Encoder-Assistance for Additive Noise Channels

Amos Lapidoth and Gian Marti
Signal and Information Processing Laboratory
ETH Zurich, 8092 Zurich, Switzerland
{lapidoth, marti}@isi.ee.ethz.ch

2020 IEEE Information Theory Workshop (ITW) | 978-1-7281-5962-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ITW46852.2021.9457638

***Abstract*—Flash helping has recently been shown to be an effective technique for describing additive noise to a decoder. It is shown here to be effective also in assisting the encoder: it achieves the helper capacity on the single-user Gaussian channel, on the multiple-access Gaussian channel, on the Exponential channel, and on the discrete modulo-additive noise channel. Most of the results hold irrespective of whether the helper observes the noise causally or noncausally.**

I. Introduction

Flash helping has recently been proposed as a capacity-achieving technique that allows a helper observing the channel noise to describe it over a rate-limited bit-pipe to the receiver in order to facilitate reliable communication [2]. This technique was extended in [3] to some multi-terminal settings, but with the help still being provided to the decoder(s). Here we consider help to the encoder. In this setting, one distinguishes between two cases: the *noncausal case*, where the encoder receives the description of the entire noise sequence before it begins to transmit, and the *causal case* where the time-k transmitted 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 settings. As a matter of fact, in all but one of the cases we consider, the two settings lead to identical capacities.[1] In those cases, we prove achievability for the causal case and the converse for the noncausal case.

The idea behind flash helping is to satisfy the rate constraint on by providing the help with great precision but infrequently. To see why this can outperform schemes that provide help with moderate precision continuously, consider a Gaussian noise channel where the noise variance is $\mathsf{N} > 0$, the maximal-allowed average power is $\mathsf{P} > 0$, and the helper's rate is $R_\mathrm{h} > 0$. For concreteness, 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 [4]). The estimate, which is known to the encoder prior to transmission, can be viewed as "dirt" in Costa's writing-on-dirty-paper setting and can be effectively canceled without any power penalty [5]. The remaining effective noise is the estimation error. This moderate-but-steady approach thus leads to an achievable rate of

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

An extended version of this paper has appeared in [1].

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

Alternatively, one could 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 description rate averaged over the block length n satisfies

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

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 following 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 but by flash helping, where r_k is zero for all k's other than some ℓ for which it equals nR_h. In other words, $r_k = nR_\mathrm{h}\cdot\mathbb{1}\{k=\ell\}$, where ℓ is in $[1:n]$ (the set $\{1,\dots,n\}$), and $\mathbb{1}\{\text{statement}\}$ 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*

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

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

Proof: Omitted. For a proof, see [1, Appendix B]. ■

The rest of the paper is organized as follows. In Section II we consider the single-user additive noise channel, and in Section III the multiple-access channel (MAC). 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 [6]. 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, thereby highlighting the structural similarity of the problems for finite and infinite alphabets.

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 $\mathsf{N} > 0$.

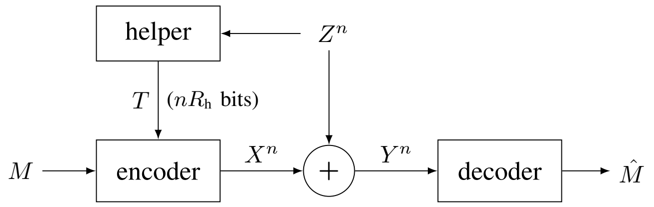

Fig. 1. The encoder-assisted single-user additive noise channel.

A rate-R message set $\mathcal{M}$ for a blocklength-n transmission is a set with 2^{nR} elements. For concreteness we assume that $\mathcal{M} = \{1, \ldots, 2^{nR}\}$. 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_h}, which, for concreteness, is assumed to be the set $\{1, \ldots, 2^{nR_\text{h}}\}$. 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 $(A_1, \ldots, A_k)$, 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 $\mathbb{E}\big[\|\mathbf{x}(m, T)\|^2\big] \leq n\mathsf{P}$. Here, $\|\cdot\|$ denotes the Euclidean norm, and $\mathsf{P} > 0$ 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 $(t_1, \ldots, t_n)$, where t_k takes value in a 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 total description length does not exceed nR_h bits, we impose the cardinality bound

$$|\mathcal{T}_1 \times \cdots \times \mathcal{T}_n| \leq 2^{nR_\text{h}}. \tag{7}$$

The time-k channel input $x_k(m, t_1, \ldots, t_k)$ produced by the encoder is determined by the message m and by the descriptions $t_1, \ldots, t_k$ received by time k. A blocklength-n causal encoder is thus described using n mappings $\{\phi^{(k)}_\text{c-enc}\}_{k=1}^n$ where

$$\phi^{(k)}_\text{c-enc} \colon \mathcal{M} \times \mathcal{T}_1 \times \cdots \times \mathcal{T}_k \to \mathbb{R} \tag{8}$$

and we require that

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\Big[x_k\big(m, t_1(Z^1), t_2(Z^2), \ldots, t_k(Z^k)\big)^2\Big] \leq \mathsf{P} \tag{9}$$

for all $m \in \mathcal{M}$. The supremum of all rates that allow for arbitrarily small probability of error (averaged over all messages) is the capacity $C(R_\text{h})$. Imposing a causality constraint cannot, of course, increase capacity.

Theorem 2. *The capacity of the average-power constrained additive Gaussian noise channel with a noncausal helper is*

$$C(R_\text{h}) = \frac{1}{2} \log\Big(1 + \frac{\mathsf{P}}{\mathsf{N}}\Big) + R_\text{h} \tag{10}$$

$$= C(0) + R_\text{h} \tag{11}$$

and can also be achieved with a causal helper.

We shall prove achievability for a causal helper and the converse for a noncausal one. The achievability part of the proof relies on Bennett's [7] classical result on high-resolution scalar quantization, which we quote from [8, Theorem 6.2]:

Theorem 3 (High-resolution scalar quantization). *Let Z be a random variable satisfying $\mathbb{E}\big[Z^{2+\delta}\big] < \infty$ for some $\delta > 0$ and having a density $f_Z(\cdot)$ satisfying $\|f_Z\|_{1/3} \triangleq \big(\int_\mathbb{R} f_Z(z)^{1/3}\, dz\big)^3 < \infty$.* [2] *Then there exists an L-level scalar quantizer which quantizes Z to $\hat{Z}$ so that*

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

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 $\lfloor 2^{R_\text{h}/\tau} \rfloor$-level scalar quantizer of the noise. Here $0 < \tau < 1$ is arbitrary but will later approach zero from above. 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. The helper describes the k-th noise sample Z_k using a MSE-minimizing L-level quantizer, where

$$L = \lfloor 2^{R_\text{h}/\tau} \rfloor, \tag{13}$$

[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.

and the sample is reconstructed from this description as $\hat{Z}_k$, where $\hat{Z}_k$ is the conditional expectation of Z_k given the description, so

$$\mathbb{E}\big[\hat{Z}_k^2\big] \leq \mathsf{N}, \tag{14}$$

and the quantization error $\tilde{Z}_k = Z_k - \hat{Z}_k$ satisfies (12). 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 $x_k(m) - \hat{Z}_k$. This, by (14), 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

$$\frac{1}{2}\log\left(1 + \frac{\mathsf{P}}{\mathbb{E}\big[(Z-\hat{Z})^2\big]}\right) \tag{15}$$

can be transmitted reliably [9]. The achievable rate with time sharing is thus

$$(1-\tau)\,C(0) + \tau\,\frac{1}{2}\log\left(1 + \frac{\mathsf{P}}{\mathbb{E}\big[(Z-\hat{Z})^2\big]}\right) \tag{16}$$

with power

$$(1-\tau)\mathsf{P} + \tau(\mathsf{P}+\mathsf{N}). \tag{17}$$

(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 $\mathsf{P} > 0$.) It follows from (13) and (12) that the expression in (16) converges to the RHS of (11) as $\tau \downarrow 0$, so that the achievability proof is complete.

Converse: The converse follows from the upper bound in [10, Sec. VI] by substituting zero for σ_Z^2. ■

Note that the achievability proof does not rely on the noise being Gaussian: It suffices for the quantization MSE of the optimal scalar quantizer to have the proper high-resolution asymptotic behavior, namely,

$$\varlimsup_{L\to\infty} L^2\,\mathbb{E}\Big[(Z-\hat{Z})^2\Big] < \infty. \tag{18}$$

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 a causal helper is then bounded by

$$C(R_\mathrm{h}) \geq C(0) + R_\mathrm{h}, \tag{19}$$

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

To prove the converse, it was required in [10, Sec. VI] that the conditional entropy $n^{-1}H(M|\mathbf{Y},T)$ tends to zero. This 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 (10) 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. We focus on the additive-noise MAC of Figure 2, whose time-k output Y_k is

$$Y_k = x_{1,k} + x_{2,k} + Z_k, \tag{20}$$

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 the average power constraints

$$\mathbb{E}\big[\|\mathbf{X}_i(m_i, T_i)\|^2\big] \leq n\mathsf{P}_i, \quad i = 1, 2, \tag{21}$$

be satisfied, where $\mathsf{P}_1, \mathsf{P}_2 > 0$ are the maximal-allowed average powers for the two users. The total description rate is denoted R_h,

$$R_\mathrm{h} = R_{\mathrm{h}_1} + R_{\mathrm{h}_2} \tag{22}$$

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

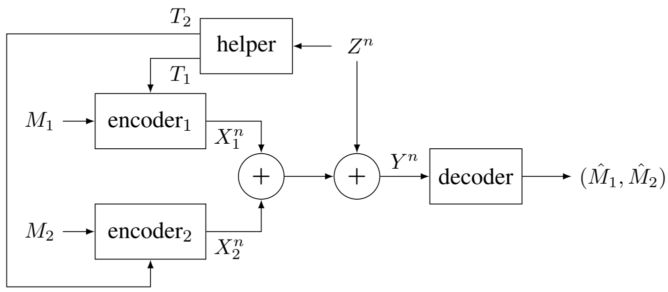

Fig. 2. The additive noise MAC with encoder-assistance.

As in the single-user case, our achievability result holds for causal help and arbitrary noise distribution (satisfying the hypotheses of Theorem 3) and the converse for noncausal help but with Gaussian noise. 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. In stating the result we shall use "+" to denote Minkowski set addition.

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

$$\mathcal{C}(0,0) + \big\{(R_1, R_2) \in \mathbb{R}_+ \times \mathbb{R}_+ :\ R_1 + R_2 \leq R_\mathrm{h}\big\} \tag{23}$$

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

Proof: Omitted. For a proof, see [1, Theorem 6] ■

IV. The Exponential Channel

The inputs to the channel we consider next must be nonnegative. This 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 nonetheless 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

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[X_k] \leq \mathsf{P}, \tag{24}$$

where $\mathsf{P} > 0$. The noise, which need not be nonnegative, is assumed to be of mean $\mathsf{N} \geq 0$. This channel reduces to the Exponential channel [11], [12] when N is positive and Z_k is a mean-N Exponential, i.e., of density

$$f_Z(z) = \frac{1}{\mathsf{N}}\exp\Big(-\frac{z}{\mathsf{N}}\Big)\mathbb{1}\{z \geq 0\}. \tag{25}$$

It reduces to the free-space optical channel [6] when Z_k is a centered Gaussian.

In the absence of help, the capacity of the Exponential channel is [11]

$$\frac{1}{2}\log\bigg(1+\frac{\mathsf{P}}{\mathsf{N}}\bigg). \tag{26}$$

We show that, also for this channel, the availability of 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*

$$C(R_{\mathrm{h}}) = \frac{1}{2}\log\left(1+\frac{\mathsf{P}}{\mathsf{N}}\right) + R_{\mathrm{h}} \tag{27}$$

$$= C(0) + R_{\mathrm{h}}, \tag{28}$$

regardless of whether the help is provided to the encoder causally or non-causally 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 (28) *is achievable with causal help, provided we interpret $C(0)$ as the capacity of the channel of said noise without help.*

Proof: 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. Adding a constant to the channel input does not impair communication.

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

$$F_k = \mathbb{1}\big\{|Z_k| \leq \mathsf{A}\big\}, \tag{29}$$

and note that the hypotheses of Theorem 3 guarantee that

$$\lim_{\mathsf{A}\to\infty}\Pr(F_k = 1) = 1. \tag{30}$$

Transmit

$$x_k(m) + \mathsf{A} - \hat{Z}_k, \tag{31}$$

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 given $|Z| \leq \mathsf{A}$. 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

$$L = 2^{R_{\mathrm{h}}/\tau}. \tag{32}$$

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

$$0 \leq \mathsf{A} - \hat{Z}_k \leq 2\mathsf{A}. \tag{33}$$

It follows from (31) and (33) 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 $(1-\tau)\mathsf{P} + \tau(\mathsf{P}+2\mathsf{A})$, 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

$$\tilde{Y}_k = x_k + \tilde{Z}_k, \tag{34}$$

where $\tilde{Z}_k = Z_k - \hat{Z}_k$. Let $\tilde{\Delta}^2$ denote the conditional MSE estimation error given "no overflow,"

$$\tilde{\Delta}^2 = \mathbb{E}\big[\tilde{Z}_k^2 | F_k = 1\big]. \tag{35}$$

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

$$\lim_{L\to\infty} L^2 \cdot \tilde{\Delta}^2 = \frac{1}{12}\|\tilde{f}\|_{1/3}. \tag{36}$$

To study the achievable rates in the with-help scheme, we study the mutual information $I(X;Y)$. Using the chain rule and upper-bounding $I(X;F|Y)$ by $H(F)$, we obtain the following bound:

$$I(X;Y) \geq I(X;Y,F) - H(F) \tag{37}$$

$$= I(X;Y|F) - H(F) \tag{38}$$

$$\geq I(X;Y|F=1)\Pr(F=1) - H(F) \tag{39}$$

$$= I(X;\tilde{Y}|F=1)\Pr(F=1) - H(F). \tag{40}$$

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,

$$\tau I(X;\tilde{Y}|F=1) \geq \tau h(P_X) - \frac{\tau}{2}\log\big(2\pi e\tilde{\Delta}^2\big). \tag{41}$$

By (36) and (32),

$$\varliminf_{\tau\downarrow 0} \tau I(X;\tilde{Y}|F=1) \geq R_{\mathrm{h}}, \tag{42}$$

and the overall achievable rate is thus lower-bounded by

$$C(0) + R_{\mathrm{h}} \Pr(F=1). \tag{43}$$

The achievability proof of $C(0) + R_{\mathrm{h}}$ is now concluded by letting A tend to infinity and recalling (30).

Converse: The converse follows by exploiting the Markovity $\mathbf{Z} \multimap T \multimap T$ and the entropy-maximizing properties of Exponential random variables; *cf.* [10, Sec. VI]. ■

V. The Modulo-Additive Noise Channel

By studying the modulo-additive noise channel, we demonstrate that a discrete-alphabet variant of flash helping can also achieve the capacity of some channels with finite alphabets. Here, however, a noncausal helper is used in the achievability proof.

Consider the modulo-additive noise channel

$$Y_k = x_k + Z_k, \tag{44}$$

where 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 sequence $\{Z_k\}$ is distributed IID according to P_Z of entropy $H(Z)$. The helper observes the noise sequence noncausally and describes it to the encoder at an average rate not exceeding R_{h}. 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 at the encoder is*

$$C(R_{\mathrm{h}}) = \min\{\log|\mathcal{X}| - H(Z) + R_{\mathrm{h}},\, \log|\mathcal{X}|\} \tag{45}$$

$$= \min\{C(0) + R_{\mathrm{h}},\, \log|\mathcal{X}|\}. \tag{46}$$

Proof: Achievability is shown exactly as in the case where help is available at the decoder [3]: the helper describes the noise (almost) losslessly at entropy-rate $H(Z)$ during $R_{\mathrm{h}}/H(Z)$ of the time, and does not describe it at all during the remaining time (assuming $R_{\mathrm{h}} \leq H(Z)$). During the "with-help" phase, the encoder cancels the noise based on the helper's description while communicating noise-free at rate $\log|\mathcal{X}|$. During the "no-help" phase, the channel is treated as a standard modulo-additive noise channel with achievable rate $C(0) = \log|\mathcal{X}| - H(Z)$ [4, Thm. 7.2.1]. The average achievable rate is thus

$$\frac{R_{\mathrm{h}}}{H(Z)}\log|\mathcal{X}| + \Big(1 - \frac{R_{\mathrm{h}}}{H(Z)}\Big)\big(\log|\mathcal{X}| - H(Z)\big) \tag{47}$$

which simplifies to $C(0) + R_{\mathrm{h}}$. If R_{h} exceeds $H(Z)$, then the first scheme is used throughout, thereby achieving $\log|\mathcal{X}|$.

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:

$$
\begin{aligned}
nR &= H(M) && (48)\\
&= I(M;\mathbf{Y},T) + H(M|\mathbf{Y},T) && (49)\\
&\leq I(M;\mathbf{Y},T) + n\delta_n && (50)\\
&= \underbrace{I(M;T)}_{=0} + I(M;\mathbf{Y}|T) + n\delta_n && (51)\\
&= H(\mathbf{Y}|T) - H(\mathbf{Y}|M,T) + n\delta_n && (52)\\
&\leq H(\mathbf{Y}) - H(\mathbf{Y}|M,T,\mathbf{X}) + n\delta_n && (53)\\
&= H(\mathbf{Y}) - H(\mathbf{Z}|T) + n\delta_n && (54)\\
&\leq H(\mathbf{Y}) - \max\{nH(Z) - nR_{\mathrm{h}},\, 0\} + n\delta_n && (55)\\
&\leq n\big(\log|\mathcal{X}| - \max\{H(Z) - R_{\mathrm{h}},\, 0\} + \delta_n\big) && (56)\\
&= n\big(\min\{C(0) + R_{\mathrm{h}},\, \log|\mathcal{X}|\} + \delta_n\big) && (57)
\end{aligned}
$$

where (50) holds for some $\{\delta_n\}$ tending to zero by Fano's inequality; (51) holds since the helper is incognizant of the message, so that T is independent of M; (54) follows because $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$ and because $(M, \mathbf{X}) \multimap T \multimap \mathbf{Z}$ is a Markov chain; and (55) holds because $I(\mathbf{Z};T) \leq H(T) \leq nR_{\mathrm{h}}$ and because (conditional) entropy is nonnegative. The converse follows by dividing by n and letting n tend to infinity. ■

References

[1] A. Lapidoth and G. Marti, "Encoder-assisted communications over additive noise channels," *IEEE Trans. on Inform. Theory*, vol. 66, no. 11, pp. 6607–6616, 2020.

[2] S. I. Bross and A. Lapidoth, "The additive noise channel with a helper," in *2019 IEEE Information Theory Workshop (ITW)*. Visby, Sweden: IEEE, Aug. 2019.

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

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. John Wiley & Sons, 2006.

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

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

[7] W. Bennett, "Spectra of quantized signals," *Bell Systems Technical Journal*, vol. 27, pp. 446–472, July 1948.

[8] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*. Springer, 2007.

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

[10] R. Tandon and S. Ulukus, "On the rate-limited Gelfand-Pinsker problem," in *2009 IEEE International Symposium on Information Theory*. IEEE, 2009, pp. 1963–1967.

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

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