formally, a *noncausal* blocklength-n helper-encoder pair $\phi_{\text{nc-help}}$, $\phi_{\text{nc-enc}}$ can be described as follows: The helper is a mapping $\phi_{\text{nc-help}}\colon \mathbb{R}^n \to \mathcal{T}$, where $\mathcal{T}$ is a set of size $2^{nR_{\text{h}}}$, which, for concreteness, is assumed to be the set $\{1,\dots,2^{nR_{\text{h}}}\}$. We refer to the result of applying $\phi_{\text{nc-help}}$ to the noise sequence Z^n as the latter's description T. (Here and throughout we use A^k to denote $(A_1,\dots,A_k)$, and we use A^n and $\mathbf{A}$ interchangeably.) The noncausal encoder $\phi_{\text{nc-enc}}\colon \mathcal{M}\times\mathcal{T} \to \mathbb{R}^n$ is presented with the message m to be transmitted and with the description t of the noise sequence z^n. It then produces the length-n sequence $\mathbf{x}(m,t)$, which for every $m \in \mathcal{M}$ must satisfy the average power constraint $\mathbb{E}\left[\|\mathbf{x}(m,T)\|^2\right] \leq n\mathsf{P}$, where $\|\cdot\|$ denotes the Euclidean norm, and

$$\mathsf{P} > 0 \tag{7}$$

is the maximal-allowed average power, which is assumed throughout to be strictly positive. (Otherwise the capacity is, of course, zero.)

Unlike the noncausal helper, a causal helper cannot see the entire noise sequence before describing it. It provides the description piece by piece, with the piece provided at time-k being a function of the noise sequence only up to time k. The encoder, for its part, cannot wait for all the pieces before commencing with transmission: the symbol it sends at time k can only depend on the message and the pieces it received by that time. More formally, a *causal* helper describes the noise sequence z^n by an n-tuple $(t_1,\dots,t_n)$, where t_k takes values in the set $\mathcal{T}_k$ and can depend only on the noise samples z^k through time k. A blocklength-n causal helper is thus described by n functions $\{\phi^{(k)}_{\text{c-help}}\}_{k=1}^{n}$, where $\phi^{(k)}_{\text{c-help}}\colon \mathbb{R}^k \to \mathcal{T}_k$ maps z^k to $t_k \in \mathcal{T}_k$. To guarantee that the overall description length does not exceed nR_{h} bits, we impose the cardinality bound

$$\left|\mathcal{T}_1\times\cdots\times\mathcal{T}_n\right| \leq 2^{nR_{\text{h}}}. \tag{8}$$

(The converse we shall present goes through also under the weaker constraint that the entropy $H(T_1,\dots,T_n)$ not exceed nR_{h}.) The time-k channel input $x_k(m,t_1,\dots,t_k)$ produced by the encoder is determined by the message m and by the descriptions $t_1,\dots,t_k$ received by time k. A blocklength-n causal encoder is thus described using n mappings $\{\phi^{(k)}_{\text{c-enc}}\}_{k=1}^{n}$ where

$$\phi^{(k)}_{\text{c-enc}}\colon \mathcal{M}\times\mathcal{T}_1\times\cdots\times\mathcal{T}_k \to \mathbb{R} \tag{9}$$

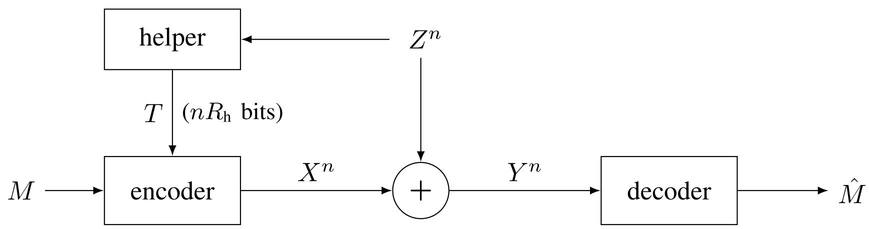

Fig. 1. The encoder-assisted single-user additive noise channel.

and we require that

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[x_k\bigl(m,t_1(Z^1),t_2(Z^2),\dots,t_k(Z^k)\bigr)^2\right]\le \mathsf{P},\qquad \forall m\in\mathcal{M}. \tag{10}$$

(The converse we shall present goes through also under the weaker constraint that the average over $m\in\mathcal{M}$ of the left-hand side of (10) does not exceed P.)

The supremum of all rates that, with the message m chosen uniformly at random from $\mathcal{M}$, allow for arbitrarily small probability of error is the capacity $C(R_{\mathrm{h}})$.

Imposing a causality constraint cannot, of course, increase capacity. Indeed, given a causal helper-encoder pair, we can define a noncausal pair of equal performance by considering the noncausal helper with $\mathcal{T}$ being $\mathcal{T}_1\times\cdots\times\mathcal{T}_n$ and with $T(Z^n)$ being $\bigl(T_1(Z^1),T_2(Z^2),\dots,T_n(Z^n)\bigr)$ and by considering the noncausal encoder whose time-k output corresponding to m and T is $x_k\bigl(m,t_1(Z^1),t_2(Z^2),\dots,t_k(Z^k)\bigr)$.

Theorem 2: The capacity of the average-power constrained additive Gaussian noise channel with noncausal or causal help is

$$C(R_{\mathrm{h}})=\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\mathrm{h}} \tag{11}$$

$$=C(0)+R_{\mathrm{h}}. \tag{12}$$

We shall prove achievability assuming causal help and the converse assuming noncausal help. The achievability part of the proof relies on Bennett's [10] classical result on high-resolution scalar quantization, which we quote from [11, Theorem 6.2]:

Theorem 3 (High-Resolution Scalar Quantization): Let the random variable Z satisfy $\mathbb{E}\left[Z^{2+\delta}\right]<\infty$ for some $\delta>0$ and have a density $f_Z(\cdot)$ satisfying[2]

$$\|f_Z\|_{1/3}\triangleq\left(\int_{\mathbb{R}}f_Z(z)^{1/3}\,\mathrm{d}z\right)^3<\infty. \tag{13}$$

For every positive integer L there then exists an L-level scalar quantizer $\hat{Z}_L$ for Z such that

$$\lim_{L\to\infty}L^2\cdot\mathbb{E}\bigl[(Z-\hat{Z}_L)^2\bigr]=\frac{1}{12}\|f_Z\|_{1/3}. \tag{14}$$

Proof of Theorem 2:

Achievability: We consider time-sharing between two schemes: the "no-help" scheme and the "with-help" scheme. The former is used $(1-\tau)$ of the time without help and the latter τ of the time with the help of a $\lfloor 2^{R_{\mathrm{h}}/\tau}\rfloor$-level scalar quantizer of the noise. Here $0<\tau<1$ is arbitrary but will later approach zero from above after we let the blocklength tend to infinity.[3] The data rate in the no-help scheme can be arbitrarily close to $C(0)$ while using an average transmission power not exceeding P. Its contribution to the overall achievable rate is thus $(1-\tau)C(0)$ and will, when we later let $\tau\downarrow 0$, converge to $C(0)$.

Consider now the with-help scheme. Let f_Z denote the density of the variance-N Gaussian distribution. Since it is bounded and decays exponentially, it has a finite order-$1/3$ norm. In fact, its order-$1/3$ quasinorm $\|f_Z\|_{1/3}$ equals $3^{3/2}\cdot 2\pi\mathsf{N}$.

The helper describes the k-th noise sample Z_k using a MSE-minimizing L-level quantizer, where

$$L=\lfloor 2^{R_{\mathrm{h}}/\tau}\rfloor \tag{15}$$

and the sample Z_k is reconstructed from its description as $\hat{Z}_{L,k}$, where $\hat{Z}_{L,k}$ is the conditional expectation of Z_k given its description, so

$$\mathbb{E}\bigl[\hat{Z}_{L,k}^2\bigr]\le\mathsf{N} \tag{16}$$

and the quantization error $\tilde{Z}_k=Z_k-\hat{Z}_{L,k}$ behaves asymptotically as in (14). The encoder uses a codebook whose codewords $\{\mathbf{x}(m)\}$ are drawn independently and uniformly over the n-dimensional sphere of radius $\sqrt{n\mathsf{P}}$, where $\mathbf{x}(m)$ denotes the m-th codeword and $x_k(m)$ its k-th component. To transmit the message m, it produces at time k the channel input $x_k(m)-\hat{Z}_{L,k}$. This, by (16), requires power at most $\mathsf{P}+\mathsf{N}$. The receiver observes the sum of this input and Z_k, i.e., $x_k(m)+\tilde{Z}_k$. Using nearest-neighbor decoding, rates arbitrarily close to

$$\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathbb{E}\bigl[(Z-\hat{Z})^2\bigr]}\right) \tag{17}$$

can be transmitted reliably [12]. Here and for the rest of the achievability proof we write Z for Z_k and $\hat{Z}$ for $\hat{Z}_{L,k}$. The achievable rate with time sharing is thus

$$(1-\tau)\,C(0)+\tau\,\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathbb{E}\bigl[(Z-\hat{Z})^2\bigr]}\right) \tag{18}$$

[2]In fact, the noise distribution need not have a density. It suffices that in its Lebesgue decomposition the part that is absolutely continuous with respect to the Lebesgue measure have a density of finite order-$1/3$ norm.

[3]We want to allow coding also over the symbols that are transmitted during the "with-help" phase. For the random coding argument to apply, the "with-help" phase must operate in the large blocklength regime, i.e. $n\tau$ must tend to infinity. The order of limits is thus crucial: we fix $\tau>0$ and let n tend to infinity (in order to be in the large blocklength regime) and only then let τ approach zero.

with power

$$(1-\tau)\mathsf{P}+\tau(\mathsf{P}+\mathsf{N}). \tag{19}$$

(The excess power $\tau\mathsf{N}$ can be eliminated by using power $\mathsf{P}-\tau\mathsf{N}/(1-\tau)$ in the no-help phase. This will reduce the achievable rate in the no-help phase by an amount that vanishes as $\tau\downarrow 0$ because, being concave, the capacity $C(0)$ is continuous in $\mathsf{P}>0$.)

We next conclude the achievability part of the proof by showing that (18) tends to $C(0)+R_{\mathrm{h}}$ as $\tau\downarrow 0$. Since the first term in (18) converges to $C(0)$, we need to show that the second is asymptotically bounded from below by R_{h}. This we do using (15) and (14):

$$\begin{aligned}
&\frac{\tau}{2}\log\left(1+\frac{\mathsf{P}}{\mathbb{E}\big[(Z-\hat{Z})^2\big]}\right)\\
&= \frac{\tau}{2}\log\left(1+\frac{\mathsf{P}}{L^{-2}\big(\frac{1}{12}\|f_Z\|_{1/3}+o_L(1)\big)}\right) &(20)\\
&= \frac{\tau}{2}\log\left(1+\frac{\big(\lfloor 2^{R_{\mathrm{h}}/\tau}\rfloor\big)^2\mathsf{P}}{\frac{1}{12}\|f_Z\|_{1/3}+o_L(1)}\right) &(21)\\
&\geq \frac{\tau}{2}\log\left(1+\frac{(2^{R_{\mathrm{h}}/\tau}-1)^2\mathsf{P}}{\frac{1}{12}\|f_Z\|_{1/3}+o_L(1)}\right) &(22)\\
&= \frac{\tau}{2}\log\left(2^{2R_{\mathrm{h}}/\tau}\left(2^{-2R_{\mathrm{h}}/\tau}+\frac{(1-2^{-R_{\mathrm{h}}/\tau})^2\mathsf{P}}{\frac{1}{12}\|f_Z\|_{1/3}+o_L(1)}\right)\right) &(23)\\
&= R_{\mathrm{h}}+\frac{\tau}{2}\log\left(2^{-2R_{\mathrm{h}}/\tau}+\frac{(1-2^{-R_{\mathrm{h}}/\tau})^2\mathsf{P}}{\frac{1}{12}\|f_Z\|_{1/3}+o_L(1)}\right) &(24)\\
&\overset{\tau\downarrow 0}{\to} R_{\mathrm{h}}, &(25)
\end{aligned}$$

where $o_L(1)$ stands for a term that tends to zero as $L\to\infty$ (which happens as a consequence of $\tau\downarrow 0$). Here (20) holds by (14); (21) follows from (15); and in the last step we used the fact that the argument of the logarithm in (24) converges to $\frac{\mathsf{P}}{\frac{1}{12}\|f_Z\|_{1/3}}$ as $\tau\downarrow 0$. This concludes the proof of achievability.

Converse: The converse is very similar to one that appears in [13] and can, in fact, be deduced from [13] by substituting zero for σ_Z^2 there. Nevertheless, we include the argument for completeness and because the proofs of the other converses in this paper build on it.

Consider a sequence of noncausal helper-encoder pairs and corresponding decoders for which the probability of error tends to zero as the blocklength tends to infinity. For each blocklength, apply the helper-encoder pair to a message M that is drawn uniformly over $\mathcal{M}$, so $H(M)=nR$ and hence,

$$\begin{aligned}
nR &= I(M;\mathbf{Y},T)+H(M|\mathbf{Y},T) &(26)\\
&\leq I(M;\mathbf{Y},T)+H(M|\mathbf{Y}) &(27)\\
&\leq I(M;\mathbf{Y},T)+n\delta_n &(28)\\
&= I(M;\mathbf{Y}|T)+n\delta_n &(29)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Y}|T,M)+n\delta_n &(30)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Y}|T,M,\mathbf{X})+n\delta_n &(31)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Y}|T,\mathbf{X})+n\delta_n &(32)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Z}|T)+n\delta_n &(33)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Z})+I(\mathbf{Z};T)+n\delta_n &(34)\\
&\leq h(\mathbf{Y}|T)-h(\mathbf{Z})+H(T)+n\delta_n &(35)\\
&\leq h(\mathbf{Y}|T)-\frac{n}{2}\log(2\pi e\mathsf{N})+nR_{\mathrm{h}}+n\delta_n, &(36)
\end{aligned}$$

where (28) holds for some $\delta_n\overset{n\to\infty}{\to}0$ by Fano's inequality and by assuming that the probability of error vanishes as $n\to\infty$; (29) holds because T is independent of M; and (33) holds because $\mathbf{Y}=\mathbf{X}+\mathbf{Z}$ where $(\mathbf{X},M)\multimap T\multimap\mathbf{Z}$ forms a Markov chain. We now bound $h(\mathbf{Y}|T)$ in terms of the probability mass function (PMF) $p_T(\cdot)$ of T. Key is that, conditional on T, the random variables X_k and Z_k are independent. Using $\mathrm{Var}[X|Y]$ to denote the variance of X given Y (and hence being a random variable that is computable from Y),

$$\begin{aligned}
&h(\mathbf{Y}|T)\\
&\leq \sum_{k=1}^{n}h(Y_k|T) &(37)\\
&= \sum_{k=1}^{n}\sum_{t\in\mathcal{T}}p_T(t)\,h(Y_k|T=t) &(38)\\
&\leq \sum_{k=1}^{n}\sum_{t\in\mathcal{T}}p_T(t)\,\frac{1}{2}\log\Big(2\pi e\mathrm{Var}[Y_k|T=t]\Big) &(39)\\
&\leq \sum_{k=1}^{n}\frac{1}{2}\log\left(2\pi e\sum_{t\in\mathcal{T}}p_T(t)\mathrm{Var}[Y_k|T=t]\right) &(40)\\
&= \sum_{k=1}^{n}\frac{1}{2}\log\Big(2\pi e\,\mathbb{E}\big[\mathrm{Var}[Y_k|T]\big]\Big) &(41)\\
&= \sum_{k=1}^{n}\frac{1}{2}\log\Big(2\pi e\big(\mathbb{E}\big[\mathrm{Var}[X_k|T]\big]+\mathbb{E}\big[\mathrm{Var}[Z_k|T]\big]\big)\Big) &(42)\\
&\leq \sum_{k=1}^{n}\frac{1}{2}\log\Big(2\pi e\big(\mathrm{Var}[X_k]+\mathrm{Var}[Z_k]\big)\Big) &(43)\\
&= \frac{n}{2}\sum_{k=1}^{n}\frac{1}{n}\log\Big(2\pi e\big(\mathrm{Var}[X_k]+\mathsf{N}\big)\Big) &(44)\\
&\leq \frac{n}{2}\log\left(2\pi e\sum_{k=1}^{n}\frac{1}{n}\big(\mathrm{Var}[X_k]+\mathsf{N}\big)\right) &(45)\\
&\leq \frac{n}{2}\log\big(2\pi e\big(\mathsf{P}+\mathsf{N}\big)\big), &(46)
\end{aligned}$$

where (40) follows from Jensen's inequality and the concavity of the logarithm; (42) holds because $\mathbf{X}\multimap T\multimap\mathbf{Z}$ forms a Markov chain; (43) holds by the law of total variance (conditioning reduces variance) [14, p. 69]; (45) follows again from Jensen's inequality; and (46) follows from the power constraint on the channel input. Thus, returning to (36) and using (46),

$$\begin{aligned}
R &\leq \frac{1}{n}h(\mathbf{Y}|T)-\frac{1}{2}\log(2\pi e\mathsf{N})+R_{\mathrm{h}}+\delta_n &(47)\\
&\leq \frac{1}{2}\log\big(2\pi e\big(\mathsf{P}+\mathsf{N}\big)\big)-\frac{1}{2}\log(2\pi e\mathsf{N})+R_{\mathrm{h}}+\delta_n &(48)\\
&= \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\mathrm{h}}+\delta_n. &(49)
\end{aligned}$$

As n tends to infinity, δ_n approaches zero, and the right-hand side (RHS) converges to the RHS of (11). ∎

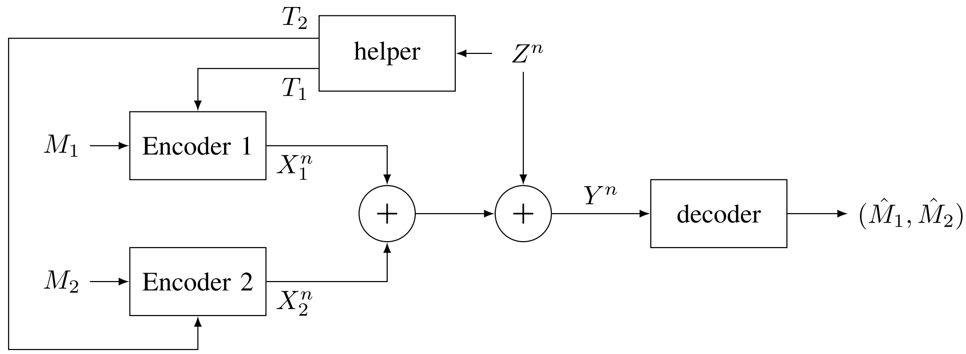

Fig. 2. The additive noise multiple-access channel with help to the encoders.

Inspecting our achievability proof, we note that it does not rely on the noise being Gaussian: it suffices that the quantization MSE of the optimal scalar quantizer have the proper high-resolution asymptotic behavior, namely,

$$\varlimsup_{L\to\infty} L^2 \cdot \mathbb{E}\big[(Z-\hat{Z})^2\big] < \infty. \tag{50}$$

Hence:

Remark 4 (Non-Gaussian Noise): Consider the additive non-Gaussian noise channel where the noise $\{Z_k\}$ is IID with a distribution satisfying the hypotheses of Theorem 3. Its capacity $C(R_{\mathrm{h}})$ with causal help is then lower-bounded by

$$C(R_{\mathrm{h}}) \geq C(0) + R_{\mathrm{h}}, \tag{51}$$

where $C(0)$ is the channel's capacity in the absence of help.

The tightness of (51) for non-Gaussian noise is discussed in [15]. It is shown there [15, Theorems 4.12 and 4.13] that for general noise,

$$C(R_{\mathrm{h}}) \leq I_{\mathrm{G}}(0) + R_{\mathrm{h}} + \min\left\{\frac{1}{2}, \frac{1}{2}\log\left(1+\frac{\mathsf{N}}{\mathsf{P}}\right)\right\}, \tag{52}$$

where $I_{\mathrm{G}}(0)$, which cannot exceed $C(0)$, is the mutual information in the absence of help between the channel input and output when the input is $\mathcal{N}(0, \mathsf{P})$. The difference between this bound and the RHS of (51) never exceeds half a bit, and tends to zero as P/N tends to infinity.

To prove the converse we needed to establish that the normalized conditional entropy $n^{-1}H(M|\mathbf{Y}, T)$ tends to zero; see (28). This, we showed, follows from Fano's inequality and the convergence to zero of the probability of error in guessing M based on $\mathbf{Y}$. But it also follows from Fano's inequality and the convergence to zero of the probability of error in guessing M based on $\mathbf{Y}$ *and* T. Hence, the converse would apply also if the decoder were cognizant not only of the channel output sequence $\mathbf{Y}$ but also of the noise's description T. Thus:

Remark 5: On the Gaussian channel, no rate exceeding the RHS of (11) is achievable even if the noise's description presented to the encoder is also presented to the decoder.

III. The Multiple-Access Channel

We next consider help to the encoders on a multiple-access channel (MAC). We focus on the additive noise MAC of Figure 2, whose time-k output Y_k is

$$Y_k = x_{1,k} + x_{2,k} + Z_k, \tag{53}$$

where $x_{1,k}$ and $x_{2,k}$ are the time-k channel inputs, and the noise $\{Z_k\}$ is IID. Depending on the scenario, the helper observes the noise causally or noncausally and provides its rate-R_{h_1} description T_1 to Encoder 1 and its rate-R_{h_2} description T_2 to Encoder 2. (In the noncausal case T_1 and T_2 are functions of Z^n; in the causal case they are n-tuples whose k-th component is a function of Z^k.) Based on the respective descriptions of the noise and on the respective messages, the encoders produce the inputs $\mathbf{X}_1(m_1, T_1)$ and $\mathbf{X}_2(m_2, T_2)$. (In the causal case the k-th component of $\mathbf{X}_1(m_1, T_1)$ must be a function of m_1 and the first k components of T_1 and likewise $\mathbf{X}_2(m_2, T_2)$.) We require that—irrespective of the transmitted messages (m_1, m_2)—the average power constraints

$$\mathbb{E}\big[\|\mathbf{X}_i(m_i, T_i)\|^2\big] \leq n\mathsf{P}_i, \quad i = 1, 2, \tag{54}$$

be satisfied, where

$$\mathsf{P}_1, \mathsf{P}_2 > 0 \tag{55}$$

are the maximal-allowed average powers for the two users. The total description rate is denoted R_{h},

$$R_{\mathrm{h}} = R_{\mathrm{h}_1} + R_{\mathrm{h}_2} \tag{56}$$

and the capacity region $\mathcal{C}(R_{\mathrm{h}_1}, R_{\mathrm{h}_2})$.

As in the single-user case, our achievability result in the following theorem holds for causal help and arbitrary noise distribution (satisfying the hypotheses of Theorem 3) and the converse for noncausal help but with Gaussian noise. In stating the result we use "+" to denote Minkowski set addition. Noteworthy is that, for Gaussian noise, the capacity region $\mathcal{C}(R_{\mathrm{h}_1}, R_{\mathrm{h}_2})$ depends on the rates R_{h_1} and R_{h_2} only via their sum R_{h}.

Theorem 6: If the noise satisfies the hypotheses of Theorem 3, then all rate pairs (R_1, R_2) in the set

$$\mathcal{C}(0,0) + \big\{(\rho_1, \rho_2) \in \mathbb{R}_+ \times \mathbb{R}_+ \colon \rho_1 + \rho_2 \leq R_{\mathrm{h}}\big\} \tag{57}$$

are achievable with causal help. If the noise is Gaussian, then no rate pair outside this set is achievable even with noncausal help; this set is then the capacity region.

Proof:

Achievability: We first prove achievability (for causal help). Again we consider a time-sharing scheme between two schemes: the "no-help" scheme, which is used $(1-\tau)$ of the time, and the "with-help" scheme, which is used τ of the time.

No help is provided in the "no-help" scheme, and its contribution to the overall rate region is thus $(1-\tau)\,\mathcal{C}(0,0)$ (with $\mathcal{C}(0,0)$ being the MAC's no-help capacity region for powers $\mathsf{P}_1, \mathsf{P}_2$). This will tend to $\mathcal{C}(0,0)$ when we later let τ approach zero.

In the "with-help" scheme we differentiate between the inputs at odd times and even times. At odd times, Encoder 1 receives help at rate $2R_{\mathrm{h}_1}/\tau$ and Encoder 2 receives none.[4] The help to Encoder 1 is in the form of a scalar quantization of the noise Z_{o} with

$$L_1 = 2^{2R_{\mathrm{h}_1}/\tau} \tag{58}$$

levels, with corresponding noise reconstruction $\hat{Z}_{\mathrm{o}}$, and corresponding estimation error $\tilde{Z}_{\mathrm{o}}$ $(= Z_{\mathrm{o}} - \hat{Z}_{\mathrm{o}})$ satisfying

$$\lim_{L_1\to\infty} L_1^2 \cdot \mathbb{E}\big[(Z_{\mathrm{o}} - \hat{Z}_{\mathrm{o}})^2\big] = \frac{1}{12}\|f_Z\|_{1/3}. \tag{59}$$

(As we did in the single-user case, we are dropping the time indices.) At these odd times, Encoder 1 subtracts its estimate $\hat{Z}_{\mathrm{o}}$ of the noise Z_{o} (with a power penalty that vanishes as $\tau \downarrow 0$).

At the even times the roles are reversed: the help is provided to Encoder 2, and it is Encoder 2 that subtracts its estimate of the noise from its input. The corresponding number of quantization levels and noise estimates are defined in an analogous way.

We now consider the "super-MAC" with Encoder 1 input $\tilde{X}_1 \triangleq (X_{1,\mathrm{o}}, X_{1,\mathrm{e}})$, Encoder 2 input $\tilde{X}_2 \triangleq (X_{2,\mathrm{o}}, X_{2,\mathrm{e}})$, and output $\tilde{Y} \triangleq (Y_{\mathrm{o}}, Y_{\mathrm{e}})$. The inputs $X_{1,\mathrm{o}}$, $X_{2,\mathrm{o}}$, $X_{1,\mathrm{e}}$, and $X_{2,\mathrm{e}}$ we consider are IID according to some distribution $\tilde{P}_X$ that satisfies both power constraints and that has finite differential entropy, e.g., $\mathcal{N}(0, \min\{\mathsf{P}_1, \mathsf{P}_2\})$. The terms defining the capacity of the super-MAC, namely, $I(\tilde{X}_1;\tilde{Y}|\tilde{X}_2)$, $I(\tilde{X}_2;\tilde{Y}|\tilde{X}_1)$, and $I(\tilde{X}_1,\tilde{X}_2;\tilde{Y})$ can now be lower bounded by lower-bounding the differential entropy of the output by that of the input and by upper-bounding the differential entropy of the equivalent noise (i.e., estimation error) by that of a Gaussian of equal second moment. For example,

$$\begin{aligned} &I(\tilde{X}_1;\tilde{Y}|\tilde{X}_2)\\ &= h(\tilde{Y}|\tilde{X}_2) - h(\tilde{Y}|\tilde{X}_1,\tilde{X}_2) && (60)\\ &\geq h(\tilde{X}_1) - h(\tilde{Z}_{\mathrm{o}}) - h(\tilde{Z}_{\mathrm{e}}) && (61)\\ &\geq 2\, h(\tilde{P}_X) - \tfrac{1}{2}\log\big(2\pi e\,\mathbb{E}\big[\tilde{Z}_{\mathrm{o}}^2\big]\big) - \tfrac{1}{2}\log\big(2\pi e\,\mathbb{E}\big[\tilde{Z}_{\mathrm{e}}^2\big]\big), && (62) \end{aligned}$$

with the other terms, namely $I(\tilde{X}_2;\tilde{Y}|\tilde{X}_1)$ and $I(\tilde{X}_1,\tilde{X}_2;\tilde{Y})$ treated similarly.

When we scale these lower bounds by one-half (to account for the fact that in using the super-MAC once we are using the original channel twice) and by τ (to account for the fact that we use the "with-help" scheme only τ of the time) we obtain—upon taking τ to zero and recalling (59) and (58)—the second term in the Minkowski sum on the RHS of (57).

[4]Describing the same noise sample Z_k at rate $\tau^{-1}R_{\mathrm{h}_1}$ to Encoder 1 and at rate $\tau^{-1}R_{\mathrm{h}_2}$ to Encoder 2 is suboptimal because it is not as beneficial as describing it to only one of them at rate $\tau^{-1}(R_{\mathrm{h}_1}+R_{\mathrm{h}_2})$. The presented scheme allows us to pool the resources. An equivalent approach would be to break up the "with-help" phase into two subphases: in one subphase the noise would be described only to Encoder 1 and in the other only to Encoder 2.

For example, we obtain from (62)

$$\begin{aligned} &\frac{\tau}{2} I(\tilde{X}_1;\tilde{Y}|\tilde{X}_2)\\ &\geq \tau h(\tilde{P}_X) - \frac{\tau}{4}\log\big(2\pi e\,\mathbb{E}\big[\tilde{Z}_{\mathrm{o}}^2\big]\big) - \frac{\tau}{4}\log\big(2\pi e\,\mathbb{E}\big[\tilde{Z}_{\mathrm{e}}^2\big]\big) && (63)\\ &= \tau\big(h(\tilde{P}_X) - \frac{1}{2}\log(2\pi e)\big) - \frac{\tau}{4}\log\big(\mathbb{E}\big[\tilde{Z}_{\mathrm{o}}^2\big]\big)\\ &\quad - \frac{\tau}{4}\log\big(\mathbb{E}\big[\tilde{Z}_{\mathrm{e}}^2\big]\big) && (64)\\ &= \tau\left(h(\tilde{P}_X) - \frac{1}{2}\log(2\pi e)\right)\\ &\quad - \frac{\tau}{4}\log\left(L_1^{-2}\left(\frac{1}{12}\|f_Z\|_{1/3} + o_L(1)\right)\right)\\ &\quad - \frac{\tau}{4}\log\left(L_2^{-2}\left(\frac{1}{12}\|f_Z\|_{1/3} + o_L(1)\right)\right) && (65)\\ &= \frac{\tau}{2}\log L_1 + \frac{\tau}{2}\log L_2\\ &\quad + \tau\left(h(\tilde{P}_X) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\|f_Z\|_{1/3} + o_L(1)\right)\right) && (66)\\ &\overset{\tau\downarrow 0}{\to} R_{\mathrm{h}_1} + R_{\mathrm{h}_2} && (67) \end{aligned}$$

with a similar asymptotic analysis for $(\tau/2)I(\tilde{X}_2;\tilde{Y}|\tilde{X}_1)$ and $(\tau/2)I(\tilde{X}_1,\tilde{X}_2;\tilde{Y})$.

Converse: We now turn to the converse for noncausal help on the Gaussian channel. Define $T \triangleq (T_1, T_2)$. Let M_1 and M_2 be drawn independently and uniformly over their corresponding support sets, and let $\mathbf{X}_1$ and $\mathbf{X}_2$ be their encoding.

$$\begin{aligned} nR_1 &= H(M_1) && (68)\\ &= I(M_1;\mathbf{Y},T) + H(M_1|\mathbf{Y},T) && (69)\\ &\leq I(M_1;\mathbf{Y},T) + H(M_1,M_2|\mathbf{Y},T) && (70)\\ &\leq I(M_1;\mathbf{Y},T,\mathbf{X}_2) + n\delta_n && (71)\\ &= I(M_1;\mathbf{Y}|T,\mathbf{X}_2) + n\delta_n && (72)\\ &= h(\mathbf{Y}|T,\mathbf{X}_2) - h(\mathbf{Y}|T,\mathbf{X}_2,\mathbf{X}_1,M_1) + n\delta_n && (73)\\ &= h(\mathbf{Y}|T,\mathbf{X}_2) - h(\mathbf{Z}|T,\mathbf{X}_2,\mathbf{X}_1,M_1) + n\delta_n && (74)\\ &= h(\mathbf{X}_1+\mathbf{Z}|T) - h(\mathbf{Z}|T) + n\delta_n && (75)\\ &\leq h(\mathbf{X}_1+\mathbf{Z}|T) - h(\mathbf{Z}) + nR_{\mathrm{h}} + n\delta_n && (76)\\ &\leq \frac{n}{2}\log(2\pi e(\mathsf{P}_1+\mathsf{N})) - \frac{n}{2}\log(2\pi e\mathsf{N}) + nR_{\mathrm{h}} + n\delta_n && (77)\\ &= n\left(\frac{1}{2}\log\left(1+\frac{\mathsf{P}_1}{\mathsf{N}}\right) + R_{\mathrm{h}} + \delta_n\right), && (78) \end{aligned}$$

where (71) holds for some $\delta_n \to 0$ by Fano's inequality and by introducing $\mathbf{X}_2$; (72) holds because M_1 is independent of $(T,\mathbf{X}_2)$; (73) holds because $\mathbf{X}_1$ is computable from (M_1,T); (74) holds because $\mathbf{Z} = \mathbf{Y} - (\mathbf{X}_1+\mathbf{X}_2)$; (75) holds because $\mathbf{Z} \multimap T \multimap (M_1,\mathbf{X}_1,\mathbf{X}_2)$ and $X_1 \multimap T \multimap X_2$ form Markov chains; and (77) can be derived along the lines leading to (46).

Analogously,

$$R_2 \leq \frac{1}{2}\log\left(1+\frac{\mathsf{P}_2}{\mathsf{N}}\right) + R_{\mathrm{h}} + \delta_n. \tag{79}$$

As to the sum of the rates,

$$
\begin{aligned}
&n(R_1+R_2)\\
&= I(M_1,M_2;\mathbf{Y},T)+H(M_1,M_2|\mathbf{Y},T) &(80)\\
&\le I(M_1,M_2;\mathbf{Y},T)+n\delta_n &(81)\\
&= I(M_1,M_2;\mathbf{Y}|T)+n\delta_n &(82)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Y}|T,M_1,M_2,\mathbf{X}_1,\mathbf{X}_2)+n\delta_n &(83)\\
&= h(\mathbf{Y}|T)-h(\mathbf{Z}|T)+n\delta_n &(84)\\
&\le h(\mathbf{X}_1+\mathbf{X}_2+\mathbf{Z}|T)-h(\mathbf{Z})+nR_{\mathrm{h}}+n\delta_n &(85)\\
&\le \frac{n}{2}\log(2\pi e(\mathsf{P}_1+\mathsf{P}_2+\mathsf{N}))-\frac{n}{2}\log(2\pi e\mathsf{N})+nR_{\mathrm{h}}+n\delta_n &(86)\\
&= n\left(\frac{1}{2}\log\left(1+\frac{\mathsf{P}_1+\mathsf{P}_2}{\mathsf{N}}\right)+R_{\mathrm{h}}+\delta_n\right), &(87)
\end{aligned}
$$

where (83) holds because $\mathbf{X}_1$ and $\mathbf{X}_2$ are computable from (T,M_1,M_2); (84) holds because $\mathbf{Z}=\mathbf{Y}-(\mathbf{X}_1+\mathbf{X}_2)$ and $\mathbf{Z} \multimap\!\!- T \multimap\!\!- (M_1,M_2,\mathbf{X}_1,\mathbf{X}_2)$; (85) can be derived as in (33)–(36); (86) holds because, conditional on T, the vectors $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{Z}$ are independent (so their covariance matrices add) and because the IID Gaussian maximizes differential entropy subject to a trace constraint on the covariance matrix (this is in direct analogy to (37)–(46)); and (87) follows from straightforward algebra.

Letting $n\to\infty$ and recalling the capacity region $\mathcal{C}(0,0)$ of the Gaussian MAC without help [16, Theorem 4.4]

$$
\mathcal{C}(0,0)=\Bigg\{(R_1,R_2)\in\mathbb{R}_+\times\mathbb{R}_+:\; R_1\le\frac{1}{2}\log\left(1+\frac{\mathsf{P}_1}{\mathsf{N}}\right),\; R_2\le\frac{1}{2}\log\left(1+\frac{\mathsf{P}_2}{\mathsf{N}}\right),\; R_1+R_2\le\frac{1}{2}\log\left(1+\frac{\mathsf{P}_1+\mathsf{P}_2}{\mathsf{N}}\right)\Bigg\} \qquad (88)
$$

concludes the proof. ■

IV. The Exponential Channel

The inputs to our next channel must be nonnegative. This constraint makes it tricky for the encoder to subtract its estimate of the noise, because this subtraction might lead to a negative input. Nevertheless, a minor modification of our technique can achieve capacity. The channel we study is the additive noise channel of (6), but with different constraints: The input x_k is *nonnegative* and we require that its time-averaged expectation, i.e., "average power" be bounded by the maximal-allowed average power P

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left[X_k\right]\le\mathsf{P}, \qquad (89)$$

where $\mathsf{P}>0$. The noise, which need not be nonnegative, is assumed to be of mean $\mathsf{N}\in\mathbb{R}$. This channel reduces to the Exponential channel [17], [18] when N is positive and Z_k is a mean-N Exponential, i.e., of density

$$f_Z(z)=\frac{1}{\mathsf{N}}\exp\left(-\frac{z}{\mathsf{N}}\right)\mathbf{1}\{z\ge 0\}. \qquad (90)$$

It reduces to the free-space optical channel [5]–[9] when Z_k is a centered Gaussian.

In the absence of help, the capacity of the Exponential channel is [17]

$$\frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right). \qquad (91)$$

We next show that, also for this channel, the availability of causal or noncausal help to the encoder increases capacity by the rate of the help.

Theorem 7: The capacity of the Exponential channel with help to the encoder is

$$
\begin{aligned}
C(R_{\mathrm{h}}) &= \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\mathrm{h}} &(92)\\
&= C(0)+R_{\mathrm{h}}, &(93)
\end{aligned}
$$

regardless of whether the help is provided to the encoder causally or noncausally and regardless of whether the help that is provided to the encoder is also provided to the decoder. Moreover, if the noise is not necessarily Exponential but satisfies the hypotheses of Theorem 3, then the RHS of (93) is achievable with causal help, provided we interpret $C(0)$ as the capacity of the channel of said noise without help.

Proof:

Achievability: We prove the achievability of the RHS of (93) for general noise and causal help and the converse for Exponential noise with noncausal help to the encoder that is also provided to the decoder.

Again we use time-sharing between a "no-help" scheme, where the channel is used for $(1-\tau)$ of the time without help, and the "with-help" scheme where the channel is used for τ of the time with help at rate R_{h}/τ. In the no-help scheme, the channel is used with average power P and rate $C(0)$ and therefore contributes to the overall rate $(1-\tau)C(0)$. We will again let $\tau\downarrow 0$ and thus drive this contribution to $C(0)$. In the with-help scheme, which we describe next, we add to the channel input a constant A that (most of the time) allows the subtraction of the encoder's noise estimate. The constant can be subtracted by the receiver and therefore does not impair communication.

Fix some $\mathsf{A}>0$. (After letting $\tau\downarrow 0$ we shall let A tend to infinity.) Define

$$F_k=\mathbf{1}\big\{|Z_k|\le\mathsf{A}\big\}, \qquad (94)$$

and note that the hypotheses of Theorem 3 guarantee that

$$\lim_{\mathsf{A}\to\infty}\Pr(F_k=1)=1 \qquad (95)$$

and hence

$$\lim_{\mathsf{A}\to\infty}H(F_k)=0. \qquad (96)$$

Let the time-k transmitted symbol be

$$x_k(m)+\mathsf{A}-\hat{Z}_k, \qquad (97)$$

where $\hat{Z}_k$ (to be defined shortly) is an estimate of the time-k noise sample, and where $\{\mathbf{x}(m)\}$ are codewords that are drawn independently, each with IID $\sim P_X$ components, where P_X is an input distribution that satisfies the power and nonnegativity constraints and has finite differential entropy.

Let $\tilde{f}$ be the conditional density of Z_k given $F_k = 1$, i.e., given $|Z_k| \leq \mathsf{A}$. Note that if f_Z satisfies the hypotheses of Theorem 3, then so does $\tilde{f}$. The helper uses a MSE-minimizing L-level scalar quantizer for $\tilde{f}$, where

$$L = 2^{R_\mathrm{h}/\tau} \tag{98}$$

and $\hat{Z}_k$ denotes the conditional expectation under $\tilde{f}$ of Z_k given its description. Since the support of $\tilde{f}$ is contained in the interval $[-\mathsf{A}, \mathsf{A}]$,

$$0 \leq \mathsf{A} - \hat{Z}_k \leq 2\mathsf{A}. \tag{99}$$

It follows from (97) and (99) that the transmitted power in the with-help scheme is upper bounded by $\mathsf{P} + 2\mathsf{A}$. The overall power is thus upper-bounded by $(1-\tau)\mathsf{P} + \tau(\mathsf{P}+2\mathsf{A})$, which approaches P as $\tau \downarrow 0$. (One can also reduce the power in the no-help scheme to $\mathsf{P} - 2\mathsf{A}\tau/(1-\tau)$ to guarantee that no excess power is used and then use the continuity of $C(0)$.)

The corresponding received symbol is $x_k(m) + \mathsf{A} - \hat{Z}_k + Z_k$. Upon subtracting the constant A, the receiver obtains

$$\tilde{Y}_k = x_k(m) + \tilde{Z}_k, \tag{100}$$

where

$$\tilde{Z}_k = Z_k - \hat{Z}_k. \tag{101}$$

Let $\tilde{\Delta}^2$ denote the conditional MSE estimation error given "no overflow,"

$$\tilde{\Delta}^2 = \mathbb{E}\big[\tilde{Z}_k^2 \,\big|\, F_k = 1\big]. \tag{102}$$

Conditional on $F_k = 1$, the density of Z_k is $\tilde{f}$ for which the quantizer was designed, so

$$\lim_{L\to\infty} L^2 \cdot \tilde{\Delta}^2 = \frac{1}{12}\|\tilde{f}\|_{1/3}. \tag{103}$$

To study the achievable rates in the with-help scheme, we study $I(X;Y)$. (As in the previous proofs, we now drop the time indices.) Using the chain rule and upper-bounding $I(X;F|Y)$ by $H(F)$,

$$\begin{aligned} I(X;Y) &\geq I(X;Y,F) - H(F) && (104)\\ &= I(X;Y|F) - H(F) && (105)\\ &\geq I(X;Y|F=1)\Pr(F=1) - H(F) && (106)\\ &= I(X;\tilde{Y}|F=1)\Pr(F=1) - H(F). && (107) \end{aligned}$$

The contribution of the with-help scheme to the overall rate is thus at least $\tau I(X;\tilde{Y}|F=1)\Pr(F=1) - \tau H(F)$. The second term will vanish as $\tau \downarrow 0$, so we focus on the first. Lower-bounding the (conditional) differential entropy of $\tilde{Y}$ by that of X, and recalling that the differential entropy is upper-bounded by that of a Gaussian of equal variance,

$$\tau I(X;\tilde{Y}|F=1) \geq \tau h(P_X) - \frac{\tau}{2}\log\big(2\pi e \tilde{\Delta}^2\big). \tag{108}$$

This inequality, (103), and (98) imply that

$$\varliminf_{\tau\downarrow 0} \tau I(X;\tilde{Y}|F=1) \geq R_\mathrm{h}, \tag{109}$$

and the overall achievable rate is thus lower-bounded by

$$C(0) + R_\mathrm{h}\Pr(F=1). \tag{110}$$

The achievability proof of $C(0) + R_\mathrm{h}$ is now concluded by letting A tend to infinity (while recalling (95)).

Converse: Assume Exponential noise, that M is drawn uniformly from its support set, that $\mathbf{X}$ is the result of encoding M, and that the help T is provided to both encoder and decoder.

$$\begin{aligned} nR &= I(M;\mathbf{Y},T) + H(M|\mathbf{Y},T) && (111)\\ &\leq I(M;\mathbf{Y},T) + n\delta_n && (112)\\ &= I(M;\mathbf{Y}|T) + n\delta_n && (113)\\ &= h(\mathbf{Y}|T) - h(\mathbf{Y}|M,T,\mathbf{X}) + n\delta_n && (114)\\ &= h(\mathbf{Y}|T) - h(\mathbf{Z}|M,T,\mathbf{X}) + n\delta_n && (115)\\ &= h(\mathbf{X}+\mathbf{Z}|T) - h(\mathbf{Z}|T) + n\delta_n && (116)\\ &\leq h(\mathbf{X}+\mathbf{Z}|T) - h(\mathbf{Z}) + nR_\mathrm{h} + n\delta_n && (117)\\ &\leq h(\mathbf{X}+\mathbf{Z}) - h(\mathbf{Z}) + nR_\mathrm{h} + n\delta_n && (118)\\ &\leq \sum_{k=1}^{n} h(X_k+Z_k) - n\log(e\mathsf{N}) + nR_\mathrm{h} + n\delta_n && (119)\\ &\leq \sum_{k=1}^{n} \log(e\,\mathbb{E}[X_k+Z_k]) - n\log(e\mathsf{N}) + nR_\mathrm{h} + n\delta_n && (120)\\ &\leq n\log\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] + \mathsf{N}\right) - n\log\mathsf{N} + nR_\mathrm{h} + n\delta_n && (121)\\ &\leq n\log(\mathsf{P}+\mathsf{N}) - n\log\mathsf{N} + nR_\mathrm{h} + n\delta_n, && (122) \end{aligned}$$

where (119) holds by the chain rule (and the fact that conditioning reduces entropy) and because the mean-N Exponential distribution is of differential entropy $\log(e\mathsf{N})$; (120) holds because, of all distributions on the nonnegatives with mean μ, the Exponential distribution maximizes differential entropy [3] and has differential entropy $\log(e\mu)$; (121) follows from the concavity of the logarithm; and (122) follows from the expectation constraint on the channel input (89). Dividing both sides of (122) by n and then letting n tend to infinity establishes that

$$R \leq \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}}\right) + R_\mathrm{h} \tag{123}$$

and thus concludes the proof of the converse. ■

V. The Modulo-Additive Noise Channel

We next study the modulo-additive noise channel and demonstrate that a discrete-alphabet variant of flash helping can also achieve the capacity of some channels with finite alphabets. We only consider noncausal helping.

The time-k output Y_k of the modulo-additive noise channel of time-k input x_k is

$$Y_k = x_k + Z_k, \tag{124}$$

where Z_k denotes the time-k noise; x_k, Z_k, and Y_k all take values in the modulo-additive group $\mathcal{X} = \{0, 1, \ldots, |\mathcal{X}|-1\}$; and "+" denotes addition modulo $|\mathcal{X}|$. Irrespective of the input sequence $\{x_k\}$, the noise samples $\{Z_k\}$ are IID according to some PMF P_Z of entropy $H(Z)$. In a blocklength-n communication, the helper observes the noise sequence noncausally and describes it to the encoder prior to transmission with at most nR_{h} bits. The setup is thus similar to the one of Section II with a noncausal helper, except for the finite channel alphabet and the absence of a power constraint.

Theorem 8: The capacity of the modulo-additive noise channel with noncausal help to the encoder is

$$C(R_{\mathrm{h}}) = \min\{\log|\mathcal{X}| - H(Z) + R_{\mathrm{h}},\, \log|\mathcal{X}|\} \tag{125}$$
$$= \min\{C(0) + R_{\mathrm{h}},\, \log|\mathcal{X}|\}. \tag{126}$$

Proof:
Achievability: Achievability is similar to the one for the case where the help is provided to the decoder [2] because, in the absence of power considerations, it does not matter whether it is the decoder or encoder that subtracts the noise: If R_{h} exceeds $H(Z)$, the helper can describe the noise sequence (almost) losslessly and the encoder can subtract it from its codeword to achieve (almost) noise-free communication. Otherwise, the helper describes the noise (almost) losslessly at rate $H(Z)$ during $R_{\mathrm{h}}/H(Z)$ of the time, and does not describe it at all during the remaining time. During the "with-help" phase, the encoder subtracts the noise based on the helper's description and communicates (nearly) noise-free at rate $\log|\mathcal{X}|$. During the "no-help" phase, the channel is treated as a modulo-additive noise channel without a helper, a channel whose capacity is $C(0) = \log|\mathcal{X}| - H(Z)$ [3, Theorem 7.2.1]. The aggregate achievable rate is thus

$$\frac{R_{\mathrm{h}}}{H(Z)} \log|\mathcal{X}| + \left(1 - \frac{R_{\mathrm{h}}}{H(Z)}\right)\bigl(\log|\mathcal{X}| - H(Z)\bigr) \tag{127}$$

which simplifies to $C(0) + R_{\mathrm{h}}$.

Converse: We establish the following chain of inequalities for a uniformly drawn message M and any sequence of coding schemes of vanishing probabilities of error:

$$\begin{aligned}
nR &= H(M) && (128)\\
&= I(M;\mathbf{Y},T) + H(M|\mathbf{Y},T) && (129)\\
&\le I(M;\mathbf{Y},T) + n\delta_n && (130)\\
&= \underbrace{I(M;T)}_{=0} + I(M;\mathbf{Y}|T) + n\delta_n && (131)\\
&= H(\mathbf{Y}|T) - H(\mathbf{Y}|M,T) + n\delta_n && (132)\\
&\le H(\mathbf{Y}) - H(\mathbf{Y}|M,T,\mathbf{X}) + n\delta_n && (133)\\
&= H(\mathbf{Y}) - H(\mathbf{Z}|T) + n\delta_n && (134)\\
&\le H(\mathbf{Y}) - \max\{nH(Z) - nR_{\mathrm{h}},\, 0\} + n\delta_n && (135)\\
&\le n\bigl(\log|\mathcal{X}| - \max\{H(Z) - R_{\mathrm{h}},\, 0\} + \delta_n\bigr) && (136)\\
&= n\bigl(\min\{C(0) + R_{\mathrm{h}},\, \log|\mathcal{X}|\} + \delta_n\bigr) && (137)
\end{aligned}$$

where (130) holds for some $\{\delta_n\}$ tending to zero by Fano's inequality; (131) holds because the helper is incognizant of the message and, consequently, T is independent of M; (133) holds because conditioning reduces entropy; (134) follows because $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ and because $(M,\mathbf{X}) \multimap T \multimap \mathbf{Z}$ forms a Markov chain; and (135) holds because $I(\mathbf{Z};T) \le H(T) \le nR_{\mathrm{h}}$ and because (conditional) entropy is nonnegative. The converse now follows by dividing by n and letting n tend to infinity. ■

Appendix A
No Cut-Set Bound for Encoder Assistance

We provide an example of a state-dependent discrete memoryless channel (SD-DMC) (albeit not an additive noise channel) whose encoder-assisted capacity can exceed its unaided capacity by more than the helper rate, i.e., for which $C(R_{\mathrm{h}}) > C(0) + R_{\mathrm{h}}$. This explains why the Cut-Set bound, which is so useful in analyzing decoder assistance, is not applicable to encoder assistance.

Consider the SD-DMC whose state S is uniform over the set $\mathcal{S} = \{\text{"red"}, \text{"blue"}\}$. The receiver observes the state in the sense that the channel output has the form (S, Y), where S is the channel state. Here Y is the "output number," which is an element of $[1:\eta]$. The channel input is a pair (σ, x), where $\sigma \in \mathcal{S}$ is the "input color," and $x \in [1:\eta]$ is the "input number." If the input color matches the state, then the output number is identical to the input number. Otherwise, it is uniformly distributed over $[1:\eta]$.

Suppose now that help at rate $R_{\mathrm{h}} = 1$ is provided causally to the encoder. Denoting the resulting capacity $C(R_{\mathrm{h}})$,

$$C(R_{\mathrm{h}}) \ge \log\eta, \tag{138}$$

because, when $R_{\mathrm{h}} = 1$, the helper can describe the state to the encoder precisely, and the latter can then choose its input color to match the state so as to obtain a clean channel from the input number to the output number.

Next consider $C(0)$. Given the input color σ and the input number x, the output color (which equals the state) is uniform over $\mathcal{S}$. If it is equal to the input color, then the output number is equal to the input number, else it is uniform over $[1:\eta]$. Thus, for $\tau, \sigma \in \mathcal{S}$ and $x, y \in [1:\eta]$,

$$W\bigl((\tau,y)\big|(\sigma,x)\bigr) = \begin{cases} \frac{1}{2} & \text{if } (\tau = \sigma) \wedge (y = x),\\ 0 & \text{if } (\tau = \sigma) \wedge (y \ne x),\\ \frac{1}{2\eta} & \text{if } \tau \ne \sigma. \end{cases} \tag{139}$$

Since S is independent of the input,

$$\begin{aligned}
&I\bigl((\Sigma,X);(S,Y)\bigr)\\
&\quad= I(\Sigma,X;Y|S) && (140)\\
&\quad= H(Y|S) - H(Y|\Sigma,S,X) && (141)\\
&\quad= H(Y|S) - \Pr(\Sigma \ne S) H(Y|\Sigma,S,X,\Sigma \ne S) && (142)\\
&\quad= H(Y|S) - \frac{1}{2}\log\eta && (143)\\
&\quad\le \frac{1}{2}\log\eta. && (144)
\end{aligned}$$

We conclude that

$$C(1) \ge C(0) + \frac{1}{2}\log\eta, \tag{145}$$

with $C(1) - C(0)$ being unbounded in the number of states.

APPENDIX B
PROOF OF PROPOSITION 1

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}2^{-2r_k}}\right) \\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}}\cdot 2^{2r_k}\right) && (146)\\
&\leq \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(\left(1+\frac{\mathsf{P}_k}{\mathsf{N}}\right)\cdot 2^{2r_k}\right) && (147)\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}}\right)+r_k\right] && (148)\\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}}\right)+\frac{1}{n}\sum_{k=1}^{n}r_k && (149)\\
&\leq \frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}}\right)+R_{\mathsf{h}} && (150)\\
&\leq \frac{1}{2}\log\left(\frac{1}{n}\sum_{k=1}^{n}\left[1+\frac{\mathsf{P}_k}{\mathsf{N}}\right]\right)+R_{\mathsf{h}} && (151)\\
&= \frac{1}{2}\log\left(1+\frac{1}{\mathsf{N}}\cdot\frac{1}{n}\sum_{k=1}^{n}\mathsf{P}_k\right)+R_{\mathsf{h}} && (152)\\
&\leq \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\mathsf{h}}, && (153)
\end{aligned}
$$

where (147) follows because $r_k \geq 0$ implies $2^{2r_k} \geq 1$ and because the logarithm is monotonically increasing, (150) follows from the condition (2), (151) follows from the concavity of the logarithm, and finally (153) follows from the condition (3) and again the fact that the logarithm is increasing.

As to the conditions guaranteeing equality as $n \to \infty$, let $\mathsf{P}_k \equiv \mathsf{P}$ and $r_k \equiv nR_{\mathsf{h}} \cdot \mathbf{1}\{k=\ell\}$ for some integer ℓ. Then

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n}\frac{1}{2}\log\left(1+\frac{\mathsf{P}_k}{\mathsf{N}2^{-2\ r_k}}\right) \\
&= \frac{n-1}{2n}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+\frac{1}{2n}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\cdot 2^{2nR_{\mathsf{h}}}\right) && (154)\\
&= \frac{n-1}{2n}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+\frac{1}{2n}\log\left(\left(2^{-2nR_{\mathsf{h}}}+\frac{\mathsf{P}}{\mathsf{N}}\right)\cdot 2^{2nR_{\mathsf{h}}}\right) && (155)\\
&= \frac{n-1}{2n}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right)+\frac{1}{2n}\log\left(2^{-2nR_{\mathsf{h}}}+\frac{\mathsf{P}}{\mathsf{N}}\right)+R_{\mathsf{h}}, && (156)
\end{aligned}
$$

which approaches the RHS of (5) as n tends to infinity.

ACKNOWLEDGMENT

The authors thank L. Wang for helpful discussions.

REFERENCES

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

[2] S. I. Bross, A. Lapidoth, and G. Marti, "Decoder-assisted communications over additive noise channels," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4150–4161, Jul. 2020.

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

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

[5] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.

[6] S. Hranilovic and F. R. Kschischang, "Capacity bounds for power- and band-limited optical intensity channels corrupted by Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 784–795, May 2004.

[7] A. Thangaraj, G. Kramer, and G. Bocherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, Jul. 2017.

[8] S. M. Moser, L. Wang, and M. Wigger, "Capacity results on multiple-input single-output wireless optical channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6954–6966, Nov. 2018.

[9] A. Lapidoth, S. M. Moser, and M. A. Wigger, "On the capacity of free-space optical intensity channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4449–4461, Oct. 2009.

[10] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446–472, Jul. 1948.

[11] S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*. Berlin, Germany: Springer, 2007.

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

[13] R. Tandon and S. Ulukus, "On the rate-limited Gelfand-Pinsker problem," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1963–1967.

[14] G. Grimmett and D. Stirzaker, *Probability and Random Processes*. New York, NY, USA: Oxford Univ. Press, 2001.

[15] G. Marti, "Channels with a helper," M.S. thesis, Dept. Inf. Technol. Elect. Eng., ETH Zürich, Zürich, Switzerland, 2019.

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

[17] S. Verdú, "The exponential distribution in information theory," *Problemy Peredachi Inform.*, vol. 32, no. 1, pp. 100–111, 1996.

[18] A. Martinez, "Communication by energy modulation: The additive exponential noise channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3333–3351, Jun. 2011.

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 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 the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland His research interests are in digital communications and information theory. He is the author of the textbook *A Foundation in Digital Communication*, second edition (Cambridge University Press, 2017).

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