On the Zero-Error Capacity of the Modulo-Additive Noise Channel With Help

Amos Lapidoth, *Fellow, IEEE*, and Yiming Yan, *Graduate Student Member, IEEE*

***Abstract*— The zero-error helper capacity of the modulo-additive noise channel is studied both in the presence and in the absence of feedback. In its presence, a complete solution of said capacity is provided. In its absence, a solution is provided when the alphabet size is prime. For all other cases, upper and lower bounds are derived, and a necessary and sufficient condition for positivity is provided. Thanks to the help, the zero-error capacity may increase by more than the help's rate, and it can be positive yet smaller than one bit.**

***Index Terms*— Feedback, helper, list decoding, modulo-additive noise channel, zero-error capacity.**

I. Introduction

THIS paper investigates the extent to which the zero-error capacity can benefit from a rate-limited description of the noise. We study both encoder assistance, where the description is provided to the encoder before transmission begins, and decoder assistance, where it is provided to the decoder. We show that, perhaps paradoxically, the zero-error helper capacity can be calculated as a function of the description rate even for some channels whose no-help zero-error capacity is unknown. This is not a contradiction, because a zero-rate description is not tantamount to no description: it still allows for a binary description of length that is sublinear in the blocklength. In fact, as we shall see, the solution of the zero-rate help case is the key to the general solution.

We focus on the memoryless modulo-additive noise channel (MMANC) 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 $\mathcal{A} = \{0, 1, \dots, |\mathcal{A}| - 1\}$; and "$\oplus$" denotes mod-$|\mathcal{A}|$ addition. The channel law $Q_{Y|X}(\cdot|\cdot)$ is thus

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

where "$\ominus$" denotes mod-$|\mathcal{A}|$ subtraction. A key role is played by the cardinality $|\mathcal{S}|$ of the support set $\mathcal{S}$ of Q_Z

$$\mathcal{S} = \{z \in \mathcal{A} \colon Q_Z(z) > 0\}. \tag{3}$$

Example 1: When $|\mathcal{S}| = 2$, the MMANCs corresponding to $|\mathcal{A}|$ being 3, 5, and 7 are, respectively, the Triangle channel, Shannon's Pentagon channel [1], and the Heptagon channel (a.k.a. the $3/2$, $5/2$, and $7/2$ channels, respectively).

In the presence of a noiseless feedback link from the receiver to the encoder, we calculate the zero-error helper capacity both for encoder and for decoder assistance (Theorem 3). In its absence, we derive upper and lower bounds on the zero-error helper capacity (Theorem 5) and establish a positivity result for the zero-error capacity (Corollary 3): if the assistance rate is positive, then so is the capacity; otherwise, the capacity is positive if and only if (iff) the support $\mathcal{S}$ of the noise is a strict subset of $\mathcal{A}$. When the cardinality of $\mathcal{A}$ is prime (as in Example 1) we calculate the zero-error helper capacity in Theorem 4 using structured codes. Calculating the zero-error helper capacity without feedback when $|\mathcal{A}|$ is not prime is left as an open problem.

These results add to the body of literature on the benefits of helpers as measured in terms of the rare-error capacity[1] [3], [4], [5], [6], [7], error exponents [8], erasures-only capacity [5], listsize capacity [5], [9], and secrecy [10].

The rest of the paper is organized as follows. Section II introduces some notation, defines the key quantities of interest, and surveys some of the literature that touches on this work. Section III presents the paper's main results and some of their consequences. The proof of Theorem 3 pertaining to feedback is presented in Section IV. The proofs of Theorems 4 and 5 pertaining to the no-feedback setting are presented in Sections V-A and V-B respectively.

II. Preliminaries

A. Notation

Unless stated otherwise, all logarithms in this paper are to base 2. The positive integers are denoted $\mathbb{Z}^+$, and if $n \in \mathbb{Z}^+$, then $[n]$ denotes the set $\{1, 2, \dots, n\}$. The cardinality of a set $\mathcal{K}$ is denoted $|\mathcal{K}|$, and the set of all probability mass functions (PMFs) on it $\mathcal{P}(\mathcal{K})$.

Mod-$|\mathcal{A}|$ addition "$\oplus$" and mod-$|\mathcal{A}|$ substraction "$\ominus$" are extended to n-tuples, which are usually designated in boldface, componentwise:

$$\mathbf{x} \oplus \mathbf{y} = (x_1 \oplus y_1, \dots, x_n \oplus y_n) \tag{4}$$

Manuscript received 8 November 2023; accepted 29 February 2024. Date of publication 13 March 2024; date of current version 18 June 2024. This work was supported by the Swiss National Science Foundation (SNSF) under Grant 200021-215090. An earlier version of this paper was presented in part at the 2023 International Symposium on Information Theory (ISIT) [DOI: 10.1109/ISIT54713.2023.10206953]. *(Corresponding author: Yiming Yan.)*

The authors are with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch; yan@isi.ee.ethz.ch).

Communicated by V. Kostina, Associate Editor for Shannon Theory and Information Measures.

Digital Object Identifier 10.1109/TIT.2024.3377601

[1] Throughout this paper, "rare-error capacity" and "rare-error feedback capacity" refer to the supremum of the achievable rates, in the sense that the probability of error tends to zero as the blocklength tends to infinity [2]. We refrain from calling it Shannon capacity lest it be confused with the Shannon capacity of a graph.

and likewise for $\mathbf{x} \ominus \mathbf{y}$. If $\mathcal{B} \subseteq \mathcal{A}^n$ is a set of n-tuples, then $\mathcal{B}^*$ denotes $\mathcal{B} \setminus \{\mathbf{0}\}$, i.e., $\mathcal{B}$ without the all-zero n-tuple. For $\mathcal{B}, \mathcal{B}' \subseteq \mathcal{A}^n$, we denote the sumset and the difference set by

$$\mathcal{B} \oplus \mathcal{B}' = \bigl\{\mathbf{b} \oplus \mathbf{b}' \colon \mathbf{b} \in \mathcal{B},\, \mathbf{b}' \in \mathcal{B}'\bigr\} \tag{5}$$

$$\mathcal{B} \ominus \mathcal{B}' = \bigl\{\mathbf{b} \ominus \mathbf{b}' \colon \mathbf{b} \in \mathcal{B},\, \mathbf{b}' \in \mathcal{B}'\bigr\}; \tag{6}$$

and for $\mathbf{x} \in \mathcal{A}^n$, we write $\mathbf{x} \oplus \mathcal{B}$ and $\mathbf{x} \ominus \mathcal{B}$ for $\{\mathbf{x}\} \oplus \mathcal{B}$ and $\{\mathbf{x}\} \ominus \mathcal{B}$. We use $\{\xi\}^+$ to denote $\max\{0, \xi\}$.

B. Definitions and Preliminaries

A blocklength-n code for a (general) discrete memoryless channel (DMC) $Q_{Y|X}(\cdot|\cdot)$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, consists of a message set $\mathcal{M} = \bigl\{1, 2, \dots, |\mathcal{M}|\bigr\}$ and an encoding function $f \colon \mathcal{M} \to \mathcal{X}^n$, $m \mapsto \mathbf{x}(m) = (x_1(m), \dots, x_n(m))$. Since the codewords $\mathbf{x}(1), \dots, \mathbf{x}(|\mathcal{M}|)$ need not be distinct, the codebook $\mathcal{C} = \bigl\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(|\mathcal{M}|)\bigr\}$ is a multiset (i.e., an unordered collection of elements that may repeat). Its cardinality is $|\mathcal{M}|$. A sequence of codes, indexed by the blocklength n, is said to have transmission rate $\liminf_{n\to\infty} \frac{1}{n} \log |\mathcal{M}|$.

The *zero-error capacity* C_0 [1] is the supremum of rates R for which there exists a sequence of rate-R codes, indexed by the blocklength, that fulfill the zero-error requirement that to every output sequence $\mathbf{y} \in \mathcal{Y}^n$ there correspond at most one compatible message, i.e., a message $m \in \mathcal{M}$ for which

$$Q^n_{Y|X}(\mathbf{y}|\mathbf{x}(m)) > 0. \tag{7}$$

Here $Q^n_{Y|X}(\mathbf{y}|\mathbf{x})$ stands for $\prod_{i=1}^n Q_{Y|X}(y_i|x_i)$.

A necessary and sufficient condition for C_0 to be positive is that there exist channel inputs $x, x' \in \mathcal{X}$ such that $Q_{Y|X}(y|x) \cdot Q_{Y|X}(y|x') = 0$ for every $y \in \mathcal{Y}$ [1]. This characterization can be used, for example, to conclude that C_0 is zero for the Triangle channel. It also shows that, whenever C_0 is positive, we can transmit a bit by using the channel once (with the input x or x'). Consequently, C_0 cannot be positive yet smaller than one. As we shall see, this is not the case in the presence of help (Remark 3).

Determining the zero-error capacity for general DMCs is an open combinatorial problem and is one of the holy grails of information theory. It is known for some specific channels, including the Pentagon channel: Shannon showed that $\frac{1}{2}\log 5 \leq C_0 \leq \log \frac{5}{2}$ in 1959 [1], and Lovász proved, using algebraic graph theory, that the lower bound is tight in 1979 [11]. The zero-error capacity of the 7/2 channel is to date unknown.

The problem is greatly simplified if the time-i channel input may depend not only on the message m but also on the past channel outputs y^{i-1} that are revealed to the encoder via a feedback link from the channel output to the encoder. A blocklength-n encoder now consists of n functions $f_i \colon \mathcal{M} \times \mathcal{Y}^{i-1} \to \mathcal{X}$, $(m, y^{i-1}) \mapsto x_i(m, y^{i-1})$, one for each $i \in [n]$, and the *zero-error feedback capacity* $C_{0\mathrm{F}}$ is defined like C_0 except that $\mathbf{x}(m)$ in (7) is replaced by $\mathbf{x}(m, \mathbf{y}) = (x_1(m), x_2(m, y_1), \dots, x_n(m, y^{n-1}))$.[2] Since the encoder may ignore the feedback link,

$$C_{0\mathrm{F}} \geq C_0. \tag{8}$$

The zero-error feedback capacity $C_{0\mathrm{F}}$ was determined by Shannon:

Theorem 1 ([1]): On a DMC, if $C_0 = 0$, then the zero-error feedback capacity $C_{0\mathrm{F}}$ is also zero. Else, $C_{0\mathrm{F}} = -\log \pi_0$, where

$$\pi_0 = \min_{P \in \mathcal{P}(\mathcal{X})} \max_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}_y} P(x), \tag{9}$$

and $\mathcal{X}_y$ comprises the inputs that can induce the output letter y with positive probability:

$$\mathcal{X}_y = \{x \in \mathcal{X} \colon Q_{Y|X}(y|x) > 0\}. \tag{10}$$

Note that, since $C_{0\mathrm{F}} > 0$ iff $C_0 > 0$, and since $C_{0\mathrm{F}} \geq C_0$, also $C_{0\mathrm{F}}$ cannot be positive yet strictly smaller than one. We shall see that this is not true in the presence of zero-rate help (Remark 3).

Applying Theorem 1 to the MMANC yields the following corollary.

Corollary 1: On the MMANC, if $C_0 = 0$, then the zero error feedback capacity $C_{0\mathrm{F}}$ is also zero. Else,

$$C_{0\mathrm{F}} = \log |\mathcal{A}| - \log |\mathcal{S}|. \tag{11}$$

Proof: We can lower-bound π_0 by lower bounding the maximum over $y \in \mathcal{A}$ by the average:

$$\pi_0 = \min_{P \in \mathcal{P}(\mathcal{A})} \max_{y \in \mathcal{A}} \sum_{x \in \mathcal{X}_y} P(x) \tag{12}$$

$$\geq \frac{1}{|\mathcal{A}|} \min_{P \in \mathcal{P}(\mathcal{A})} \sum_{y \in \mathcal{A}} \sum_{x \in \mathcal{X}_y} P(x) \tag{13}$$

$$= \frac{|\mathcal{S}|}{|\mathcal{A}|} \min_{P \in \mathcal{P}(\mathcal{A})} \sum_{x \in \mathcal{A}} P(x) \tag{14}$$

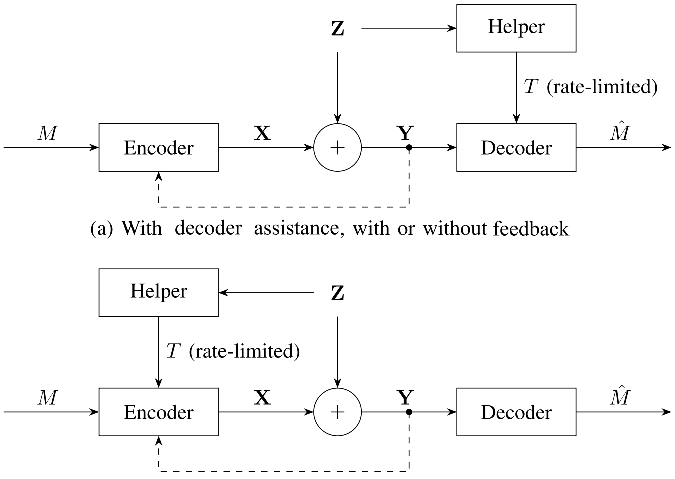
$$= \frac{|\mathcal{S}|}{|\mathcal{A}|}, \tag{15}$$

where in (13) we lower-bounded the maximum over $\mathcal{A}$ by the arithmetic average; and (14) holds since in the double sum, each $x \in \mathcal{A}$ is contained in $\mathcal{X}_y$ for exactly $|\mathcal{S}|$ different y's in $\mathcal{A}$. The corollary follows by noting that this lower bound is tight as can be seen by considering P equiprobable. ■

Henceforth, we focus on MMANCs. A helper is an altruistic party that has no message to send and only wishes to assist the transmission. To do so, it observes the noise sequence $\mathbf{Z} = Z^n$ (noncausally); it produces a rate-limited description T of it; and it reveals the description to the encoder, or to the decoder, or to both. It is incognizant of the transmitted message. More formally, a blocklength-n helper, represented by the helping function $h \colon \mathcal{A}^n \to \mathcal{T}$, observes the noise sequence $\mathbf{Z}$ and describes it as $T = h(\mathbf{Z})$, with T taking values in a finite set $\mathcal{T}$. For a given sequence of coding schemes, the help rate R_h is 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 description T is revealed to the decoder, as in Fig. 1a. In this scenario we use $C_{0,\mathrm{dec}}(R_\mathrm{h})$ to denote the supremum of rates R

[2] Also in the presence of feedback, the blocklength n is deterministic: we do not consider coding schemes with random transmission durations.



(a) With decoder assistance, with or without feedback

(b) With encoder assistance, with or without feedback

Fig. 1. Modulo-additive noise channels.

for which there exists a sequence of zero-error coding schemes (without feedback) with transmission rate at least R, and with help rate no larger than R_{h}. By zero-error we now mean that for any $\mathbf{y} \in \mathcal{A}^n$ and $t \in \mathcal{T}$, at most one message m is compatible with $(\mathbf{y}, t)$ in the sense that

$$Q^n_{Y|X}(\mathbf{y}|\mathbf{x}(m)) > 0 \text{ and } h(\mathbf{y} \ominus \mathbf{x}(m)) = t. \tag{16}$$

In the presence of feedback, we denote the analogous capacity $C_{\text{0F,dec}}(R_{\text{h}})$: we merely replace $\mathbf{x}(m)$ with $\mathbf{x}(m, \mathbf{y})$ in (16).

Encoder assistance corresponds to the scenario where T is revealed noncausally to the encoder, as in Fig. 1b. In the absence of feedback, the encoding function is $f\colon \mathcal{M} \times \mathcal{T} \to \mathcal{A}^n$, $(m, t) \mapsto \mathbf{x}(m, t)$, and $C_{\text{0,enc}}(R_{\text{h}})$ is defined with the requirement that to every $\mathbf{y} \in \mathcal{A}^n$ there correspond at most one compatible message m in the sense that[3]

$$\exists t \in \mathcal{T} \text{ s.t. } Q^n_{Y|X}(\mathbf{y}|\mathbf{x}(m,t)) > 0 \text{ and } h(\mathbf{y} \ominus \mathbf{x}(m,t)) = t. \tag{17}$$

With feedback, the encoder employs functions $f_i\colon \mathcal{M} \times \mathcal{T} \times \mathcal{A}^{i-1} \to \mathcal{A}$, $(m, t, y^{i-1}) \mapsto x_i(m, t, y^{i-1})$ for $i \in [n]$, and $C_{\text{0F,enc}}(R_{\text{h}})$ is defined analogously by replacing $\mathbf{x}(m, t)$ with $\mathbf{x}(m, t, \mathbf{y}) = (x_1(m, t), x_2(m, t, y_1), \ldots, x_n(m, t, y^{n-1}))$ in (17).

Assumption 1: We shall assume throughout that

$$|\mathcal{S}| > 1 \tag{18}$$

and

$$R_{\text{h}} \leq \log |\mathcal{S}|. \tag{19}$$

If $|\mathcal{S}| = 1$, then the noise is deterministic and even without feedback or help, the zero-error capacity is $\log |\mathcal{A}|$. And, for any sequence of helpers $h\colon \mathcal{A}^n \to \mathcal{T}$ of rate $R_{\text{h}} > \log |\mathcal{S}|$, one can construct another sequence of helpers h' of rate $R'_{\text{h}} = \log |\mathcal{S}|$ with $h'(\mathbf{z}) = h(\mathbf{z})$ for all $\mathbf{z} \in \mathcal{S}^n$, so that $h(\mathbf{Z})$ and $h'(\mathbf{Z})$ are identical with probability one.

In this paper, we present results on $C_{\text{0,dec}}(R_{\text{h}})$ and $C_{\text{0,enc}}(R_{\text{h}})$, the zero-error capacity in the presence of decoder or encoder assistance, and on $C_{\text{0F,dec}}(R_{\text{h}})$ and $C_{\text{0F,enc}}(R_{\text{h}})$, the analogous quantities in the presence of feedback.

C. Related Work

Related to our work is the following theorem [12] on the case where the noise is of full support:

Theorem 2 (MMANC With Noise of Full Support [12]): On the MMANC with rate-R_{h} decoder or encoder assistance, if $\mathcal{S} = \mathcal{A}$, then

$$C_{\text{0,dec}}(R_{\text{h}}) = C_{\text{0,enc}}(R_{\text{h}}) = R_{\text{h}}, \tag{20a}$$

and in particular,

$$C_{\text{0,dec}}(0) = 0 \tag{20b}$$

and

$$C_{\text{0,enc}}(0) = 0. \tag{20c}$$

Our present work extends this result by studying the general case where the noise need not be of full support. As we shall see ahead (Corollary 3), the condition $\mathcal{S} = \mathcal{A}$ is also necessary for (20b) to hold, and likewise for (20c). We also study the effect of feedback on the zero-error helper capacity.

Also related to our results is the work of Merhav on error exponents [8]. To see the relevance, note that on a DMC, the Reliability Function $E(R)$ equals infinity iff R can be achieved with zero error. The intuition is the following. Let the transition matrix be $Q(y|x)$ and the input sequence be x^n, then all output sequences of positive probability have probability at least α^n, where

$$\alpha = \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}\colon Q(y|x) > 0} Q(y|x).$$

Consequently,

$$\Big(\Pr(\text{error}|M = m) > 0\Big) \implies \Big(\Pr(\text{error}|M = m) \geq \alpha^n\Big), \tag{21}$$

whose contrapositive can be rewritten as

$$\Big(-\frac{1}{n} \log \Pr(\text{error}|M = m) > \log \frac{1}{\alpha}\Big) \implies \Big(\Pr(\text{error}|M = m) = 0\Big). \tag{22}$$

For our MMANC, Merhav [8, Eq. (57)] derived an upper bound on the Reliability Function for $R < \log |\mathcal{A}|$,

$$E(R) \leq \min_{\tilde{Q}_Z\colon H(\tilde{Q}_Z) > \log |\mathcal{A}| + R_{\text{h}} - R} D(\tilde{Q}_Z \| Q_Z). \tag{23}$$

This upper bound is finite iff $R > \log |\mathcal{A}| - \log |\mathcal{S}| + R_{\text{h}}$: only in this range of rates, there exists a PMF $\tilde{Q}_Z$ that is feasible in the minimization and satisfies $\operatorname{supp}(\tilde{Q}_Z) \subseteq \mathcal{S}$. This implies an upper bound on the zero-error capacity

$$C_0(R_{\text{h}}) \leq \log |\mathcal{A}| - \log |\mathcal{S}| + R_{\text{h}}. \tag{24}$$

When $|\mathcal{A}|$ is a prime, our results (Theorem 4) show that this upper bound is tight.

III. Main Results

A. Feedback Link Present

The following theorem addresses the zero-error helper capacity of the MMANC in the presence of a feedback link.

Theorem 3 (Assistance and Feedback): On the MMANC with feedback and rate-R_{h} decoder or encoder assistance,

$$C_{\text{0F,dec}}(R_{\text{h}}) = C_{\text{0F,enc}}(R_{\text{h}}) = \log |\mathcal{A}| - \log |\mathcal{S}| + R_{\text{h}}. \tag{25}$$

[3] This condition is equivalent to $Q_{\mathbf{Y}|M}(\mathbf{y}|m) > 0$, where $Q_{\mathbf{Y}|M}(\mathbf{y}|m) = \sum_{t \in \mathcal{T}} Q_T(t) Q_{\mathbf{Y}|\mathbf{X},T}(\mathbf{y}|\mathbf{x}(m,t), t)$.

B. Feedback Link Absent

When the cardinality $|\mathcal{A}|$ of the alphabet $\mathcal{A}$ is prime, the zero-error helper capacity without feedback is determined in the following theorem:

Theorem 4 (Assistance Without Feedback: Prime Cardinality): On the MMANC with rate-R_h decoder or encoder assistance, if $|\mathcal{A}|$ is prime, then

$$C_{0,\text{dec}}(R_\text{h}) = C_{0,\text{enc}}(R_\text{h}) = \log|\mathcal{A}| - \log|\mathcal{S}| + R_\text{h}. \tag{26}$$

When $|\mathcal{A}|$ is not necessarily a prime, we provide the following upper and lower bounds:

Theorem 5 (Assistance Without Feedback: General Cardinality): On the MMANC with rate-R_h decoder assistance,

$$C_{0,\text{dec}}(R_\text{h}) \le \log|\mathcal{A}| - \log|\mathcal{S}| + R_\text{h} \tag{27}$$

and

$$C_{0,\text{dec}}(R_\text{h}) \ge \frac{\log|\mathcal{A}|}{\log|\mathcal{S}|}\cdot R_\text{h} + \left(1 - \frac{R_\text{h}}{\log|\mathcal{S}|}\right)\cdot \max\left\{C_0,\ \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|}\right\}. \tag{28}$$

These bounds also hold for $C_{0,\text{enc}}(R_\text{h})$.

Theorem 5 has two corollaries. The first characterizes the zero-error helper capacity when the noise support "tessellates" the alphabet $\mathcal{A}$ and thus strengthens Theorem 2.

Corollary 2 (Special MMANCs): If $C_0 = \log|\mathcal{A}| - \log|\mathcal{S}|$, then

$$C_{0,\text{dec}}(R_\text{h}) = C_{0,\text{enc}}(R_\text{h}) = \log|\mathcal{A}| - \log|\mathcal{S}| + R_\text{h}. \tag{29}$$

Proof of Corollary 2: Follows from Theorem 5 by substituting $\log|\mathcal{A}| - \log|\mathcal{S}|$ for C_0 on the RHS of (28) and noting that the result matches the RHS of (27). ■

When $\mathcal{S} = \mathcal{A}$, which implies that $C_0 = 0$ and hence that $C_0 = \log|\mathcal{A}| - \log|\mathcal{S}|$, the corollary recovers Theorem 2. But see Corollary 3 ahead for a stronger statement.

For another application of this corollary, consider the MMANCs with $|\mathcal{A}| = 4$ and with $\mathcal{S} = \{0,1\}$ or $\mathcal{S} = \{0,2\}$ (so $|\mathcal{S}| = 2$). In both cases $C_0 = 1$ $(= \log|\mathcal{A}| - \log|\mathcal{S}|)$, so the corollary yields that $C_{0,\text{dec}}(R_\text{h}) = C_{0,\text{enc}}(R_\text{h}) = 1 + R_\text{h}$.

The second corollary to Theorem 5 provides a necessary and sufficient condition for the positivity of the zero-error helper capacity.

Corollary 3 (Positivity): The following statements are equivalent:

i) $C_{0,\text{dec}}(R_\text{h}) = 0$;
ii) $C_{0,\text{enc}}(R_\text{h}) = 0$;
iii) $C_{0\text{F},\text{dec}}(R_\text{h}) = 0$;
iv) $C_{0\text{F},\text{enc}}(R_\text{h}) = 0$;
v) $R_\text{h} = 0$ and $\mathcal{S} = \mathcal{A}$.

Proof of Corollary 3: The equivalence of iii), iv), and v) follows from Theorem 3 on feedback. The implications v) $\Longrightarrow$ i) and v) $\Longrightarrow$ ii) follow from (27) and the analogous result for $C_{0,\text{enc}}(R_\text{h})$; the implication i) $\Longrightarrow$ v) follows from (28) as follows:

$$C_{0,\text{dec}}(R_\text{h}) \ge \frac{\log|\mathcal{A}|}{\log|\mathcal{S}|}\cdot R_\text{h} + \left(1 - \frac{R_\text{h}}{\log|\mathcal{S}|}\right)\cdot\frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|} \tag{30}$$

$$= \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|} + \left(\frac{\log|\mathcal{A}|}{2\log|\mathcal{S}|} + \frac{1}{2}\right)\cdot R_\text{h} \tag{31}$$

$$\ge \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|} + R_\text{h}; \tag{32}$$

the proof that ii) $\Longrightarrow$ v) is similar. ■

The above theorems and corollaries have some noteworthy implications:

Remark 1 (Benifit of Assistance): Assistance can increase the zero-error capacity by more than its rate. Even zero-rate assistance can increase the zero-error capacity: on the Pentagon channel, it raises the zero-error capacity from Lovász's $\frac{1}{2}\log 5$ to $\log\frac{5}{2}$, i.e., to $C_{0\text{F}}$ (Corollary 1); on the Triangle channel, it raises the zero-error capacity from zero to $\log\frac{3}{2}$, which even exceeds $C_{0\text{F}}$ (the latter being zero for this channel).

Remark 2: Thanks to Theorem 4, the zero-error capacity with a helper can sometimes be determined even if it is unknown in the absence of help, e.g., for the Heptagon channel.

Remark 3 (Less Than One Bit): As on the Gel'fand-Pinsker channel with feedback [13], in all cases (with or without feedback, and with decoder or encoder assistance), the zero-error capacity can be positive yet smaller than 1 bit. This is not the case in the absence of assistance.

IV. Feedback Link Present

In this section, we study the zero-error feedback capacity with helper and establish Theorem 3; see Fig. 1a and 1b with the feedback link. To this end, we need the following lemma, stating that feedback does not increase the helper rare-error capacity on the MMANC.

Lemma 1: On the MMANC with feedback and rate-R_h decoder or encoder assistance, the rare-error capacities are given by

$$C_{\text{F,dec}}(R_\text{h}) = C_{\text{F,enc}}(R_\text{h}) = \log|\mathcal{A}| - \left\{H(Q_Z) - R_\text{h}\right\}^+. \tag{33}$$

Proof: In light of [3, Theorem 12] and [4, Theorem 8], which establish that the RHS of (33) can be achieved without feedback, we only need to prove a converse. To that end, we prove the stronger claim that—even if the description T is presented to both encoder and decoder—the rare-error feedback capacity does not exceed the RHS of (33). We assume $R_\text{h} \le H(Q_Z)$, because otherwise the result is obvious.

Consider a message M that is drawn equiprobably from the message set $\mathcal{M}$. For any sequence of coding schemes of rate R with rate-R_h assistance and vanishing probabilities of error,

$$\log|\mathcal{M}| = H(M) \tag{34}$$

$$= I(M;\mathbf{Y},T) + H(M|\mathbf{Y},T) \tag{35}$$

$$\le I(M;\mathbf{Y},T) + n\delta_n \tag{36}$$

$$= I(M;\mathbf{Y}|T) + n\delta_n \tag{37}$$

$$= H(\mathbf{Y}|T) - H(\mathbf{Y}|M,T) + n\delta_n \tag{38}$$

$$\le H(\mathbf{Y}) - H(\mathbf{Y}|M,T) + n\delta_n \tag{39}$$

$$\le H(\mathbf{Y}) - H(\mathbf{Z}|M,T) + n\delta_n \tag{40}$$

$$= H(\mathbf{Y}) - H(\mathbf{Z}|T) + n\delta_n \tag{41}$$

$$= H(\mathbf{Y}) - H(\mathbf{Z}) + I(\mathbf{Z};T) + n\delta_n \tag{42}$$

$$\begin{aligned} &\le H(\mathbf{Y}) - H(\mathbf{Z}) + \log|\mathcal{T}| + n\delta_n && (43)\\ &\le n\log|\mathcal{A}| - nH(Q_Z) + \log|\mathcal{T}| + n\delta_n, && (44)\end{aligned}$$

where (36) holds for some $\{\delta_n\}$ tending to zero by Fano's inequality; (37) and (41) hold because T is a function of $\mathbf{Z}$, so $(\mathbf{Z}, T)$ is independent of M; and (40) holds because, in the presence of feedback and help, $\mathbf{Z}$ is a function of $(\mathbf{Y}, M, T)$ namely $Z_i = Y_i \ominus f_i(M, T, Y^{i-1})$ for $i \in [n]$. Dividing the inequalities by n and letting n tend to infinity establishes the converse. ■

Proof of Theorem 3: We first establish the converse for decoder assistance. If $\tilde{Q}_{Y|X}$ is any auxiliary MMANC over $\mathcal{A}$ of noise PMF $\tilde{Q}_Z \in \mathcal{P}(\mathcal{A})$ that is absolutely continuous with respect to Q_Z (i.e., whose support is contained in $\mathcal{S}$, denoted by $\tilde{Q}_Z \ll Q_Z$), then its rare-error feedback capacity with decoder assistance $\tilde{C}_{\text{F,dec}}(R_\text{h})$ forms an upper bound on $C_{\text{0F,dec}}(R_\text{h})$, because any error-free coding scheme for the original channel is also error-free on the auxiliary channel. Indeed, for any $\mathbf{y} \in \mathcal{A}^n$ and $t \in \mathcal{T}$, the absolute continuity hypothesis implies that

$$\Big(\tilde{Q}^n_{Y|X}(\mathbf{y}|\mathbf{x}(m,\mathbf{y})) > 0\Big) \implies \Big(Q^n_{Y|X}(\mathbf{y}|\mathbf{x}(m,\mathbf{y})) > 0\Big) \tag{45}$$

so if a message m is compatible with $(\mathbf{y}, t)$ on the auxiliary channel (in the sense that $\tilde{Q}^n_{Y|X}(\mathbf{y}|\mathbf{x}(m,\mathbf{y})) > 0$ and $h(\mathbf{y} \ominus \mathbf{x}(m,\mathbf{y})) = t$), then it is also compatible with $(\mathbf{y}, t)$ on the original channel.

Therefore, upon minimizing over the choice of $\tilde{Q}_{Y|X}$ to get the tightest bound,

$$\begin{aligned} C_{\text{0F,dec}}(R_\text{h}) &\le \min_{\tilde{Q}_Z:\, \tilde{Q}_Z \ll Q_Z} \tilde{C}_{\text{F,dec}}(R_\text{h}) && (46)\\ &= \min_{\tilde{Q}_Z:\, \tilde{Q}_Z \ll Q_Z} \Big\{\log|\mathcal{A}| - \big\{H(\tilde{Q}_Z) - R_\text{h}\big\}^+\Big\} && (47)\\ &= \log|\mathcal{A}| - \big\{\log|\mathcal{S}| - R_\text{h}\big\}^+, && (48)\end{aligned}$$

where (47) follows from Lemma 1, and (48) holds because, subject to a support constraint, the uniform PMF maximizes entropy.

Similar arguments apply also to encoder assistance.

We now turn to the direct part.

- Case 1: $R_\text{h} \ge \log|\mathcal{S}|$. In this case feedback is unnecessary. The codebook comprises all the distinct sequences in $\mathcal{A}^n$. Using $\lceil n\log|\mathcal{S}|\rceil$ bits, the helper describes the noise sequence $\mathbf{Z}$ precisely. The decoder (resp. encoder) subtracts the noise from the received sequence (resp. from the codeword to be transmitted), so the codeword and the message can be received error-free. This establishes the achievability of $\log|\mathcal{A}|$ bits per channel use.
- Case 2: $R_\text{h} = 0$. A two-phase coding scheme is proposed. In Phase 1, we follow the construction (for a uniform input distribution) in Shannon's proof of Theorem 1 in [1], where the encoder sequentially reduces the decoder's ambiguity. In the i-th channel use, thanks to the feedback, the encoder reconstructs the list of messages compatible with Y^{i-1} and evenly assigns them to different input symbols (in a manner agreed upon with the decoder prior to transmission). Only $|\mathcal{S}|$ of the $|\mathcal{A}|$ input symbols are compatible with Y_i, and the number of compatible messages is reduced by a factor of roughly $\frac{|\mathcal{S}|}{|\mathcal{A}|}$. More precisely, Shannon showed that if $|\mathcal{M}| = \big\lfloor \big(\frac{|\mathcal{S}|}{|\mathcal{A}|}\big)^{-n}\big\rfloor$, then after n channel uses, the number of compatible messages is at most $|\mathcal{A}|^2$. The final ambiguity is removed in Phase 2, where the helper comes into play. Since the messages that are compatible with the outputs from Phase 1 are known to the encoder, and since their number does not exceed $|\mathcal{A}|^2$, the encoder can inform the decoder which compatible message was sent in two additional clean channel uses. To clean these two channel uses, the helper informs the decoder (resp. encoder) of the exact value of $Z_{n+1}^{n+2} \in \mathcal{S}^2$ and the decoder (resp. encoder) subtracts the noise after (resp. before) the transmission. The rate of help is therefore

$$\lim_{n\to\infty} \frac{1}{n+2}\log|\mathcal{S}|^2 = 0 \tag{49}$$

and the transmission rate

$$\lim_{n\to\infty}\frac{1}{n+2}\log|\mathcal{M}| = \lim_{n\to\infty}\frac{\log\big\lfloor\big(\frac{|\mathcal{S}|}{|\mathcal{A}|}\big)^{-n}\big\rfloor}{n+2} = \log\frac{|\mathcal{A}|}{|\mathcal{S}|}. \tag{50}$$

- Case 3: $0 < R_\text{h} < \log|\mathcal{S}|$. We divide the transmission block into two parts of relative length $\frac{R_\text{h}}{\log|\mathcal{S}|}$ and $1 - \frac{R_\text{h}}{\log|\mathcal{S}|}$. We then apply the aforementioned coding schemes for helper rates of $\log|\mathcal{S}|$ and zero, respectively. The total rate achieved by this time-sharing scheme is

$$\begin{aligned} &\frac{R_\text{h}\log|\mathcal{A}|}{\log|\mathcal{S}|} + \bigg(1 - \frac{R_\text{h}}{\log|\mathcal{S}|}\bigg)\big(\log|\mathcal{A}| - \log|\mathcal{S}|\big)\\ &\quad = \log|\mathcal{A}| - \log|\mathcal{S}| + R_\text{h}. && (51)\end{aligned}$$

■

V. Feedback Link Absent

In this section, we provide proofs pertaining to the zero-error helper capacity in the absence of feedback; see Fig. 1a and 1b without the feedback link.

A. Prime Cardinality

We begin with the case where $|\mathcal{A}|$ is a prime, which we denote p. We denote the cardinality-p finite field $\mathbb{F}_p$ and identify it with the set $\mathbb{Z}_p = \{0, \ldots, p-1\}$ with mod-p arithmetic.

Proof of Theorem 4: Since feedback cannot hurt, it follows from Theorem 3 that we only need to prove the direct part. This is trivial unless $|\mathcal{S}| < |\mathcal{A}|$, which we proceed to assume. We first focus on decoder assistance.

- Case 1: $R_\text{h} \ge \log|\mathcal{S}|$. The achievability in this case is as in the proof of Theorem 3, where the feedback link is ignored.
- Case 2: $R_\text{h} = 0$. We will construct a sequence of blocklength-n codebooks of rate $\big(\log\frac{|\mathcal{A}|}{|\mathcal{S}|} - \epsilon_n\big)$ that can be decoded error-free utilizing rate-ϵ'_n decoder assistance, for some $\{\epsilon_n\}$ and $\{\epsilon'_n\}$ tending to zero.

The codes we construct have two key properties. The first is that they are L-*list-decodable* [14], [15], [16] where $\mathsf{L} \in \mathbb{Z}^+$ grows subexponentially with n. That is, every $\mathbf{y} \in \mathcal{A}^n$ is compatible with at most L messages. This guarantees that the decoder's ambiguity could be eliminated with a sublinear number of bits. Elias [14] established the existence such codebooks of rate $\log \frac{|\mathcal{A}|}{|\mathcal{S}|} - \Theta(\mathrm{L}^{-1})$. But this is not enough, because, in the absence of feedback, neither the transmitter nor the helper can determine the list facing the decoder. This is where the second property comes in: To overcome this issue and enable the helper to remove the ambiguity, we shall impose a linear structure on the code, and this is where the assumption that $|\mathcal{A}|$ is a prime will be essential: it will allow us to view $\mathcal{A}$ as a field.

The existence of *linear* L-list-decodable codes can be established using a variation on a theme by Elias [14] using tools that were used successfully in the analysis of random linear codes (e.g. [17], [18], [19]).

Lemma 2: Consider a MMANC with $|\mathcal{A}| = p$, where p is prime. Given $\mathrm{L} \in \mathbb{Z}^+$, define

$$R_{\mathrm{L}} = \max\left\{0,\, \log \frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{\log^2 |\mathcal{A}|}{\log(\mathrm{L}+1)}\right\}. \tag{52}$$

Then, for any $n \in \mathbb{Z}^+$, there exists a blocklength-n linear code over the field $\mathbb{F}_p$ of rate $\frac{\log |\mathcal{A}|}{n} \lfloor \frac{n R_{\mathrm{L}}}{\log |\mathcal{A}|} \rfloor$ that is L-list-decodable.

Proof: Assume $R_{\mathrm{L}} > 0$ (because otherwise there is nothing to prove). Given some blocklength n, let the message set be $\mathcal{M} = \mathbb{F}_p^k$, with $k \in \mathbb{Z}^+$ to be specified later. A generic message $\mathbf{m} \in \mathcal{M}$ is thus represented by a k-vector, and the transmission rate is k/n in base-p logarithm, or $(k/n)\log_2 p$ bits per channel use.

Pick a random $(n \times k)$-matrix $\mathbb{A}$ whose entries are drawn IID equiprobably from $\mathbb{F}_p$, and consider the encoding rule $\mathbf{m} \mapsto \mathbf{X}(\mathbf{m}) = \mathbb{A}\mathbf{m}$. Let $\mathcal{C}$ be the random linear code (multiset) it induces. This encoding rule maps any ℓ linearly independent messages to independent codewords, each having IID equiprobable random components.

Among any $(\mathrm{L}+1)$ messages, at least $\ell \triangleq \lceil \log_p(\mathrm{L}+1) \rceil$ are linearly independent, so the probability that there exists some $\mathbf{y} \in \mathcal{A}^n$ compatible with $(\mathrm{L}+1)$ messages is upper bounded by the probability that there exists some $\mathbf{y} \in \mathcal{A}^n$ that is compatible with ℓ linearly independent messages. The latter, by the Union Bound, is strictly smaller than

$$\begin{aligned}
&\sum_{\mathbf{y} \in \mathcal{A}^n} \sum_{\substack{\mathbf{m}_1,\ldots,\mathbf{m}_\ell \in \mathcal{M} \\ \text{linearly independent}}} \Pr \bigcap_{i=1}^{\ell} \bigcap_{j=1}^{n} \left[X_j(\mathbf{m}_i) \in \mathcal{X}_{y_j}\right] \\
&= \sum_{\mathbf{y} \in \mathcal{A}^n} \sum_{\substack{\mathbf{m}_1,\ldots,\mathbf{m}_\ell \in \mathcal{M} \\ \text{linearly independent}}} \left(\frac{|\mathcal{S}|^n}{|\mathcal{A}|^n}\right)^{\ell} && (53) \\
&\le |\mathcal{A}|^n \binom{|\mathcal{M}|}{\ell} \left(\frac{|\mathcal{S}|}{|\mathcal{A}|}\right)^{n\ell} && (54) \\
&\le |\mathcal{A}|^n p^{k\ell} \left(\frac{|\mathcal{S}|}{|\mathcal{A}|}\right)^{n\ell} && (55) \\
&\le \left(|\mathcal{A}|\, p^{\left(\ell \log_p \frac{|\mathcal{A}|}{|\mathcal{S}|} - 1\right)} \left(\frac{|\mathcal{S}|}{|\mathcal{A}|}\right)^{\ell}\right)^{n} && (56) \\
&= 1, && (57)
\end{aligned}$$

where (53) holds because—when $\mathbf{m}_1, \ldots, \mathbf{m}_\ell$ are linearly independent—the codewords $\mathbf{X}(\mathbf{m}_1), \ldots, \mathbf{X}(\mathbf{m}_\ell)$ are independent, each having IID equiprobably distributed components, and because $|\mathcal{X}_y| = |\mathcal{S}|$ for every $y \in \mathcal{A}$; and in (56) we choose $k = \lfloor n\big(\log_p \frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{1}{\log_p(\mathrm{L}+1)}\big)\rfloor$, so

$$\begin{aligned}
k\ell &\le n\Big(\log_p \frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{1}{\log_p(\mathrm{L}+1)}\Big)\ell && (58) \\
&\le n\Big(\ell \log_p \frac{|\mathcal{A}|}{|\mathcal{S}|} - 1\Big). && (59)
\end{aligned}$$

Hence, with positive probability, the random linear code is L-list-decodable. The lemma then follows by noting that, in bits, the rate is $\frac{k}{n} \cdot \log |\mathcal{A}|$. ■

We now use Lemma 2 to complete the proof of Theorem 4 for the case of $R_{\mathrm{h}} = 0$. Let $\{\mathrm{L}_n\}$ be a sequence of positive integers tending to infinity subexponentially, e.g., $\mathrm{L}_n = \Theta(n)$. The lemma implies that, for every blocklength n, there exists a linear code $\mathcal{C}_n$ of rate $\frac{\log|\mathcal{A}|}{n} \lfloor \frac{n R_{\mathrm{L}_n}}{\log|\mathcal{A}|} \rfloor$ that is L_n-list-decodable. A minor annoyance is that the codewords in $\mathcal{C}_n$ need not be distinct. To overcome this, we consider the code $\mathcal{C}'_n \subseteq \mathbb{F}_p^n$ comprising all the distinct elements in $\mathcal{C}_n$. Note that (i) $\mathcal{C}'_n$ is a subgroup of $\mathbb{F}_p^n$; (ii) $\mathcal{C}'_n$ is L_n-list-decodable; and (iii)

$$|\mathcal{C}'_n| \ge \frac{|\mathcal{C}_n|}{\mathrm{L}_n} \tag{60}$$

(because $\mathcal{C}_n$, being L_n-list-decodable, contains no codeword more than L_n times). This latter property and the fact that $\{\mathrm{L}_n\}$ is subexponential imply that $\{\mathcal{C}'_n\}$ has the desired rate:

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{C}'_n| &= \lim_{n\to\infty} \frac{1}{n} \log |\mathcal{C}_n| && (61) \\
&= \lim_{n\to\infty} R_{\mathrm{L}_n} && (62) \\
&= \lim_{n\to\infty} \log \frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{\log^2 |\mathcal{A}|}{\log(\mathrm{L}_n + 1)} && (63) \\
&= \log \frac{|\mathcal{A}|}{|\mathcal{S}|}, && (64)
\end{aligned}$$

where (61) follows from (60) and the fact that $\{\mathrm{L}_n\}$ is subexponential; and (64) holds because $\{\mathrm{L}_n\}$ tends to infinity.

We next show that—although the helper is incognizant of the list of messages that are compatible with the received sequence—a $\lceil \log \mathrm{L}_n \rceil$-bit description of the noise sequence (which is of zero rate as L_n is subexponential in the blocklength n) suffices to guarantee zero-error transmission of the codebook $\mathcal{C}'_n$. To this end, we propose the following helper. To simplify its description, we drop the subscript n.

For $\mathbf{z}, \mathbf{z}' \in \mathcal{S}^n$, let us write $\mathbf{z} \sim \mathbf{z}'$ if their componentwise difference is in $\mathcal{C}'$, i.e.,

$$\big(\mathbf{z} \sim \mathbf{z}'\big) \iff \big(\mathbf{z} \ominus \mathbf{z}' \in \mathcal{C}'\big), \quad \mathbf{z}, \mathbf{z}' \in \mathcal{S}^n. \tag{65}$$

Since $\mathcal{C}'$ is a subgroup of $\mathbb{F}_p^n$, this relation is an equivalence relation, and $\mathbf{z} \sim \mathbf{z}'$, i.e., $\mathbf{z}$ and $\mathbf{z}'$ are equivalent iff $\mathbf{z}$ and $\mathbf{z}'$ belong to the same coset of $\mathcal{C}'$. We shall use $[\mathbf{z}] \subseteq \mathcal{S}^n$ to denote the equivalence class containing $\mathbf{z}$.

Our proposed helper assigns labels (descriptions) only to noise sequences in $\mathcal{S}^n$, and it does so in such a way that, unless identical, *equivalent noise sequences are assigned differing labels*. Such a helper leads to zero errors, because if $\mathbf{x} \in \mathcal{C}'$ is transmitted and $\mathbf{x} \oplus \mathbf{z}$ is received (where $\mathbf{z} \in \mathcal{S}^n$), then the decoder can confuse $\mathbf{x}$ with some $\mathbf{x}'$ only if: (i) $\mathbf{x}'$ is also a codeword; (ii) $\mathbf{x} \oplus \mathbf{z} = \mathbf{x}' \oplus \mathbf{z}'$ for some $\mathbf{z}' \in \mathcal{S}^n$; and (iii) $\mathbf{z}$

and $\mathbf{z}'$ have the same label. The former two conditions imply that $\mathbf{z} \sim \mathbf{z}'$, and hence that $\mathbf{z}$ and $\mathbf{z}'$ are identical or else of differing labels. The third condition then implies that they are, in fact, identical, so $\mathbf{x}'$ equals $\mathbf{x}$.

It remains to verify that we can find a labeling rule as above with at most L different labels. This will follow once we show that, for every $\mathbf{z} \in \mathcal{S}^n$,

$$\big|[\mathbf{z}]\big| \leq \mathsf{L}. \tag{66}$$

To establish (66), we note that the L-list-decodability property of $\mathcal{C}'$, namely

$$\mathsf{L} \geq \big|(\mathbf{y} \ominus \mathcal{C}') \cap \mathcal{S}^n\big|, \quad \forall \mathbf{y} \in \mathcal{A}^n \tag{67}$$

is equivalent (because $\mathcal{C}'$ is a subgroup of $\mathbb{F}_p^n$) to

$$\mathsf{L} \geq \big|(\mathbf{y} \oplus \mathcal{C}') \cap \mathcal{S}^n\big|, \quad \forall \mathbf{y} \in \mathcal{A}^n, \tag{68}$$

i.e., to every coset of $\mathcal{C}'$ intersecting $\mathcal{S}^n$ in at most L points. This establishes (66) and concludes the achievability proof.

- Case 3: $0 < R_\mathrm{h} < \log|\mathcal{S}|$. Analogous to (51), the achievability follows from time sharing.

The case with encoder assistance is essentially identical. If $R_\mathrm{h} \geq \log|\mathcal{S}|$, the rate $\log|\mathcal{A}|$ is achievable as in the proof of Theorem 3. If $R_\mathrm{h} = 0$, the relation

$$C_{0,\mathrm{enc}}(0) \geq C_{0,\mathrm{dec}}(0) \tag{69}$$

holds because, in the presence of encoder assistance, any zero-rate help to the encoder can be conveyed to the decoder with negligible extra help and negligible loss in rate: the encoder simply appends a frame to convey the help, with the frame being of sublinear length (because the help to be conveyed is of zero rate); it requests that the helper provide it with a precise description of the noise affecting the frame (with the extra help being negligible because the frame is short); and it subtracts that noise from the transmission in that frame so as to render it noise free. For intermediate values of R_h, the achievability follows by time sharing. ■

Remark 4 (Gap to Capacity vs L*):* By Lemma 2, as the number of labels L tends to infinity, it is possible to communicate error-free at transmission rates that converge to the zero-error helper capacity, with the gap to capacity decaying in L like $\mathcal{O}(1/\log \mathsf{L})$. Although irrelevant to the computation of the capacity, it might be interesting to investigate whether the gap to capacity can decay faster in L.

*Remark 5 (*L*-List-Decodability and (ℓ, L) Recoverability):* On the MMANC, L-list-decodability is related to (zero-error) list-recoverability [19]: Given $\ell, \mathsf{L} \in \mathbb{Z}^+$ and a finite set $\mathcal{X}$, a codebook $\mathcal{C} \subseteq \mathcal{X}^n$ is (ℓ, L)-*list-recoverable* if for any collection of n subsets $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_n$ of $\mathcal{X}$, each of which has no more than ℓ elements,

$$\Big|\mathcal{C} \cap \big(\mathcal{S}_1 \times \cdots \times \mathcal{S}_n\big)\Big| \leq \mathsf{L}. \tag{70}$$

On the MMANC, $(|\mathcal{S}|, \mathsf{L})$-list-recoverability implies L-list-decodability, because, given any output sequence $\mathbf{y} \in \mathcal{A}^n$, we can substitute $y_i \ominus \mathcal{S}$ for each $\mathcal{S}_i$ in (70) to recover L-list-decodability.

Remark 6: If instead of defining R_L as in (52), we defined

$$R_\mathrm{L} = \log\frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{|\mathcal{S}| \cdot \log^2|\mathcal{A}|}{\log(\mathsf{L}+1)}, \tag{71}$$

then the resulting weaker version of Lemma 2, while still sufficient for our purposes, could have been recovered from the literature on (ℓ, L)-list-recoverability, specifically from the result that a random linear code of such rate is $(|\mathcal{S}|$,L)-list-recoverable with high probability [17], [20] and, a fortiori, with positive probability.

Remark 7: The factor of $|\mathcal{S}|$ in the numerator of the second term on the RHS of (71), which is absent from (52), can be improved for large $|\mathcal{A}|$ using [21, Theorem 5.1] and [20].

B. General Case

We now turn to the general case where $|\mathcal{A}|$ need not be prime and establish Theorem 5.

Proof of Theorem 5: As in the proof of Theorem 4, we only need to prove the direct part, and we focus on decoder assistance, so our goal is to establish that

$$\begin{aligned} C_{0,\mathrm{dec}}(R_\mathrm{h}) \geq{}& \frac{\log|\mathcal{A}|}{\log|\mathcal{S}|} \cdot R_\mathrm{h} \\ &+ \left(1 - \frac{R_\mathrm{h}}{\log|\mathcal{S}|}\right) \cdot \max\left\{C_0, \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|}\right\}. \end{aligned} \tag{72}$$

The achievability for encoder-assistance will then follow as in the proof of Theorem 4. To establish (72), we propose the following coding scheme based on time sharing.

- Case 1: $R_\mathrm{h} = \log|\mathcal{S}|$. That $C_{0,\mathrm{dec}}(R_\mathrm{h}) \geq \log|\mathcal{A}|$ follows from the proof for Theorem 3, where the feedback link is not utilized.
- Case 2: $R_\mathrm{h} = 0$. That $C_{0,\mathrm{dec}}(0) \geq C_0$ is obvious, because help cannot hurt.

To show that

$$C_{0,\mathrm{dec}}(0) \geq \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|}, \tag{73}$$

we first introduce some notation. Given a codebook $\mathcal{C} \subseteq \mathcal{A}^n$ and a noise sequence $\mathbf{z} \in \mathcal{S}^n$, let $\mathcal{F}_\mathbf{z}(\mathcal{C})$—or $\mathcal{F}_\mathbf{z}$ for short—be the confusion set of $\mathbf{z}$ comprising the noise sequences confusable with $\mathbf{z}$:

$$\begin{aligned} \mathcal{F}_\mathbf{z}(\mathcal{C}) &= \big\{\mathbf{z}' \in \mathcal{S}^n : \exists\, \mathbf{x}, \mathbf{x}' \in \mathcal{C} \text{ s.t. } \mathbf{x} \neq \mathbf{x}' \\ &\qquad\qquad \text{and } \mathbf{z} \oplus \mathbf{x} = \mathbf{z}' \oplus \mathbf{x}'\big\} \qquad (74) \\ &= \big\{\mathbf{z}' \in \mathcal{S}^n : \mathbf{z}' \ominus \mathbf{z} \in (\mathcal{C} \ominus \mathcal{C})^*\big\}. \qquad (75) \end{aligned}$$

The proposed helper assigns confusable noise sequences different labels: if $\mathbf{z}' \in \mathcal{F}_\mathbf{z}$, then the labels assigned to $\mathbf{z}$ and $\mathbf{z}'$ are different. This guarantees error-free recovery of the noise sequence and hence, if the code has no repeating codewords, also of the transmitted message.

As we next argue, the number of different labels required is at most

$$\max_{\mathbf{z} \in \mathcal{S}^n} \big|\mathcal{F}_\mathbf{z}\big| + 1. \tag{76}$$

Indeed, the number of required labels is the chromatic number of the confusion graph of the noise sequences, i.e., the undirected graph with vertices $\mathcal{S}^n$ and with $\mathbf{z}$ connected to $\mathbf{z}'$ if $\mathbf{z}' \in \mathcal{F}_{\mathbf{z}}$. This graph is well-defined because

$$\big(\mathbf{z}\ominus\mathbf{z}' \in (\mathcal{C}\ominus\mathcal{C})^*\big) \iff \big(\mathbf{z}'\ominus\mathbf{z} \in (\mathcal{C}\ominus\mathcal{C})^*\big). \tag{77}$$

The degree of a vertex $\mathbf{z}$ in this graph is $|\mathcal{F}_{\mathbf{z}}|$, and our claimed upper bound on the number of required labels follows from the fact that the chromatic number of any graph is upper bounded by its maximum degree plus 1.

It remains to establish the existence of a code with no repeating codewords that induces small confusion sets. This is established by the following lemma, whose proof is postponed to the appendix.

Lemma 3: On the MMANC, let $\mathsf{L} > 3$ be a positive integer, and define

$$R_{\mathsf{L}} = \max\left\{0,\ \frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{\log|\mathcal{S}|}{2\log_3 \mathsf{L}}\right\}. \tag{78}$$

Then for any $n \in \mathbb{Z}^+$, there exists a codebook of cardinality $\lfloor 2^{nR_{\mathsf{L}}}\rfloor$, of differing codewords, and for which $\max_{\mathbf{z}\in\mathcal{S}^n}|\mathcal{F}_{\mathbf{z}}| \le \mathsf{L}-1$.

With the aid of the lemma, we can conclude the achievability for zero-rate help using arguments similar to those we used in the proof of Theorem 4: Consider a sequence of blocklength-n codebooks whose existence is guaranteed by Lemma 3 when we substitute L_n for L, where $\{\mathsf{L}_n\}$ tends to infinity subexponentially in n. With these codebooks and the proposed helper, transmission is error-free, the helping rate is zero, and the transmission rate approaches $\frac{1}{2}\log\frac{|\mathcal{A}|}{|\mathcal{S}|}$, thus proving (73).

- Case 3: $0 < R_{\mathrm{h}} < \log|\mathcal{S}|$. Follows from time sharing, by dividing the transmission block into two parts of relative length $\frac{R_{\mathrm{h}}}{\log|\mathcal{S}|}$ and $1-\frac{R_{\mathrm{h}}}{\log|\mathcal{S}|}$ and applying the aforementioned schemes. ■

Appendix

Proof of Lemma 3: Without loss of generality, assume $R_{\mathsf{L}} > 0$ and $|\mathcal{S}|^n > \mathsf{L}$ (otherwise the result is obvious). Define

$$\begin{aligned} R &= \frac{1}{n}\log\lfloor 2^{nR_{\mathsf{L}}}\rfloor && (79)\\ &\le R_{\mathsf{L}} && (80)\end{aligned}$$

and let $\mathcal{M} = \{1,\ldots,2^{nR}\}$. Generate a random codebook $\mathcal{C} = \{\mathbf{X}(1),\ldots,\mathbf{X}(|\mathcal{M}|)\}$ by drawing its codewords independently, each equiprobably from $\mathcal{A}^n$. We will show that with positive probability, the properties (i) $\mathcal{C}$ contains no repeating codewords, and (ii) $\max_{\mathbf{z}\in\mathcal{S}^n}|\mathcal{F}_{\mathbf{z}}| \le \mathsf{L}-1$ hold simultaneously.

Using the Union Bound, we can upper-bound the probability that Property (i) is violated as follows:

$$\begin{aligned} &\Pr\big[\exists\, m,m'\in\mathcal{M} \text{ s.t. } m\neq m' \text{ and } \mathbf{X}(m)=\mathbf{X}(m')\big]\\ &\quad\le \sum_{1\le m<m'\le 2^{nR}} \Pr\big[\mathbf{X}(m)=\mathbf{X}(m')\big] && (81)\\ &\quad= \binom{2^{nR}}{2}\cdot\frac{1}{|\mathcal{A}|^n} && (82)\\ &\quad\le \frac{2^{2nR}}{2|\mathcal{A}|^n}. && (83)\end{aligned}$$

As for Property (ii),

$$\begin{aligned} &\Pr\big[\max_{\mathbf{z}\in\mathcal{S}^n}|\mathcal{F}_{\mathbf{z}}| > \mathsf{L}-1\big]\\ &\quad= \Pr\big[\exists\,\mathbf{z}\in\mathcal{S}^n \text{ s.t. } |\mathcal{F}_{\mathbf{z}}| \ge \mathsf{L}\big] && (84)\\ &\quad\le \sum_{\mathbf{z}\in\mathcal{S}^n}\Pr\big[|\mathcal{F}_{\mathbf{z}}|\ge\mathsf{L}\big] && (85)\\ &\quad= \sum_{\mathbf{z}\in\mathcal{S}^n}\Pr\big[\exists\,\text{distinct } \boldsymbol{\xi}'_1,\ldots,\boldsymbol{\xi}'_{\mathsf{L}}\in\mathcal{S}^n\\ &\qquad\qquad \text{s.t. } \{\boldsymbol{\xi}'_1,\ldots,\boldsymbol{\xi}'_{\mathsf{L}}\}\ominus\mathbf{z}\subseteq(\mathcal{C}\ominus\mathcal{C})^*\big] && (86)\\ &\quad= \sum_{\mathbf{z}\in\mathcal{S}^n}\Pr\big[\exists\,\text{distinct } \boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\in\mathcal{S}^n\ominus\mathbf{z}\\ &\qquad\qquad \text{s.t. } \{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}\subseteq(\mathcal{C}\ominus\mathcal{C})^*\big], && (87)\end{aligned}$$

where (85) follows from the Union Bound; (86) follows from the definition of $\mathcal{F}_{\mathbf{z}}$ in (75); and in (87) we introduced $\boldsymbol{\xi}_i = \boldsymbol{\xi}'_i\ominus\mathbf{z}$.

To analyze the probabilities appearing on the RHS of (87), we first rule out the degenerate cases. We say that a collection of n-tuples $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell\in\mathcal{A}^n$ is *tri-independent* if

$$\begin{aligned} &\Big(\sum_{i=1}^{\ell}\epsilon_i\boldsymbol{\xi}_i = 0,\ \text{with } \epsilon_i\in\{0,\pm1\},\ \forall\, i\in[\ell]\Big)\\ &\qquad\Longrightarrow\ \Big(\epsilon_i = 0,\ \forall\, i\in[\ell]\Big). && (88)\end{aligned}$$

Lemma 4: Among any L distinct n-tuples $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\in\mathcal{A}^n$, there exist at least $\lceil\log_3\mathsf{L}\rceil$ that are tri-independent.

Proof of Lemma 4: Let $\mathcal{B}\subseteq\{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}$ be a maximal tri-independent set (with respect to inclusion), and let ℓ be its cardinality: $|\mathcal{B}| = \ell$. Without loss of generality, assume $\mathcal{B} = \{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell\}$ (otherwise rearrange the tuples). We will show that every n-tuple in $\{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}$ can be expressed in the form $\sum_{i=1}^{\ell}\epsilon_i\boldsymbol{\xi}_i$, with $\epsilon_i\in\{0,\pm1\}$ for all $i\in[\ell]$. (This is obvious for the n-tuples $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell$ but requires proof for $\boldsymbol{\xi}_{\ell+1},\ldots,\boldsymbol{\xi}_{\mathsf{L}}$.) This will establish the lemma because the number of distinct expressions of said form is at most 3^ℓ, and the number of sequences is L, so 3^ℓ must be at least L, and ℓ (the number of elements of $\mathcal{B}$) must thus satisfy $\ell\ge\lceil\log_3\mathsf{L}\rceil$.

To see that any $\boldsymbol{\xi}\in\{\boldsymbol{\xi}_{\ell+1},\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}$ can be expressed in said form, note that the maximality of $\mathcal{B}$ implies that for any $\boldsymbol{\xi}\in\{\boldsymbol{\xi}_{\ell+1},\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}$, there exist nontrivial (i.e., not all zero) $\{\epsilon_i\}_{i=1}^{\ell}$ and ϵ, such that $\sum_{i=1}^{\ell}\epsilon_i\boldsymbol{\xi}_i + \epsilon\boldsymbol{\xi} = 0$. Here ϵ cannot be zero, since otherwise the relation would translate to $\sum_{i=1}^{\ell}\epsilon_i\boldsymbol{\xi}_i = 0$ for nontrivial $\{\epsilon_i\}_{i=1}^{\ell}$, which would contradict the fact that, by construction, $\mathcal{B}$ is tri-independent. Without loss of generality, assume $\epsilon = -1$ (otherwise change the sign of ϵ and all ϵ_i's), so $\boldsymbol{\xi} = \sum_{i=1}^{\ell}\epsilon_i\boldsymbol{\xi}_i$, thus expressing $\boldsymbol{\xi}$ in said form. ■

With the aid of Lemma 4, and defining

$$\ell \triangleq \lceil\log_3\mathsf{L}\rceil \ge 2, \tag{89}$$

we can return to the RHS of (87) to conclude that

$$\begin{aligned} &\Pr\big[\exists\,\text{distinct } \boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\in\mathcal{S}^n\ominus\mathbf{z}\\ &\qquad\qquad \text{s.t. } \{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{\mathsf{L}}\}\subseteq(\mathcal{C}\ominus\mathcal{C})^*\big]\\ &\quad\le \Pr\big[\exists\ \text{tri-independent } \boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell\in\mathcal{S}^n\ominus\mathbf{z}\\ &\qquad\qquad \text{s.t. } \{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell\}\subseteq(\mathcal{C}\ominus\mathcal{C})^*\big]. && (90)\end{aligned}$$

Ignoring, for now, the constraint that $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_\ell$ be in $\mathcal{S}^n\ominus\mathbf{z}$, we claim:

Claim 1: Given ℓ tri-independent n-tuples $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell \in \mathcal{A}^n$, and a random codebook $\mathcal{C}$ generated as above

$$\Pr\bigl[\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\} \subseteq (\mathcal{C} \ominus \mathcal{C})^*\bigr] \leq \left(\frac{2^{2nR}}{|\mathcal{A}|^n}\right)^\ell. \tag{91}$$

Proof: Since $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\}$ are tri-independent and hence, *a fortiori*, nonzero, the event $\bigl[\boldsymbol{\xi}_i \in (\mathcal{C} \ominus \mathcal{C})^*\bigr]$ is equivalent to the event $\bigl[\exists\, (m_i, m_i') \in \mathcal{M} \times \mathcal{M} \text{ s.t. } \mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i\bigr]$. Consequently, the event $\bigl[\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\} \subseteq (\mathcal{C} \ominus \mathcal{C})^*\bigr]$ is equivalent to the event

$$\begin{aligned}&\exists \{(m_i, m_i')\}_{i=1}^\ell \subseteq \mathcal{M} \times \mathcal{M}\\ &\qquad \text{s.t. } \Bigl(\mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i,\ \forall i \in [\ell]\Bigr).\end{aligned} \tag{92}$$

Hence, by the Union Bound,

$$\begin{aligned}&\Pr\bigl[\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\} \subseteq (\mathcal{C} \ominus \mathcal{C})^*\bigr]\\ &\leq \sum_{\{(m_i, m_i')\}_{i=1}^\ell \subseteq \mathcal{M} \times \mathcal{M}} \Pr \bigcap_{i \in [\ell]} \bigl[\mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i\bigr] &(93)\\ &\leq 2^{2nR\ell} |\mathcal{A}|^{-n\ell}, &(94)\end{aligned}$$

where (94) follows from Lemma 5 ahead and from the fact that the number of terms on the RHS of (93) is upper bounded by $2^{2nR\ell}$. ■

From (87) and (90) we now obtain

$$\begin{aligned}&\Pr\bigl[\max_{\mathbf{z} \in \mathcal{S}^n} |\mathcal{F}_{\mathbf{z}}| > \mathsf{L} - 1\bigr]\\ &\leq \sum_{\mathbf{z} \in \mathcal{S}^n} \Pr\bigl[\exists \text{ tri-independent } \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell \in \mathcal{S}^n \ominus \mathbf{z}\\ &\qquad\qquad \text{s.t. } \{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\} \subseteq (\mathcal{C} \ominus \mathcal{C})^*\bigr] &(95)\\ &\leq \sum_{\mathbf{z} \in \mathcal{S}^n} \sum_{\substack{\text{tri-independent}\\ \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell \in \mathcal{S}^n \ominus \mathbf{z}}} \Pr\bigl[\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_\ell\} \subseteq (\mathcal{C} \ominus \mathcal{C})^*\bigr] &(96)\\ &\leq |\mathcal{S}|^n \binom{|\mathcal{S}|^n}{\ell} \left(\frac{2^{2nR}}{|\mathcal{A}|^n}\right)^\ell &(97)\\ &\leq \frac{|\mathcal{S}|^{n(\ell+1)}}{\ell!} \left(\frac{2^{2nR}}{|\mathcal{A}|^n}\right)^\ell, &(98)\end{aligned}$$

where (96) follows from the Union Bound, and (97) follows from Claim 1.

From (83) and (98) (and the Union Bound) we infer that, the two properties (i) and (ii) hold simultaneously with probability strictly larger than

$$\begin{aligned}&1 - \frac{2^{2nR}}{2|\mathcal{A}|^n} - \frac{|\mathcal{S}|^{n(\ell+1)}}{\ell!} \left(\frac{2^{2nR}}{|\mathcal{A}|^n}\right)^\ell\\ &\geq 1 - \frac{1}{2} |\mathcal{S}|^{-n\frac{\ell+1}{\ell}} - \frac{1}{\ell!} &(99)\\ &\geq 1 - \frac{1}{2} - \frac{1}{\ell!} &(100)\\ &\geq 0, &(101)\end{aligned}$$

where (99) holds because, by (78) and (80)

$$2^{2nR} \leq 2^{2n\left(\frac{1}{2} \log \frac{|\mathcal{A}|}{|\mathcal{S}|} - \frac{\log |\mathcal{S}|}{2\ell}\right)} = |\mathcal{A}|^n |\mathcal{S}|^{-n\frac{\ell+1}{\ell}}. \tag{102}$$

Thus, with positive probability, the random codebook $\mathcal{C}$ satisfies both desired properties simultaneously. This concludes the proof of Lemma 3 (assuming Lemma 5 ahead). ■

We next state and prove Lemma 5.

Lemma 5: Let the ℓ n-tuples $\{\boldsymbol{\xi}_i\}_{i=1}^\ell$ be tri-independent, and let $\{(m_i, m_i')\}_{i=1}^\ell \subseteq \mathcal{M} \times \mathcal{M}$ be ℓ message tuples. Let the undirected graph $G = (V, E)$ be of vertices $V = \mathcal{M}$ and of edges $E = \{(m_i, m_i')\}_{i=1}^\ell = \{e_i\}_{i=1}^\ell$, where e_i denotes the edge (m_i, m_i'). If $\mathbf{X}(1), \dots, \mathbf{X}(|\mathcal{M}|)$ are drawn IID, each equiprobably from $\mathcal{A}^n$, then the probability

$$\Pr \bigcap_{i \in [\ell]} \bigl[\mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i\bigr] \tag{103}$$

equals $|\mathcal{A}|^{-n\ell}$ if the graph G is acyclic,[4] and equals zero otherwise.

Proof: First, assume that G contains some cycle, say of vertices $v_1, v_2, \dots, v_k, v_{k+1} \triangleq v_1$. By possibly permuting $\{\boldsymbol{\xi}_i, (m_i, m_i')\}$ we may assume without loss of generality that $\{v_1, v_2\} = \{m_1, m_1'\}$ (as sets), so either $(v_1, v_2) = (m_1, m_1')$ or $(v_1, v_2) = (m_1', m_1)$ (or both, if this is a loop). In the former case the event of interest is $\bigl[\mathbf{X}(m_1) \ominus \mathbf{X}(m_1') = \boldsymbol{\xi}_1\bigr]$ and in the latter $\bigl[\mathbf{X}(m_1') \ominus \mathbf{X}(m_1) = \boldsymbol{\xi}_1\bigr]$. We similarly assume that $\{v_i, v_{i+1}\} = \{m_i, m_{i+1}'\}$ (as sets) for all $i \in [k]$.

The probability (103) is positive iff there exist $\mathbf{x}_1, \dots, \mathbf{x}_{|\mathcal{M}|} \in \mathcal{A}^n$ such that

$$\mathbf{x}_{m_i} \ominus \mathbf{x}_{m_i'} = \boldsymbol{\xi}_i, \quad i \in [\ell], \tag{104}$$

which, as we shall see, contradicts the nondegeneracy of the n-tuples.

Define

$$\epsilon_i = \begin{cases} 1 & \text{if } (v_i, v_{i+1}) = (m_i, m_i'), \\ -1 & \text{otherwise,} \end{cases} \quad i \in [k]. \tag{105}$$

Then, (104) and (105) imply

$$\begin{aligned}\sum_{i=1}^k \epsilon_i \boldsymbol{\xi}_i &= \sum_{i=1}^k \epsilon_i \bigl(\mathbf{x}_{m_i} \ominus \mathbf{x}_{m_i'}\bigr) &(106)\\ &= \sum_{i=1}^k \bigl(\mathbf{x}_{v_i} \ominus \mathbf{x}_{v_{i+1}}\bigr) &(107)\\ &= 0 &(108)\end{aligned}$$

where the last equality holds because $v_{k+1} = v_1$.

By setting $\epsilon_i = 0$ for all $i \in [\ell] \setminus [k]$ (i.e., for all edges not in the cycle), we obtain ℓ coefficients $\{\epsilon_i\}_{i=1}^\ell$ (that are not all zero) taking values in $\{0, \pm 1\}$ such that $\sum_{j=1}^\ell \epsilon_j \boldsymbol{\xi}_j = \sum_{j=1}^k \epsilon_j \boldsymbol{\xi}_j = 0$, which contradicts the tri-independence assumption.

We next consider the case where the graph G is acyclic. *A fortiori*, G contains no loops, so $m_i \neq m_i'$ for all $i \in [\ell]$. Proving that

$$\Pr \bigcap_{i \in [\ell]} \bigl[\mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i\bigr] = \frac{1}{|\mathcal{A}|^{n\ell}} \tag{109}$$

is tantamount to proving that the events $\bigl\{\bigl[\mathbf{X}(m_i) \ominus \mathbf{X}(m_i') = \boldsymbol{\xi}_i\bigr]\bigr\}_{i \in [\ell]}$ are independent, because it can be readily

[4] G is allowed to contain loops and parallel edges, which also count as cycles.

verified that for any $m \neq m'$ and $\boldsymbol{\xi} \in \mathcal{A}^n$, $\Pr\big[\mathbf{X}(m) \ominus \mathbf{X}(m') = \boldsymbol{\xi}\big] = \frac{1}{|\mathcal{A}|^n}$.

Because the codewords are chosen independently, the events corresponding to edges in different connected components of G are independent. We can therefore focus on one non-empty connected component, say $G_1 = (V_1, E_1)$, and show that the events corresponding to E_1 are independent. That is, we need to show that

$$\Pr \bigcap_{i\colon e_i \in E_1} \big[\mathbf{X}(m_i) \ominus \mathbf{X}(m'_i) = \boldsymbol{\xi}_i\big] = \frac{1}{|\mathcal{A}|^{n|E_1|}}. \tag{110}$$

To prove this, we will show by induction on $|E_1|$ that the system of linear equations (with $|E_1|$ equations and $|V_1|$ variables) corresponding to this connected component, namely,

$$\mathbf{x}_{m_i} \ominus \mathbf{x}_{m'_i} = \boldsymbol{\xi}_i, \qquad i \in \{j\colon e_j \in E_1\} \tag{111}$$

has $|\mathcal{A}|^n$ solutions. This is to be expected because G is acyclic, so G_1 is a tree, and hence $|V_1| = |E_1| + 1$.

If $|E_1| = 1$, there are two variables and one equation, so the number of solutions is, indeed, $|\mathcal{A}|^n$. If $|E_1| \geq 2$, let m_0 be a degree one vertex of G_1 (which is guaranteed to exist because G_1 is a tree). As such, $\mathbf{x}(m_0)$ appears in only one of the equations, and it is therefore uniquely determined by the remaining $(|V_1| - 1)$ variables. The number of solutions is thus as for the system that remains when we remove $\mathbf{x}_{m_0}$ and the equation in which it appears from the system, which leaves us with $(|E_1| - 1)$ equations corresponding to the induced subgraph $G_1[V_1 \setminus \{m_0\}]$. This subgraph is still a tree, and hence, by the induction hypothesis, this restricted system with $(|E_1| - 1)$ equations and $(|V_1| - 1)$ variables has $|\mathcal{A}|^n$ solutions. This concludes the induction and establishes that, as we claimed, the system of equations (111) has $|\mathcal{A}|^n$ solutions.

We can now complete the proof of (110):

$$\begin{aligned}
\Pr &\bigcap_{i\colon e_i \in E_1} \big[\mathbf{X}(m_i) \ominus \mathbf{X}(m'_i) = \boldsymbol{\xi}_i\big] \\
&= \sum_{\substack{\{\mathbf{x}_m\}_{m \in V_1} \\ \text{solving (111)}}} \Pr \bigcap_{m \in V_1} \big[\mathbf{X}(m) = \mathbf{x}_m\big] && (112) \\
&= \sum_{\substack{\{\mathbf{x}_m\}_{m \in V_1} \\ \text{solving (111)}}} \frac{1}{|\mathcal{A}|^{n|V_1|}} && (113) \\
&= \frac{|\mathcal{A}|^n}{|\mathcal{A}|^{n|V_1|}} && (114) \\
&= \frac{1}{|\mathcal{A}|^{n|E_1|}}, && (115)
\end{aligned}$$

where the last equality holds because G_1 is a tree, so $|V_1| = |E_1| + 1$. ■

Acknowledgment

Discussions with Mary Wootters are gratefully acknowledged.

References

[1] C. Shannon, “The zero error capacity of a noisy channel,” *IEEE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.

[2] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Oct. 1948.

[3] 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.

[4] A. Lapidoth and G. Marti, “Encoder-assisted communications over additive noise channels,” *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6607–6616, Nov. 2020.

[5] A. Lapidoth, G. Marti, and Y. Yan, “Other helper capacities,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1272–1277.

[6] A. Lapidoth, L. Wang, and Y. Yan, “Message-cognizant assistance and feedback for the Gaussian channel,” 2023, *arXiv:2310.15768*.

[7] A. Lapidoth, L. Wang, and Y. Yan, “State-dependent channels with a message-cognizant helper,” 2023, *arXiv:2311.08220*.

[8] N. Merhav, “On error exponents of encoder-assisted communication systems,” *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7019–7029, Nov. 2021.

[9] A. Lapidoth and Y. Yan, “The listsize capacity of the Gaussian channel with decoder assistance,” *Entropy*, vol. 24, no. 1, p. 29, Dec. 2021. [Online]. Available: https://www.mdpi.com/1099-4300/24/1/29

[10] S. Loyka and N. Merhav, “The secrecy capacity of the Gaussian wiretap channel with rate-limited help at the decoder,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2022, pp. 1040–1045.

[11] L. Lovász, “On the Shannon capacity of a graph,” *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 1–7, Jan. 1979.

[12] G. Marti, “Channels with a helper,” M.S. thesis, ETH Zurich, Zurich, Switzerland, 2019.

[13] A. Bracher and A. Lapidoth, “The zero-error feedback capacity of state-dependent channels,” *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3538–3578, May 2018.

[14] P. Elias, “Zero error capacity under list decoding,” *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1070–1074, Sep. 1988.

[15] İ. E. Telatar, “Zero-error list capacities of discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1977–1982, Nov. 1997.

[16] M. Dalai, V. Guruswami, and J. Radhakrishnan, “An improved bound on the zero-error list-decoding capacity of the 4/3 channel,” *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 749–756, Feb. 2020.

[17] V. V. Zyablov and M. S. Pinsker, “List concatenated decoding,” *Problemy Peredachi Informatsii*, vol. 17, no. 4, pp. 29–33, 1981.

[18] P. Elias, “Error-correcting codes for list decoding,” *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 5–12, Jan. 1991.

[19] V. Guruswami, *List Decoding of Error-Correcting Codes: Winning Thesis of the 2002 ACM Doctoral Dissertation Competition*, vol. 3282. Berlin, Germany: Springer, 2004.

[20] V. Guruswami, R. Li, J. Mosheiff, N. Resch, S. Silas, and M. Wootters, “Bounds for list-decoding and list-recovery of random linear codes,” *IEEE Trans. Inf. Theory*, vol. 68, no. 2, pp. 923–939, Feb. 2022.

[21] A. Rudra and M. Wootters, “Average-radius list-recoverability of random linear codes,” in *Proc. 29th Annu. ACM-SIAM Symp. Discrete Algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2018, pp. 644–662.

Amos Lapidoth (Fellow, IEEE) received the B.A. degree (summa cum laude) in mathematics and the B.Sc. (summa cum laude) and M.Sc. degrees 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 Professor and an Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT). He was the KDD Career Development Associate Professor of communications and technology. He is currently a Professor of information theory with ETH Zurich, Switzerland. He is the author of the textbook *A Foundation in Digital Communication* (Cambridge University Press, Second Edition, 2017). His research interests include digital communications and information theory.

Dr. Lapidoth served as an Associate Editor for Shannon Theory of IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2004 and in 2009.

Yiming Yan (Graduate Student Member, IEEE) received the B.E. degree in electronic engineering from Tsinghua University in 2018 and the M.Sc. degree in electrical engineering from ETH Zurich in 2020, where she is currently pursuing the Ph.D. degree with the Signal and Information Processing Laboratory (ISI).