Other Helper Capacities

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

***Abstract*—The erasures-only capacity, the listsize capacity, and the cutoff rate are computed for the modulo-additive noise channel with a helper. In one scenario the helper provides a rate-limited description of the noise sequence to the decoder and in the other to the encoder. In both scenarios the gains in these capacities thanks to the helper can exceed the helper's rate.**

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9517712

I. Introduction

Consider the memoryless modulo-additive noise channel whose time-k output Y_k corresponding to the time-k input x_k is

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

where $\{Z_k\} \sim$ IID Q_Z is the channel noise; x_k, Z_k, and Y_k all take values in the set $\mathcal{A} = \{0, 1, \ldots, |\mathcal{A}| - 1\}$; and "$\oplus$" denotes mod-$|\mathcal{A}|$ addition. The channel law $Q_{Y|X}(y \,|\, x)$ is thus

$$Q_{Y|X}(y \,|\, x) = Q_Z(y \ominus x), \quad \forall\, x, y \in \mathcal{A}, \tag{2}$$

where "$\ominus$" denotes mod-$|\mathcal{A}|$ subtraction.

In the absence of help, a blocklength-n code consists of a message set $\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$ and an encoding function $f\colon \mathcal{M} \to \mathcal{A}^n$, $m \mapsto \mathbf{x}(m) = (x_1(m), \ldots, x_n(m))$. Given the output sequence $\mathbf{Y} = (Y_1, \ldots, Y_n) = Y^n$, an erasures-only decoder declares an erasure if the list

$$\mathcal{L}(\mathbf{y}) = \{m \in \mathcal{M}\colon Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, m) > 0\} \tag{3}$$

contains more than one message and else produces the sole message in the list. A (zero-error) list decoder returns the list $\mathcal{L}(\mathbf{Y})$. Here $Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, m) = \prod_{i=1}^n Q_{Y|X}(y_i \,|\, x_i(m)) = Q^n_{Y|X}(\mathbf{y} \,|\, \mathbf{x}(m))$.

The *erasures-only capacity* $C_{\text{e-o}}$ [1], [2] (a.k.a. *zero-undetected-error capacity* [3] and *zero-error erasure capacity* [4]) is the supremum of the rates R for which there exists a sequence of blocklength-n coding schemes with $\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{M}| = R$ and

$$\lim_{n\to\infty} \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \Pr\big[|\mathcal{L}(\mathbf{Y})| \geq 2 \,\big|\, \mathbf{X} = \mathbf{x}(m)\big] = 0. \tag{4}$$

The *listsize capacity* $C_\ell(\rho)$ [5] (a.k.a. *zero-error list capacity* [6]) for $\rho > 0$ is similarly defined but with (4) replaced with

$$\lim_{n\to\infty} \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \mathbb{E}\big[|\mathcal{L}(\mathbf{Y})|^\rho \,\big|\, \mathbf{X} = \mathbf{x}(m)\big] = 1. \tag{5}$$

The *Shannon capacity* C and the *cutoff rate* $R_{\text{cutoff}}(\rho)$ are obtained by replacing $\mathcal{L}(\mathbf{y})$ in (4) and (5) with

$$\mathcal{L}(m, \mathbf{y}) = \{\tilde{m} \in \mathcal{M}\colon Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, \tilde{m}) \geq Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, m)\}, \tag{6}$$

i.e., by replacing the list of messages of positive a-posteriori probability with the list of messages that are a-posteriori at least as likely as the transmitted message. From the definitions,

$$C_{\text{e-o}} \leq C, \quad C_\ell(\rho) \leq R_{\text{cutoff}}(\rho). \tag{7}$$

The cutoff rate can be expressed as [6]

$$R_{\text{cutoff}}(\rho) = \max_{P_X} \frac{E_0(\rho, P_X)}{\rho}, \tag{8}$$

where $E_0(\rho, P_X)$ is Gallager's function

$$E_0(\rho, P_X) = -\log \sum_{y\in\mathcal{Y}} \Big(\sum_{x\in\mathcal{X}} P_X(x) \cdot Q_{Y|X}(y \,|\, x)^{\frac{1}{1+\rho}} \Big)^{1+\rho}. \tag{9}$$

Lemma 1. *For $\rho > 0$, Gallager's E_0 function for the modulo-additive noise channel with noise PMF Q_Z satisfies*

$$\max_{P_X} \frac{E_0(\rho, P_X)}{\rho} = \log |\mathcal{A}| - \mathrm{H}_{\tilde{\rho}}(Q_Z), \tag{10}$$

where $\mathrm{H}_{\tilde{\rho}}(\cdot)$ denotes the Rényi entropy of order $\tilde{\rho} \triangleq \frac{1}{1+\rho}$.

Proof: Omitted. ∎

Consider now a helper that is incognizant of the transmitted message M, but that observes the noise sequence $\mathbf{Z}$ and describes it as T, with T taking values in a finite set $\mathcal{T}$, and the rate of help R_{h} defined as $\limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{T}|$. We distinguish between two kinds of assistance:

Decoder assistance corresponds to the scenario where the noise description T is revealed to the decoder. The capacities $C_{\text{e-o,dec}}$, $C_{\ell,\text{dec}}$, C_{dec}, $R_{\text{cutoff,dec}}$, which are now functions also of R_{h}, are then defined by replacing the decoding lists $\mathcal{L}(\mathbf{y})$ and $\mathcal{L}(m, \mathbf{y})$ with $\mathcal{L}(\mathbf{y}, t)$ and $\mathcal{L}(m, \mathbf{y}, t)$, where we replace $Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, m)$ with $Q_{\mathbf{Y},T|M}(\mathbf{y}, t \,|\, m)$, i.e., with $Q_{\mathbf{Y},T|\mathbf{X}}(\mathbf{y}, t \,|\, \mathbf{x}(m))$ in the respective definitions.

Encoder assistance corresponds to the scenario where T is revealed noncausally to the encoder. The capacities $C_{\text{e-o,enc}}$, $C_{\ell,\text{enc}}$, C_{enc}, $R_{\text{cutoff,enc}}$ are defined by replacing the encoding function with $f\colon \mathcal{M} \times \mathcal{T} \to \mathcal{A}^n$, $(m, t) \mapsto \mathbf{x}(m, t)$, and the decoding lists are constructed with respect to $Q_{\mathbf{Y}|M}(\mathbf{y} \,|\, m) = \mathbb{E}_T\big[Q_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y} \,|\, \mathbf{x}(m, T), T)\big]$.

Recent studies of the benefits of decoder assistance [7], [8] and encoder assistance [9] have shown that the Shannon capacities C_{dec} and C_{enc} of such channels are increased by the rate of help (until saturating at $\log |\mathcal{A}|$). Here we derive analogous results for the other capacities. We show that the helper raises the erasures-only capacity to the same value to which it raises the Shannon capacity, and it raises the listsize

capacity to the same level it raises the cutoff rate, namely, to the sum of the rate of help and the cutoff rate.

The results when the rate of help is zero may seem paradoxical: even if zero in the absence of help, the erasures-only capacity with zero-rate help is equal to the Shannon capacity. This paradox is resolved by noting that zero-rate help is not equivalent to no help: Zero-rate help still allows the helper to describe the noise, albeit with a number of bits that is subexponential in the blocklength. As we shall see, such a description is all it takes to raise the erasures-only capacity to the Shannon capacity. A similar observation explains why zero-rate help increases the listsize capacity to the cutoff rate.

II. Decoder Assistance

A. Erasures-Only Capacity

Theorem 2. *The erasures-only capacity of the modulo-additive noise channel with noise probability mass function (PMF) Q_Z and rate-R_h decoder assistance is*

$$C_{\text{e-o, dec}}(R_\mathrm{h}) = \log|\mathcal{A}| - \big\{\mathrm{H}(Q_Z) - R_\mathrm{h}\big\}^+, \tag{11}$$

where $\{\xi\}^+$ denotes $\max\{0,\xi\}$.

Proof: The right-hand-side (RHS) of (11) is the Shannon capacity of this channel with rate-R_h decoder assistance [8, Theorem 12]. Since the erasures-only capacity never exceeds the Shannon capacity, we only need to establish achievability. We consider three cases:

- Case 1: $R_\mathrm{h} = 0$. Let Q_X be the uniform input distribution, $Q_{X,Y}$ the joint input-output distribution it induces, and Q_Y the corresponding (uniform) output distribution. Fix $\epsilon > 0$. Generate a random codebook $\{\mathbf{X}(m)\}_{m=1,\ldots,2^{nR}}$ of independent codewords, each having components drawn IID from Q_X. To send the message $m \in \mathcal{M}$, the encoder transmits the m-th codeword. The helper produces a one-bit description $T = \mathbb{1}\big\{\mathbf{Z} \in \mathcal{A}_\epsilon^{(n)}(Q_Z)\big\}$ of the noise sequence indicating whether or not it is weakly typical. The decoder receives the tuple $(\mathbf{Y}, T)$. If $T = 0$, it declares an erasure; otherwise, it searches for a message $\tilde{m}$ such that $(\mathbf{X}(\tilde{m}), \mathbf{Y}) \in \mathcal{A}_\epsilon^{(n)}(Q_{X,Y})$. If such an $\tilde{m}$ exists and is unique, it produces $\tilde{m}$. Otherwise, it declares an erasure.

To prove that undetected errors never occur, we note that the decoder attempts to decode only if $T = 1$, and that, as we next show, in this case the transmitted codeword $\mathbf{X}(m)$ is jointly typical with the received sequence $\mathbf{Y}$. Indeed, since Q_X and Q_Y are equiprobable, it follows that for all $\mathbf{x}, \mathbf{y} \in \mathcal{A}^n$

$$-\frac{1}{n}\log Q_X^n(\mathbf{x}) = \mathrm{H}(Q_X),\ -\frac{1}{n}\log Q_Y^n(\mathbf{y}) = \mathrm{H}(Q_Y) \tag{12}$$

and, consequently, $\mathbf{X}(m)$ is jointly typical with $\mathbf{Y}$ because

$$\begin{aligned}
&\left|-\frac{1}{n}\log Q_{X,Y}^n(\mathbf{X}(m),\mathbf{Y}) - \mathrm{H}(X,Y)\right| \\
&\quad= \left|-\frac{1}{n}\log Q_{Y|X}^n(\mathbf{Y}\,|\,\mathbf{X}(m)) - \mathrm{H}(Y\,|\,X)\right| && \text{(13a)}\\
&\quad= \left|-\frac{1}{n}\log Q_Z^n(\mathbf{Z}) - \mathrm{H}(Z)\right| && \text{(13b)}\\
&\quad\le \epsilon, && \text{(13c)}
\end{aligned}$$

where (13a) follows from the chain rule; (13b) holds because $Y = X \oplus Z$, with Z independent of X; and (13c) holds because, when T is 1, $\mathbf{Z} \in \mathcal{A}_\epsilon^{(n)}(Q_Z)$.

It remains to establish that the probability of erasure vanishes as n tends to infinity. An erasure is declared only if $\mathbf{Z}$ is atypical or if $(\mathbf{X}(\tilde{m}), \mathbf{Y}) \in \mathcal{A}_\epsilon^{(n)}(Q_{X,Y})$ for some $\tilde{m} \neq m$. The probability of the former tends to zero by the AEP [10], and the probability of the latter tends to zero whenever

$$R < \mathrm{I}(Q_X, Q_{Y|X}) = \log|\mathcal{A}| - \mathrm{H}(Q_Z). \tag{14}$$

and ϵ is sufficiently small.

- Case 2: $R_\mathrm{h} > \mathrm{H}(Q_Z)$. Fix $0 < \epsilon < R_\mathrm{h} - \mathrm{H}(Q_Z)$. The codebook we use in this case comprises all the distinct sequences in $\mathcal{A}^n$. The helper indicates by the bit $T_1 = 1/0$ whether the noise sequence is typical/atypical. If it is typical, the helper provides an almost lossless description of the noise sequence by producing as T_2 its index in $\mathcal{A}_\epsilon^{(n)}(Q_Z)$. (Otherwise, T_2 is arbitrary.) The decoder, upon receiving $(\mathbf{Y}, (T_1, T_2))$, declares an erasure if the noise is atypical, as indicated by $T_1 = 0$. Otherwise, it reconstructs the noise sequence from T_2 and subtracts it from $\mathbf{Y}$ to recover the codeword and hence the message. The rate $\log|\mathcal{A}|$ is thus achievable.
- Case 3: $0 < R_\mathrm{h} \le \mathrm{H}(Q_Z)$. For any $\delta > 0$, we divide the transmission block into two parts of relative length $\frac{R_\mathrm{h}}{(1+\delta)\mathrm{H}(Q_Z)}$ and $1 - \frac{R_\mathrm{h}}{(1+\delta)\mathrm{H}(Q_Z)}$. We then apply the aforementioned coding schemes for helper rates of $(1+\delta)\mathrm{H}(Q_Z)$ and zero, respectively. The total rate achieved by this time-sharing scheme is

$$\begin{aligned}
&\frac{R_\mathrm{h}\log|\mathcal{A}|}{(1+\delta)\mathrm{H}(Q_Z)} + \left(1 - \frac{R_\mathrm{h}}{(1+\delta)\mathrm{H}(Q_Z)}\right)\big(\log|\mathcal{A}| - \mathrm{H}(Q_Z)\big)\\
&\quad= \log|\mathcal{A}| - \mathrm{H}(Q_Z) + \frac{R_\mathrm{h}}{1+\delta}.
\end{aligned} \tag{15}$$

The result follows by taking $\delta \downarrow 0$. ∎

B. Listsize Capacity

Theorem 3. *For $\rho > 0$, the listsize capacity $C_{\ell,\,\text{dec}}$ and the cutoff rate $R_{\text{cutoff, dec}}$ of the modulo-additive noise channel with noise PMF Q_Z and rate-R_h decoder assistance are*

$$\begin{aligned}
C_{\ell,\,\text{dec}}(\rho, R_\mathrm{h}) &= R_{\text{cutoff, dec}}(\rho, R_\mathrm{h}) && \text{(16a)}\\
&= \log|\mathcal{A}| - \big\{\mathrm{H}_{\tilde{\rho}}(Q_Z) - R_\mathrm{h}\big\}^+. && \text{(16b)}
\end{aligned}$$

Proof: In light of (7), it suffices to prove achievability for $C_{\ell,\,\text{dec}}$ and the converse for $R_{\text{cutoff, dec}}$. We begin with achievability and consider three cases:

- Case 1: $R_\mathrm{h} = 0$. On account of Lemma 1, we only need to show that $C_{\ell,\,\text{dec}}(\rho, 0) \ge R_{\text{cutoff}}(\rho)$. To do so, we show how—starting with a sequence of codebooks $\{\mathcal{C}_n\}$ for which $\lim_{n\to\infty}\mathbb{E}\big[|\mathcal{L}(M,\mathbf{Y})|^\rho\big] = 1$ (without help)—we can construct a zero-rate helper for which the zero-error list for said codes satisfies

$$\mathbb{E}\big[|\mathcal{L}(\mathbf{Y},T)|^\rho\big] \le \mathbb{E}\big[|\mathcal{L}(M,\mathbf{Y})|^\rho\big], \tag{17}$$

and hence $\lim_{n\to\infty}\mathbb{E}\big[|\mathcal{L}(\mathbf{Y},T)|^\rho\big] = 1$.

We begin by indexing the family $\mathcal{P}_n$ of PMFs on $\mathcal{A}$ with denominator n. To send the message $m \in \mathcal{M}$, the encoder

transmits the m-th codeword in $\mathcal{C}_n$, and the helper produces as T the index of the empirical type $\hat{P}_{\mathbf{Z}}$ of the noise sequence. Since the cardinality of $\mathcal{P}_n$ is subexponential in n, the rate of help is zero.

On our channel (2), T determines the conditional probability of $\mathbf{Y}$ given the message m [10, Theorem 11.1.2]

$$Q_{\mathbf{Y}|M}(\mathbf{Y}\,|\,m) = 2^{-n(\mathrm{H}(\hat{P}_{\mathbf{Z}})+\mathrm{D}(\hat{P}_{\mathbf{Z}}\|Q_Z))}, \tag{18}$$

so the zero-error list can only contain messages $\tilde{m}$ for which $Q_{\mathbf{Y}|M}(\mathbf{Y}\,|\,\tilde{m}) = Q_{\mathbf{Y}|M}(\mathbf{Y}\,|\,m)$, and (17) follows from the inclusion

$$\begin{aligned}\mathcal{L}(\mathbf{Y},T) &= \{\tilde{m}\in\mathcal{M}\colon Q_{\mathbf{Y},T|M}(\mathbf{Y},T\,|\,\tilde{m})>0\} &\text{(19a)}\\ &\subseteq \{\tilde{m}\in\mathcal{M}\colon Q_{\mathbf{Y}|M}(\mathbf{Y}\,|\,\tilde{m}) = Q_{\mathbf{Y}|M}(\mathbf{Y}\,|\,m)\} &\text{(19b)}\\ &\subseteq \mathcal{L}(m,\mathbf{Y}). &\text{(19c)}\end{aligned}$$

- Case 2: $R_{\mathrm{h}} > \mathrm{H}_{\tilde{\rho}}(Q_Z)$. The codebook consists of all the sequences in $\mathcal{A}^n$. To send the message $m \in \mathcal{M} = \{1,\dots,|\mathcal{A}|^n\}$, the encoder transmits the m-th codeword. As to the helper, we rely on a result on Task Encoding [11, Theorem 1.2]: if $R_{\mathrm{h}} > \mathrm{H}_{\tilde{\rho}}(Q_Z)$, then there exists a sequence of mappings $f_n\colon \mathcal{A}^n \to \{1,\dots,2^{nR_{\mathrm{h}}}\}$ with pre-images $f_n^{-1}(t) = \{\mathbf{z}\in\mathcal{A}^n\colon f_n(\mathbf{z}) = t\}$ such that

$$\lim_{n\to\infty}\mathbb{E}\big[|f_n^{-1}(f_n(\mathbf{Z}))|^{\rho}\big] = 1. \tag{20}$$

Using this result, let the helper's description of the noise be $T = f_n(\mathbf{Z})$. Based on $(\mathbf{Y},T)$, the decoder produces the list

$$\mathcal{L}(\mathbf{Y},T) = \big\{m\in\mathcal{M}\colon f_n(\mathbf{Y}\ominus\mathbf{X}(m)) = T\big\}. \tag{21}$$

Its ρ-th moment is

$$\begin{aligned}\mathbb{E}\big[|\mathcal{L}(\mathbf{Y},T)|^{\rho}\big] &= \mathbb{E}\big[\big|\{\mathbf{x}\in\mathcal{A}^n\colon f_n(\mathbf{Y}\ominus\mathbf{x}) = T\}\big|^{\rho}\big] &\text{(22a)}\\ &= \mathbb{E}\big[|f_n^{-1}(f_n(\mathbf{Z}))|^{\rho}\big], &\text{(22b)}\end{aligned}$$

which, by (20), tends to 1 as n tends to infinity.

- Case 3: $0 < R_{\mathrm{h}} \le \mathrm{H}_{\tilde{\rho}}(Q_Z)$. For this case we propose time sharing as in the proof of Theorem 2. The details are omitted. This concludes the proof of achievability.

We now prove the converse by showing that

$$R_{\text{cutoff, dec}}(\rho,R_{\mathrm{h}}) \le \log|\mathcal{A}| - \big\{\mathrm{H}_{\tilde{\rho}}(Q_Z) - R_{\mathrm{h}}\big\}^+. \tag{23}$$

The upper bound $\log|\mathcal{A}|$ holds even for the capacity, so we focus on the case $0 \le R_{\mathrm{h}} < \mathrm{H}_{\tilde{\rho}}(Q_Z)$, for which we need Arıkan's lower bound on guessing [12].

Fix any rate-R blocklength-n codebook $\mathcal{C}_n$ and rate-R_{h} helper. Given $(\mathbf{y},t)$, list the messages $m\in\mathcal{M}$ in decreasing order of the likelihood $Q_{\mathbf{Y},T|M}(\mathbf{y},t\,|\,m)$ (resolving ties in some arbitrary fixed way, e.g., ranking low numerical values of m higher), and let $G(m\,|\,\mathbf{y},t)$ denote the ranking of the message m in this list, so

$$|\mathcal{L}(m,\mathbf{y},t)| \ge G(m\,|\,\mathbf{y},t). \tag{24}$$

where the inequality can be strict because of the way ties are resolved. It follows that the ρ-th moment of $|\mathcal{L}(m,\mathbf{y},t)|$ cannot tend to one unless the ρ-th moment of $G(m\,|\,\mathbf{y},t)$ does. We now establish a necessary condition for the latter:

Create a second list where the messages $m\in\mathcal{M}$ are listed in decreasing order of their likelihood in the absence of help (i.e., according to $\tilde{Q}_{\mathbf{Y}|M}(\mathbf{y}\,|\,m) = Q^n_{Y|X}(\mathbf{y}|\mathbf{x}(m))$), and let $\tilde{G}(m|\mathbf{y})$ denote the ranking in that list of the message m given $\mathbf{y}$. The functions $G(m|\mathbf{y},t)$ and $\tilde{G}(m|\mathbf{y})$ are thus optimal guessing functions with respect to $Q_{M,\mathbf{Y},T}$ and $\tilde{Q}_{M,\mathbf{Y}}$ in the sense of minimal ρ-th moment [13, Theorem 6.4]. Since

$$\tilde{Q}_{\mathbf{Y}|M}(\mathbf{y}\,|\,m) = \sum_{t\in\mathcal{T}} Q_{\mathbf{Y},T|M}(\mathbf{y},t\,|\,m), \tag{25}$$

it follows from [13, Proposition 6.9] that

$$\big|\mathcal{T}\big|^{\rho}\cdot\mathbb{E}\big[G(M\,|\,\mathbf{Y},T)^{\rho}\big] \ge \mathbb{E}\big[\tilde{G}(M\,|\,\mathbf{Y})^{\rho}\big], \tag{26}$$

where the RHS can be lower bounded as in [12, Eq. (14)]

$$\mathbb{E}\big[\tilde{G}(M\,|\,\mathbf{Y})^{\rho}\big] \ge (1+nR)^{-\rho}\cdot 2^{n(\rho R - \max_{P_X} E_0(\rho,P_X))}, \tag{27}$$

where E_0 is Gallager's function for the channel $Q_{Y|X}$. It follows from (26) and (27) that, as n tends to infinity, $\mathbb{E}\big[G(M\,|\,\mathbf{Y},T)^{\rho}\big]$ cannot tend to one (and hence by (24) nor can $\mathbb{E}\big[|\mathcal{L}(M,\mathbf{Y},T)|^{\rho}\big]$) unless

$$\begin{aligned}R &\le R_{\mathrm{h}} + \max_{P_X}\frac{E_0(\rho,P_X)}{\rho} &\text{(28a)}\\ &= R_{\mathrm{h}} + \log|\mathcal{A}| - \mathrm{H}_{\tilde{\rho}}(Q_Z), &\text{(28b)}\end{aligned}$$

where (28b) follows from Lemma 1. This establishes the converse. ∎

Remark 4. *The proof of the converse uses the additive nature of the channel only in* (28b)*, so* (8) *and* (28a) *imply that*

$$R_{\text{cutoff, dec}}(\rho,R_{\mathrm{h}}) \le R_{\mathrm{h}} + R_{\text{cutoff}}(\rho) \tag{29}$$

holds for general channels with rate-R_{h} decoder assistance.

Remark 5. *The technique we used to prove achievability in Case 1 can be used to establish that*

$$C_{\text{e-o, dec}}(R_{\mathrm{h}}) = C(R_{\mathrm{h}}),\quad C_{\ell,\text{ dec}}(\rho,R_{\mathrm{h}}) = R_{\text{cutoff, dec}}(\rho,R_{\mathrm{h}}) \tag{30}$$

on more general DMCs with IID states $\{S_k\}$, provided that Y_k is a function of (x_k,S_k), and S_k is a function of (x_k,Y_k).

III. Encoder Assistance

A. Erasures-Only Capacity

Theorem 6. *The erasures-only capacity of the modulo-additive noise channel with noise PMF Q_Z and rate-R_{h} encoder assistance is*

$$C_{\text{e-o, enc}}(R_{\mathrm{h}}) = \log|\mathcal{A}| - \big\{\mathrm{H}(Q_Z) - R_{\mathrm{h}}\big\}^+. \tag{31}$$

Proof: Since the RHS of (31) is the channel's Shannon capacity with encoder assistance [9, Theorem 8], we only need to establish achievability. The proof builds on the schemes proposed for Theorem 2. The idea is to convey the assistance to the decoder with negligible loss in transmission rate. We consider three cases:

- Case 1: $R_{\mathrm{h}} = 0$. We propose the following blocklength-$(n+1)$ scheme. For any $R < \log|\mathcal{A}| - \mathrm{H}(Q_Z)$ and $\epsilon > 0$ sufficiently small, consider a rate-R blocklength-n codebook

that, when used with decoder assistance as in Theorem 2, yields no undetected errors and small probability of erasure. Using $(1+\lceil\log|\mathcal{A}|\rceil)$ bits, the helper conveys to the encoder the bit $T_1 = \mathbb{1}\{Q_Z^n(Z_1^n) \geq 2^{-n(\mathrm{H}(Q_Z)+\epsilon)}\}$ and the noise sample $T_2 = Z_{n+1}$. To send the message m, the encoder transmits the codeword $\mathbf{x}(m)$, followed by $X_{n+1} = T_1 \ominus T_2$, so that $Y_{n+1} = T_1$. Equipped with T_1, the decoder can proceed as with decoder assistance. The overall rate is $nR/(n+1)$, which approaches R.

- Case 2: $R_\mathrm{h} > \mathrm{H}(Q_Z)$. Fix $0 < \epsilon < R_\mathrm{h} - \mathrm{H}(Q_Z)$. As in Case 1, the helper and encoder can convey the bit $T_1 = \mathbb{1}\{Z_1^n \in \mathcal{A}_\epsilon^{(n)}(Q_Z)\}$ to the decoder. Additionally, if the noise is typical, the helper can describe it to the encoder, who can then subtract it from the codeword.
- Case 3: $0 < R_\mathrm{h} \leq \mathrm{H}(Q_Z)$. Follows by time sharing. ∎

B. Listsize Capacity

Theorem 7. *For $\rho > 0$, the listsize capacity $C_{\ell,\,\mathrm{enc}}$ and the cutoff rate $R_{\mathrm{cutoff},\,\mathrm{enc}}$ of the modulo-additive noise channel with noise PMF Q_Z and rate-R_h encoder assistance are*

$$C_{\ell,\,\mathrm{enc}}(\rho, R_\mathrm{h}) = R_{\mathrm{cutoff},\,\mathrm{enc}}(\rho, R_\mathrm{h}) \tag{32a}$$

$$= \log|\mathcal{A}| - \left\{\mathrm{H}_{\tilde{\rho}}(Q_Z) - R_\mathrm{h}\right\}^+. \tag{32b}$$

Proof: We prove achievability of $C_{\ell,\,\mathrm{enc}}$ and converse for $R_{\mathrm{cutoff},\,\mathrm{enc}}$. For achievability, we consider three cases:

- Case 1: $R_\mathrm{h} = 0$. With negligible rate loss, $\lceil\log|\mathcal{P}_n(\mathcal{A})|/\log|\mathcal{A}|\rceil$ extra channel uses can be used to convey the type of Z_1^n to the decoder, and the problem reduces to that of decoder assistance.
- Case 2: $R_\mathrm{h} > \mathrm{H}_{\tilde{\rho}}(Q_Z)$. Fix $\epsilon, \delta > 0$ and sufficiently large blocklength n. The codebook is $\mathcal{A}^n$, and coding is in three phases, with Phase 1 of length $n_1 = n$, and Phases 2 and 3 of lengths n_2 and n_3 to be specified later, but both negligible compared to n. The helper can thus describe the noise corrupting the channel in Phases 2 and 3 with negligible overhead, allowing the encoder to subtract the noise and thereby enabling error-free transmission in those two phases.

Observing the Phase 1 noise Z_1^n (denoted hereafter $\mathbf{Z}$), the helper conveys its empirical type $T_2 = \hat{P}_\mathbf{Z}$ to the encoder, who conveys it to the decoder in Phase 2. (Since the number of types is subexponential, both the length of the helper's description T_2 and the number of channel uses the encoder needs to convey T_2 to the decoder are negligible.) Thereafter, $\hat{P}_\mathbf{Z}$ is known to all parties. Consider two subcases:

(i) $R_\mathrm{h} \geq \mathrm{H}(\hat{P}_\mathbf{Z})$. In this subcase, the encoder, with the helper's assistance, can subtract the Phase 1 noise corrupting the codeword: Since $|\top_{\hat{P}_\mathbf{Z}}^{(n)}| \leq 2^{n\mathrm{H}(\hat{P}_\mathbf{Z})} \leq 2^{nR_\mathrm{h}}$, the helper can set T_1 to be the index of $\mathbf{Z}$ in $\top_{\hat{P}_\mathbf{Z}}^{(n)}$, so that T_1 and T_2 jointly determine $\mathbf{Z}$, allowing the encoder to subtract $\mathbf{Z}$ from the codeword in Phase 1. The decoder recovers the codeword as Y_1^n. Phase 3 is unnecessary and zero padding is applied to exhaust the frame.

(ii) $R_\mathrm{h} < \mathrm{H}(\hat{P}_\mathbf{Z})$. In this subcase, the noise samples in only nq locations are described and subtracted, where $q = \lceil nR_\mathrm{h}/\mathrm{H}(\hat{P}_\mathbf{Z})\rceil/n$. The issue is how to choose the locations and how to convey them to the decoder with negligible overhead. To this end, we use as location indicator a binary codeword $\mathbf{U}$ in a size-$2^{n\epsilon}$ codebook $\mathcal{C}(\hat{P}_\mathbf{Z}) \subset \{0,1\}^n$, which is specifically designed for $\hat{P}_\mathbf{Z}$, and each of whose codewords has nq components equal to 1. Prior to transmission, codebooks are designed and agreed upon by all parties for each possible $\hat{P}_\mathbf{Z}$, i.e., for each PMF in $\mathcal{P}_n(\mathcal{A})$. The index of the codeword $\mathbf{U} \in \mathcal{C}(\hat{P}_\mathbf{Z})$ (with the dependence of $\mathcal{C}$ on $\hat{P}_\mathbf{Z}$ henceforth made implicit) is described using $n\epsilon$ bits to the encoder who conveys it to the decoder in Phase 3. With $\mathbf{U}$ known to all parties, the helper can describe the noise samples at the corresponding locations, the encoder subtracts them from the codeword, and the decoder can then be certain of the value of the codeword at these locations. Paramount is that the noise samples at the locations indicated by the codeword have an empirical type that allows their description. This is where the codebook construction is critical.

To specify how $\mathcal{C}$ is constructed and how $\mathbf{U}$ is chosen, we need some notation. Given a binary n-tuple $\mathbf{u}$, we define $\mathcal{I}(\mathbf{u}) \triangleq \{i \in [1\colon n]\colon u_i = 1\}$ and $\bar{\mathbf{u}} \triangleq \mathbf{1} - \mathbf{u}$, so $\mathcal{I}(\mathbf{u}) \sqcup \mathcal{I}(\bar{\mathbf{u}}) = [1\colon n]$, where "$\sqcup$" denotes the disjoint union. If $\mathcal{I} \subset [1\colon n]$ is of elements $i_1 < i_2 < \cdots < i_{|\mathcal{I}|}$ and $\mathbf{z} \in \mathcal{A}^n$, then $\mathbf{z}(\mathcal{I})$ is the tuple $(z_{i_1}, z_{i_2}, \ldots, z_{i_{|\mathcal{I}|}}) \in \mathcal{A}^{|\mathcal{I}|}$ "picked by $\mathcal{I}$", and its empirical type is $\hat{P}_{\mathbf{z}(\mathcal{I})} \in \mathcal{P}_{|\mathcal{I}|}(\mathcal{A})$. For simplicity of notation, $\mathbf{z}(\mathcal{I}(\mathbf{u}))$ is also denoted $\mathbf{z}(\mathbf{u})$.

The codebook $\mathcal{C}$ is a type-covering [13, Lemma 2.34] rate-distortion codebook for some joint type $\hat{P}_{\mathbf{Z},\mathbf{U}} \approx \hat{P}_\mathbf{Z} \circ \mathrm{Ber}(q)$, whose Z-marginal is $\hat{P}_\mathbf{Z}$, whose U-marginal is $\mathrm{Ber}(q)$, and under which Z and U are "nearly" independent. This near independence guarantees that the empirical type of $\mathbf{Z}(\mathbf{U})$—which under $\hat{P}_{\mathbf{Z},\mathbf{U}}$ equals $P_{Z|U=1}$—is approximately $\hat{P}_\mathbf{Z}$, so

$$\left|\mathrm{H}(\hat{P}_{\mathbf{Z}(\mathbf{U})}) - \mathrm{H}(\hat{P}_\mathbf{Z})\right| < \delta. \tag{33}$$

The codebook $\mathcal{C} \subset \{0,1\}^n$, whose proof of existence is omitted, is thus such that: (1) $|\mathcal{C}| = 2^{n\epsilon}$, (2) each codeword $\mathbf{u} \in \mathcal{C}$ is of type $\mathrm{Ber}(q)$, i.e. $|\mathcal{I}(\mathbf{u})| = nq$, and (3) to each $\mathbf{z}$ of type $\hat{P}_\mathbf{Z}$, there corresponds some codeword $\mathbf{u} \in \mathcal{C}$ such that $(\mathbf{z}, \mathbf{u})$ is of type $\hat{P}_{\mathbf{Z},\mathbf{U}}$.

The coding scheme is the following. The helper sets T_3 to be the index of the codeword $\mathbf{U} \in \mathcal{C}$ whose joint type with the noise $\mathbf{Z}$ is $\hat{P}_{\mathbf{Z},\mathbf{U}}$. The index of the codeword $\mathbf{U}$ is conveyed to the decoder in Phase 3. Since the noise sequence $\mathbf{Z}(\mathbf{U})$ picked by $\mathcal{I}(\mathbf{U})$ is of type $\hat{P}_{\mathbf{Z}(\mathbf{U})}$, and since, by (33),

$$\left|\top_{\hat{P}_{\mathbf{Z}(\mathbf{U})}}^{(|\mathcal{I}(\mathbf{U})|)}\right| \leq 2^{nq\mathrm{H}(\hat{P}_{\mathbf{Z}(\mathbf{U})})} < 2^{n(R_\mathrm{h}+\delta)}, \tag{34}$$

the helper can convey $\mathbf{Z}(\mathbf{U})$ to the encoder using some universal code. In Phase 1, the encoder subtracts the noise at locations $\mathcal{I}(\mathbf{U})$ from the codeword. The decoder, with the knowledge of $\mathbf{U}$, recovers these nq components received correctly, and, with $\hat{P}_{\mathbf{Z},\mathbf{U}}$, knows the type of the subsequence at the remaining $n(1-q)$ components $\hat{P}_{\mathbf{Z}(\bar{\mathbf{U}})}$. Its list is thus of size

$$\left|\top_{\hat{P}_{\mathbf{Z}(\bar{\mathbf{U}})}}^{(|\mathcal{I}(\bar{\mathbf{U}})|)}\right| \leq 2^{n(1-q)\mathrm{H}(\hat{P}_{\mathbf{Z}(\bar{\mathbf{U}})})} \tag{35a}$$

$$\leq 2^{n\mathrm{H}(\hat{P}_\mathbf{Z}) - nq\mathrm{H}(\hat{P}_{\mathbf{Z}(\mathbf{U})})} < 2^{n(\mathrm{H}(\hat{P}_\mathbf{Z}) - R_\mathrm{h} + \delta)}, \tag{35b}$$

where the first inequality of (35b) holds by the convexity of entropy and the relation $\hat{P}_{\mathbf{Z}} = q\hat{P}_{\mathbf{Z}(\mathbf{U})} + (1-q)\hat{P}_{\mathbf{Z}(\bar{\mathbf{U}})}$.

The duration of Phase 2 and 3 is

$$n_2 = \left\lceil \frac{\log|\mathcal{P}_n(\mathcal{A})|}{\log|\mathcal{A}|} \right\rceil, \quad n_3 = \left\lceil \frac{n\epsilon}{\log|\mathcal{A}|} \right\rceil. \tag{36}$$

to allow the description of $\hat{P}_{\mathbf{Z}}$ in Phase 2 and of the index of the locator codeword in Phase 3. By letting $\epsilon \downarrow 0$ as $n \to \infty$, the overall assistance rate and transmission rate tend to R_{h} and $\log|\mathcal{A}|$, respectively.

We next prove that the ρ-th moment of the output list tends to 1 in both subcases. Conditional on $\hat{P}_{\mathbf{Z}}$, the ρ-th moment of listsize is upper bounded by

$$\mathbb{E}\big[|\mathcal{L}(Y^{n+n_2+n_3})|^{\rho} \,\big|\, \hat{P}_{\mathbf{Z}}\big] \le 1 + 2^{n\rho\left(\mathrm{H}(\hat{P}_{\mathbf{Z}}) - R_{\mathrm{h}} + \delta\right)}. \tag{37}$$

Taking the expectation of (37) over $\hat{P}_{\mathbf{Z}}$, upper bounding the probability of a type class [10, Theorem 11.1.4], and recalling that the number of types is subexponential in n, the ρ-th moment of the listsize can be bounded from above by

$$1 + \max_{P_Z \in \mathcal{P}_n(\mathcal{A})} 2^{n\rho\left(\mathrm{H}(P_Z) - \rho^{-1}\mathrm{D}(P_Z\|Q_Z) - R_{\mathrm{h}} + \delta + o(1)\right)}. \tag{38}$$

This expression tends to 1 as n tends to infinity (for δ small enough) because

$$R_{\mathrm{h}} > \mathrm{H}_{\tilde{\rho}}(Q_Z) = \max_{P_Z \in \mathcal{P}(\mathcal{A})} \big\{\mathrm{H}(P_Z) - \rho^{-1}\mathrm{D}(P_Z\|Q_Z)\big\}. \tag{39}$$

- Case 3: $0 < R_{\mathrm{h}} \le \mathrm{H}_{\tilde{\rho}}(Q_Z)$. Follows by time sharing.

The converse is implied by the following stronger statement.

Claim 8. *The cutoff rate of the modulo-additive noise channel with rate-R_{h} assistance that is provided to both the decoder and the encoder is upper bounded (for any $\rho > 0$) by*

$$R_{\text{cutoff, both}}(\rho, R_{\mathrm{h}}) \le \log|\mathcal{A}| - \big\{\mathrm{H}_{\tilde{\rho}}(Q_Z) - R_{\mathrm{h}}\big\}^{+}. \tag{40}$$

Here $R_{\text{cutoff, both}}$ is defined with the encoding function having the form $f(m,t)$ and the decoding list $\mathcal{L}(m,\mathbf{y},t)$ corresponding to $Q_{\mathbf{Y},T|M}(\mathbf{y},t\,|\,m) = Q_T(t)\cdot Q_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}\,|\,\mathbf{x}(m,t),t)$.

From this claim, the converse follows because

$$R_{\text{cutoff, enc}}(\rho, R_{\mathrm{h}}) \le R_{\text{cutoff, both}}(\rho, R_{\mathrm{h}}). \tag{41}$$

Proof of Claim 8. Since the cutoff rate is always upper bounded by $\log|\mathcal{A}|$, we focus on the case $0 \le R_{\mathrm{h}} < \mathrm{H}_{\tilde{\rho}}(Q_Z)$. Conditional on $T = t$, the channel reduces to the non-IID modulo-additive noise channel $Q_{\mathbf{Y}|\mathbf{X},T=t}$ with the noise distribution $Q_{\mathbf{Z}|T=t}$, and the encoding function $m \mapsto \mathbf{x}(m,t)$ can be described using the codebook $\mathcal{C}_t = \big\{\mathbf{x}(m,t)\colon m \in \mathcal{M}\big\}$.

Fix any rate-R_{h} helper and any family of rate-R blocklength-n codebooks $\{\mathcal{C}_t\}_{t\in\mathcal{T}}$. Given t and $\mathbf{y}$, list the messages in decreasing order of the likelihood $Q_{\mathbf{Y}|M,T}(\mathbf{y}\,|\,m,t) = Q_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}\,|\,\mathbf{x}(m,t),t)$, and denote the ranking of message m in the list by $G(m\,|\,\mathbf{y},t)$, so

$$|\mathcal{L}(m,\mathbf{y},t)| \ge G(m\,|\,\mathbf{y},t). \tag{42}$$

By Arıkan's inequality on guessing [12, Section III], the conditional ρ-th moment of the RHS can be lower bounded by

$$\mathbb{E}\big[G(M\,|\,\mathbf{Y},T)^{\rho}\,\big|\,T=t\big] \ge (1+nR)^{-\rho}\cdot 2^{n\rho R - \max_{P_{\mathbf{X}}} E_0^{(n)}(\rho, P_{\mathbf{X}}, Q_{\mathbf{Y}|\mathbf{X},T=t})}, \tag{43}$$

where $E_0^{(n)}$ is Gallager's function of $Q_{\mathbf{Y}|\mathbf{X},T=t}$, the modulo-additive noise channel over length-n superalphabets. By applying Lemma 1 to superalphabets of length n,

$$\max_{P_{\mathbf{X}}} E_0^{(n)}(\rho, P_{\mathbf{X}}, Q_{\mathbf{Y}|\mathbf{X},T=t}) = \rho\log|\mathcal{A}|^n - \rho\mathrm{H}_{\tilde{\rho}}(Q_{\mathbf{Z}|T=t}). \tag{44}$$

It follows from (43) and (44) by averaging over all $t \in \mathcal{T}$ that

$$\mathbb{E}\big[G(M\,|\,\mathbf{Y},T)^{\rho}\big] \ge (1+nR)^{-\rho}\cdot 2^{n\rho(R - \log|\mathcal{A}|) + \rho\mathrm{H}_{\tilde{\rho}}(\mathbf{Z}\,|\,T)}. \tag{45}$$

The RHS of (45) cannot tend to one as n tends to infinity unless

$$R \le \log|\mathcal{A}| - \frac{1}{n}\mathrm{H}_{\tilde{\rho}}(\mathbf{Z}\,|\,T) \tag{46a}$$

$$\le \log|\mathcal{A}| - \frac{1}{n}(-\log\big|\mathcal{T}\big| + \mathrm{H}_{\tilde{\rho}}(\mathbf{Z})) \tag{46b}$$

$$= \log|\mathcal{A}| - \mathrm{H}_{\tilde{\rho}}(Z) + R_{\mathrm{h}} + o(1), \tag{46c}$$

where (46b) follows from the chain rule of Rényi entropy [14, Theorem 3]. This proves Claim 8 and hence the converse. ■

Remark 9. *Claim 8 notwithstanding, for general channels, $R_{\text{cutoff, enc}}(\rho, R_{\mathrm{h}}) - R_{\text{cutoff}}(\rho)$ can be arbitrarily large, as can be illustrated by an example like the one in [9, Appendix A].*

Remark 10. *Some of the paper's results extend to general cost-constrained additive-noise channels*

$$Y_k = x_k + Z_k, \tag{47}$$

where the noise $\{Z_k\}$ is IID with density f_Z; the symbols x_k, Z_k, $Y_k \in \mathbb{R}$ are real; addition is over the reals; and each codeword $\mathbf{x}(m)$ satisfies

$$\frac{1}{n}\sum_{k=1}^{n} g(x_k(m)) \le \Gamma \tag{48}$$

and

$$x_k(m) \in [a,b], \quad \forall\, k \in [1\colon n], \tag{49}$$

for given $g\colon \mathbb{R} \to \mathbb{R}^+$, $\Gamma > 0$, and $-\infty \le a < b \le +\infty$.

For such channels with rate-R_{h} decoder assistance, the erasures-only capacity $C_{\text{e-o, dec}}$ is given by

$$C_{\text{e-o, dec}}(R_{\mathrm{h}}) = C_{\text{dec}}(R_{\mathrm{h}}) \tag{49a}$$

$$= \max_{f_X\,:\,\mathbb{E}[g(X)]\le\Gamma} \mathrm{I}(f_X; f_{Y|X}) + R_{\mathrm{h}}; \tag{49b}$$

and with rate-R_{h} encoder assistance, the erasures-only capacity $C_{\text{e-o, enc}}$ satisfies

$$C_{\text{e-o, enc}}(R_h) \ge \max_{f_X\,:\,\mathbb{E}[g(X)]\le\Gamma} \mathrm{I}(f_X; f_{Y|X}) + R_{\mathrm{h}}, \tag{50}$$

with the lower bound being tight when f_Z is Gaussian, g quadratic, $a = -\infty$, and $b = +\infty$.

REFERENCES

[1] M. S. Pinsker and A. Y. Sheverdjaev, "Transmission capacity with zero error and erasure," *Problemy Peredachi Informatsii*, vol. 6, no. 1, pp. 20–24, 1970.

[2] I. Csiszar and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 35–43, 1995.

[3] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity approaches the Sperner capacity," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3825–3833, 2014.

[4] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 55–62, 1996.

[5] C. Bunte and A. Lapidoth, "On the listsize capacity with feedback," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6733–6748, 2014.

[6] I. E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1977–1982, 1997.

[7] Y. Kim, "Capacity of a class of deterministic relay channels," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1328–1329, 2008.

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

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

[10] T. M. Cover, *Elements of information theory*, 2nd ed. John Wiley & Sons, Inc., 2006.

[11] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5065–5076, 2014.

[12] E. Arıkan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[13] S. M. Moser, *Advanced Topics in Information Theory (Lecture Notes)*, 4th ed., 2019, https://moser-isi.ethz.ch/docs/atit_script_v45.pdf.

[14] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6801–6810, 2014.