Assisted Identification over Modulo-Additive Noise Channels

Journal Article

Author(s):
Lapidoth, Amos; Ni, Baohua

Publication date:
2023-09

Permanent link:
https://doi.org/10.3929/ethz-b-000634657

Rights / license:
Creative Commons Attribution 4.0 International

Originally published in:
Entropy 25(9), https://doi.org/10.3390/e25091314

entropy

MDPI

Article

Assisted Identification over Modulo-Additive Noise Channels

Amos Lapidoth * **and Baohua Ni**

Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland; baohni@isi.ee.ethz.ch
* Correspondence: lapidoth@isi.ee.ethz.ch

Abstract: The gain in the identification capacity afforded by a rate-limited description of the noise sequence corrupting a modulo-additive noise channel is studied. Both the classical Ahlswede–Dueck version and the Ahlswede–Cai–Ning–Zhang version, which does not allow for missed identifications, are studied. Irrespective of whether the description is provided to the receiver, to the transmitter, or to both, the two capacities coincide and both equal the helper-assisted Shannon capacity.

Keywords: erasures-only capacity; helper; identification capacity; modulo-additive noise; rate-limited; zero-undetected-error capacity

1. Introduction

If a helper can observe the additive noise corrupting a channel and can describe it to the decoder, then the latter can subtract it and thus render the channel noiseless. However, for this to succeed, the description must be nearly lossless and hence possibly of formidable rate. It is thus of interest to study scenarios where the description rate is limited, and to understand how the rate of the help affects performance.

When performance is measured in terms of the Shannon capacity, the problem was solved for a number of channel models [1–3], where the former two address assistance to the decoder and the latter to the encoder. When performance was measured in terms of the erasures-only capacity or the list-size capacity, the problem was solved in [4,5]. Error exponents with assistance were studied in [6]. Here we study how rate-limited help affects the identification capacity [7].

We focus on the memoryless modulo-additive 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 is the time-k noise sample; the channel input x_k, the channel output Y_k, and the noise Z_k all take values in the set $\mathcal{A}$—also denoted $\mathcal{X}$, or $\mathcal{Y}$, or $\mathcal{Z}$—comprising the $|\mathcal{A}|$ elements $\{0, \ldots, |\mathcal{A}| - 1\}$; and $\oplus$ and $\ominus$ denote mod-$|\mathcal{A}|$ addition and subtraction, respectively. The noise sequence $\{Z_k\}$ is IID $\sim P_Z$, where P_Z is some PMF on $\mathcal{A}$.

Irrespective of whether the help is provided to the encoder, to the decoder, or to both, the Shannon capacity of this channel coincides with its erasures-only capacity, and both are given by [3] (Section V) and [4] (Theorems 2 and 6):

$$C_{\text{e-o}}(R_\text{h}) = C_\text{Sh}(R_\text{h}) = \log|\mathcal{A}| - \{H(P_Z) - R_\text{h}\}^+ \tag{2}$$

where $\{\xi\}^+$ denotes $\max\{0, \xi\}$, and $H(P_Z)$ is the Shannon entropy of P_Z.

Here we study two versions of the identification capacity of this channel: Ahlswede and Dueck's original identification capacity C_ID [7], and the identification capacity subject to no missed-identifications $C_{\text{ID},0}$ [8]. Our main result is that—irrespective of whether the help is provided to the encoder, to the decoder, or to both—the two identification capacities coincide and both equal the right-hand side (RHS) of (2).

check for updates

Citation: Lapidoth, A.; Ni, B. Assisted Identification over Modulo-Additive Noise Channels. *Entropy* **2023**, *25*, 1314. https://doi.org/10.3390/e25091314

Academic Editor: Igal Sason

Received: 15 June 2023
Revised: 28 August 2023
Accepted: 1 September 2023
Published: 8 September 2023

2. Problem Formulation

The identification-over-a-channel problem is parameterized by the blocklength n, which tends to infinity in the definition of the identification capacity. The n-length noise sequence $Z^n \in \mathcal{A}^n$ is presented to a helper, which produces its nR_h-bit description $t(Z^n)$:

$$t(z^n) \in \mathcal{T} \tag{3}$$

where:

$$\mathcal{T} = \{0,1\}^{nR_\mathrm{h}}. \tag{4}$$

We refer to the set $\mathcal{N} = \{1, \ldots, \mathsf{N}\}$ as the set of identification messages and to its cardinality N as the number of identification messages. The identification rate is defined (for N sufficiently large) as:

$$\frac{1}{n} \log \log \mathsf{N}. \tag{5}$$

A generic element of $\mathcal{N}$—namely, a generic identification message—is denoted i.

If no help is provided to the encoder, then the latter is specified by a family $\{P^i_{X^n}\}_{i\in\mathcal{N}}$ of PMFs on $\mathcal{A}^n$ that are indexed by the identification messages, with the understanding that, to convey the identification message (IM) i, the encoder transmits a random sequence in $\mathcal{A}^n$ that it draws according to the PMF $P^i_{X^n}$. If help $T = t(Z^n) \in \mathcal{T}$ is provided to the encoder, then the encoder's operation is specified by a family of PMFs $\{P^i_{X^n|t}\}_{(i,t)\in\mathcal{N}\times\mathcal{T}}$ that is now indexed by pairs of identification messages and noise descriptions, with the understanding that, to convey IM i given the description $T = t(Z^n)$, the encoder produces a random n-length sequence of channel inputs that is distributed according to $P^i_{X^n|T}$. In either case, the channel output sequence Y^n is:

$$Y^n = X^n \oplus Z^n \tag{6}$$

componentwise.

If help is provided to the encoder, and if IM i is to be conveyed, then the joint distribution of (X^n, Z^n, Y^n, T) has the form:

$$P_{Z^n}(z^n)\, P_{T|Z^n}(t|z^n)\, P^i_{X^n|T}(x^n|t)\, P_{Y^n|X^n,Z^n}(y^n|x^n, z^n) \tag{7}$$

where

$$P_{Y^n|X^n,Z^n}(y^n|x^n, z^n) = \mathbb{1}\{y^n = x^n \oplus z^n\} \tag{8}$$

and where

$$P_{T|Z^n}(t|z^n) = \mathbb{1}\{t = t(z^n)\} \tag{9}$$

because we are assuming that the noise description is a deterministic function of the noise sequence. (The results also hold if we allow randomized descriptions: our coding schemes employ deterministic descriptions and the converse allows for randomization.) Here $\mathbb{1}\{\text{statement}\}$ equals 1 if the statement holds and equals 0 otherwise. In the absence of help, the joint distribution has the form:

$$P_{Z^n}(z^n)\, P_{T|Z^n}(t|z^n)\, P^i_{X^n}(x^n)\, P_{Y^n|X^n,Z^n}(y^n|x^n, z^n). \tag{10}$$

Based on the data available to it—Y^n in the absence of help to the decoder and $(Y^n, t(Z^n))$ in its presence—the receiver performs N binary tests indexed by $i \in \mathcal{N}$, where the i-th test is whether or not the IM is i. It accepts the hypothesis that the IM is i if Y^n is in its acceptance region, which we denote $\mathcal{D}_i(t) \in \mathcal{A}^n$ in the presence of decoder assistance $t \in \mathcal{T}$ and $\mathcal{D}_i \in \mathcal{A}^n$ in its absence.

When the help $t \in \mathcal{T}$ is provided to the receiver, the probability of missed detection associated with IM $i \in \mathcal{N}$ is thus:

$$p^i_{\mathrm{MD}}(t) = 1 - P^i_{Y^n|T=t}\big(\mathcal{D}_i(t)\big) \tag{11}$$

and the worst-case false alarm associated with it is:

$$p^i_{\text{FA}}(t) = \max_{j \in \mathcal{N} \setminus \{i\}} P^j_{Y^n|T=t}(\mathcal{D}_i(t)). \tag{12}$$

Note that, given $t \in \mathcal{T}$, the acceptance regions $\{\mathcal{D}_i(t)\}_{i \in \mathcal{N}}$ of the different tests need not be disjoint. We define:

$$p_{\text{MD,max}} = \max_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} P_T(t)\, p^i_{\text{MD}}(t) \tag{13}$$

and:

$$p_{\text{FA,max}} = \max_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} P_T(t)\, p^i_{\text{FA}}(t). \tag{14}$$

In the absence of help to the receiver, the probability of missed detection associated with IM i is:

$$p^i_{\text{MD}} = 1 - \sum_{t \in \mathcal{T}} P_T(t) P^i_{Y^n|T=t}(\mathcal{D}_i) = 1 - P^i_{Y^n}(\mathcal{D}_i) \tag{15}$$

and the worst-case probability of false alarm associated with it is:

$$p^i_{\text{FA}} = \sum_{t \in \mathcal{T}} P_T(t) \max_{j \in \mathcal{N} \setminus \{i\}} P^j_{Y^n|T=t}(\mathcal{D}_i). \tag{16}$$

In this case, we define:

$$p_{\text{MD,max}} = \max_{i \in \mathcal{N}} p^i_{\text{MD}} \tag{17}$$

and:

$$p_{\text{FA,max}} = \max_{i \in \mathcal{N}} p^i_{\text{FA}}. \tag{18}$$

In both cases we say that a scheme is of zero missed detectionsif $p_{\text{MD,max}}$ is zero.

A rate R is an achievable identification rate if, for every $\gamma > 0$ and every $\epsilon > 0$, there exists some positive integer n_0 such that, for all blocklengths n exceeding n_0, there exists a scheme with:

$$\mathsf{N} = \lceil 2^{2^{n(R-\gamma)}} \rceil \tag{19}$$

identification messages for which:

$$\max\{p_{\text{MD,max}}, p_{\text{FA,max}}\} < \epsilon. \tag{20}$$

The supremum of achievable rates is the identification capacity with a helper $C_{\text{ID}}(R_\text{h})$. Replacing requirement (20) with:

$$p_{\text{MD,max}} = 0, \qquad p_{\text{FA,max}} < \epsilon \tag{21}$$

leads to the definition of the zero missed-identification capacity $C_{\text{ID},0}(R_\text{h})$.

Remark 1. *Writing out $p_{\text{FA,max}}$ of (14) as:*

$$p_{\text{FA,max}} = \max_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} P_T(t) \max_{j \in \mathcal{N} \setminus \{i\}} P^j_{Y^n|T=t}(\mathcal{D}_i(t)) \tag{22}$$

highlights that (prior to maximizing over i) we first maximize over j and then average the result over t. In this sense, the help—even if provided to both encoder and decoder—cannot be viewed as "common randomness" in the sense of [9–11] where the averaging over the common randomness is performed before taking the maximum. Our criterion is more demanding of the direct part (code construction) and less so of the converse.

Both criteria are interesting. Ours allows for the notion of "outage", namely, descriptions that indicate that identification might fail and that therefore call for retransmission. The other criterion *highlights the interplay between the noise description and the generation of common randomness (particularly when the help is provided to both transmitter and receiver).*

The following theorem is the main result of this paper.

Theorem 1. *On the modulo additive noise channel—irrespective of whether the help is provided to the transmitter, to the receiver, or to both—the identification capacity with a helper $C_{\mathrm{ID}}(R_{\mathrm{h}})$ and the zero missed-identification capacity with a helper $C_{\mathrm{ID},0}(R_{\mathrm{h}})$ are equal and coincide with the Shannon capacity:*

$$C_{\mathrm{ID}}(R_{\mathrm{h}}) = C_{\mathrm{ID},0}(R_{\mathrm{h}}) = C_{\mathrm{Sh}}(R_{\mathrm{h}}) \tag{23}$$

where the latter is given in (2).

We prove this result by establishing in Section 3 that $C_{\mathrm{ID},0}(R_{\mathrm{h}}) \geq C_{\mathrm{Sh}}(R_{\mathrm{h}})$ using a slight strengthening of recent results in [4] in combination with the code construction proposed in [8]. The converse is proved in Section 4, where we use a variation on a theme by Watanabe [12] to analyze the case where the assistance is provided to both transmitter and receiver.

3. Direct Part: Zero Missed Detection

In this section we prove that:

$$C_{\mathrm{ID},0}(R_{\mathrm{h}}) \geq C_{\mathrm{Sh}}(R_{\mathrm{h}}) \tag{24}$$

by proposing identification schemes of no missed detections and of rates approaching $C_{\mathrm{Sh}}(R_{\mathrm{h}})$. To this end, we extend to the helper setting the connection—due to Ahlswede, Cai, and Zhang [8]—between the zero-missed-detection identification capacity $C_{\mathrm{ID},0}$ and the erasures-only capacity $C_{\text{e-o}}$. We then call on recent results [4] to infer that, on the modulo-additive noise channel with a helper, the Erasures-Only capacity is equal to the Shannon capacity. We treat encoder-only assistance and decoder-only assistance separately. Either case also proves achievability when the assistance is provided to both encoder and decoder.

Recall that an erasures-only decoder produces a list $\mathcal{L}$ comprising the messages under which the observation is of positive likelihood and then act as follows: If the list contains only one message, it produces that message; otherwise, it declares an erasure. Since the list always contains the transmitted message, this decoder never errs. The erasures-only capacity is defined like the Shannon capacity, but with the additional requirement that the decoder be the erasures-only decoder. This notion extends in a natural way to settings with a helper [4].

3.1. Encoder Assistance

A rate-R, blocklength-n, encoder-assisted, erasures-only transmission code comprises a message set $\mathcal{M} = \{1, \ldots, \mathsf{M}\}$ with $\mathsf{M} = 2^{nR}$ messages and a collection of M mappings $\{f_m\}_{m\in\mathcal{M}}$ from $\mathcal{T}$ to $\mathcal{X}^n$, indexed by $\mathcal{M}$, with the understanding that to transmit Message m after being presented with the help $t(Z^n) \in \mathcal{T}$, the encoder produces the n-tuple of channel inputs $f_m\big(t(Z^n)\big) \in \mathcal{X}^n$. Since the decoder observes only the channel outputs (and not the help), it forms the list:

$$\mathcal{L}(y^n) = \big\{m \in \mathcal{M} \colon \exists t \in \mathcal{T} \text{ s.t. } P_{Y^n|X^n,T}\big(y^n|f_m(t), t\big) > 0\big\}. \tag{25}$$

The collection of output sequences that cause the erasures-only decoder to produce an erasure is:

$$\mathcal{Y}_{\mathrm{er}} = \big\{y^n \in \mathcal{A}^n \colon |\mathcal{L}(y^n)| > 1\big\}. \tag{26}$$

The probability of erasure associated with the transmission of Message m with encoder help t is $P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big)$. On the modulo additive noise channel with rate-R_h encoder assistance, the erasures-only capacity and the Shannon capacity coincide and [4]:

$$C_{\text{e-o}}(R_\text{h}) = C_{\text{Sh}}(R_\text{h}) = \log|\mathcal{A}| - \big\{H(P_Z) - R_\text{h}\big\}^+. \tag{27}$$

We shall need the following slightly-stronger version of the achievability part of this result, where we swap the maximization over the messages with the expectation over the help:

Proposition 1. *Consider the modulo additive noise channel with rate-R_h encoder assistance. For any transmission rate R smaller than $C_{\text{e-o}}(R_\text{h})$ of (27)), there exists a sequence of rate-R transmission codes for which:*

$$\lim_{n\to\infty}\sum_{t\in\mathcal{T}} P_T(t)\max_{m\in\mathcal{M}} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) = 0. \tag{28}$$

A similar result holds for decoder assistance.

Proof. The proof is presented in Appendix A. It is based on the construction in [4], but with a slightly finer analysis. □

The coding scheme we propose is essentially that of [8]. We just need to account for the help. For each blocklength n, we start out with a transmission code of roughly $2^{nC_{\text{e-o}}(R_\text{h})}$ codewords for which (28) holds, and use Lemma 1 ahead to construct approximately $2^{2^{nC_{\text{e-o}}(R_\text{h})}}$ lightly-intersecting subsets of its message set. We then associate an IM with each of the subsets, with the understanding that to transmit an IM we pick uniformly at random one of the messages in the subset associated with it and transmit this message with the helper's assistance.

Lemma 1 ([7] Proposition 14)**.** *Let $\mathscr{X}$ be a finite set, and let $\lambda \in (0, \frac{1}{2})$ be given. If $\epsilon > 0$ is sufficiently small so that:*

$$\lambda\log\left(\frac{1}{\epsilon} - 1\right) > 2 \qquad \text{and} \qquad \frac{1}{\epsilon} > 6 \tag{29}$$

then there exist subsets $\mathscr{A}_1, \ldots, \mathscr{A}_\mathsf{N}$ of $\mathscr{X}$ such that for all distinct $i, j \in \{1, \ldots, \mathsf{N}\}$ the following hold:

(a) $$|\mathscr{A}_i| = \lfloor\epsilon\,|\mathscr{X}|\rfloor, \tag{30}$$

(b) $$|\mathscr{A}_i \cap \mathscr{A}_j| < \lambda\lfloor\epsilon\,|\mathscr{X}|\rfloor, \tag{31}$$

(c) $$\mathsf{N} \geq |\mathscr{X}|^{-1}\cdot 2^{\lfloor\epsilon\,|\mathscr{X}|\rfloor} - 1. \tag{32}$$

With the aid of this lemma, we can now prove the achievability of $C_{\text{e-o}}(R_\text{h})$.

Proof. Given an erasures-only encoder-assisted transmission code $\{(f_m)\}_{m\in\mathcal{M}}$ where $f_m\colon \mathcal{T} \to \mathcal{X}^n$, we apply Lemma 1 to the transmission message set $\mathcal{M}$ with:

$$\epsilon = \frac{1}{n^2+2} \qquad \text{and} \qquad \lambda = \frac{1}{\log n} \tag{33}$$

to infer, for large enough n, the existence of subsets $\mathcal{F}_1, \ldots, \mathcal{F}_\mathsf{N} \subseteq \mathcal{M}$ such that for all distinct $i, j \in \{1, \ldots, \mathsf{N}\}$ with $j \neq i$:

$$|\mathcal{F}_i| = \left\lfloor\frac{\mathsf{M}}{n^2+2}\right\rfloor \tag{34}$$

$$|\mathcal{F}_i \cap \mathcal{F}_j| < \frac{1}{\log n}\left\lfloor\frac{\mathsf{M}}{n^2+2}\right\rfloor \tag{35}$$

$$\mathsf{N} \geq \mathsf{M}^{-1} \cdot 2^{\left\lfloor \frac{\mathsf{M}}{n^2+2} \right\rfloor} - 1. \tag{36}$$

Note that (36) implies that:

$$\varliminf_{n\to\infty} \left(\frac{1}{n} \log\log \mathsf{N} - \frac{1}{n} \log \mathsf{M} \right) \geq 0. \tag{37}$$

To send IM i after obtaining the assistance $t(z^n)$, the encoder picks a random element M from $\mathcal{F}_i$ equiprobably and transmits $X^n = f_M\big(t(Z^n)\big)$, so:

$$P^i_{X^n|T}(x^n|t) = \frac{1}{|\mathcal{F}_i|} \sum_{m\in\mathcal{F}_i} \mathbb{1}\{x^n = f_m(t)\}. \tag{38}$$

To guarantee no missed detections, we set the acceptance region of i-th IM to be:

$$\mathcal{D}_i = \Big\{ y^n \in \mathcal{Y}^n \colon \exists (m,t) \in \mathcal{F}_i \times \mathcal{T} \text{ s.t. } P_{Y^n|X^n,T}\big(y^n|f_m(t),t\big) > 0 \Big\}. \tag{39}$$

It now remains to analyze the scheme's maximal false-alarm probability.

$$\begin{aligned}
p_{\text{FA,max}} &= \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} P^j_{Y^n|T=t}(\mathcal{D}_i) && (40)\\
&= \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \frac{1}{|\mathcal{F}_j|} \sum_{m\in\mathcal{F}_j} P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) && (41)\\
&= \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \frac{1}{|\mathcal{F}_j|} \Bigg[\sum_{m\in\mathcal{F}_j\setminus\mathcal{F}_i} P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) \\
&\qquad\qquad + \sum_{m\in\mathcal{F}_j\cap\mathcal{F}_i} P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) \Bigg] && (42)\\
&\leq \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \frac{1}{|\mathcal{F}_j|} \Bigg[\sum_{m\in\mathcal{F}_j\setminus\mathcal{F}_i} P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) + |\mathcal{F}_j \cap \mathcal{F}_i| \Bigg] && (43)\\
&\leq \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \frac{1}{|\mathcal{F}_j|} \Bigg[\sum_{m\in\mathcal{F}_j\setminus\mathcal{F}_i} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) + |\mathcal{F}_j \cap \mathcal{F}_i| \Bigg] && (44)\\
&< \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \left\{ \frac{1}{|\mathcal{F}_j|} \sum_{m\in\mathcal{F}_j\setminus\mathcal{F}_i} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) + \frac{\left\lfloor \frac{\mathsf{M}}{n^2+2} \right\rfloor}{|\mathcal{F}_j| \log n} \right\} && (45)\\
&\leq \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \left\{ \frac{|\mathcal{F}_j \setminus \mathcal{F}_i|}{|\mathcal{F}_j|} \max_{m\in\mathcal{M}} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) \right\} + \frac{1}{\log n} && (46)\\
&\leq \max_{i\in\mathcal{N}} \sum_{t\in\mathcal{T}} P_T(t) \max_{j\in\mathcal{N}\setminus\{i\}} \left\{ \max_{m\in\mathcal{M}} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) \right\} + \frac{1}{\log n} && (47)\\
&= \sum_{t\in\mathcal{T}} P_T(t) \max_{m\in\mathcal{M}} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) + \frac{1}{\log n} && (48)
\end{aligned}$$

where in (41) we expressed $P^j_{Y^n|T=t}(\mathcal{D}_i)$ as $P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big)$ using (7); in (42) we expressed $\mathcal{F}_j$ as the disjoint union of $\mathcal{F}_j \cap \mathcal{F}_i$ and $\mathcal{F}_j \setminus \mathcal{F}_i$; in (43) we used the trivial bound:

$$P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) \leq 1; \tag{49}$$

in (44) we used the fact that whenever $m \neq i$:

$$P_{Y^n|X^n,T}\big(\mathcal{D}_i|f_m(t),t\big) \leq P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big) \tag{50}$$

which holds because, by the definition of the set $\mathcal{Y}_{\text{er}}$, any output sequence y^n that contributes to the LHS of (50), i.e., that is in $\mathcal{D}_i$ with $P_{Y^n|X^n,T}\big(y^n|f_m(t),t\big) > 0$, must also be in $\mathcal{Y}_{\text{er}}$; in (45) we used (35); in (46) we replaced each term in the sum with the global maximum (over $m \in \mathcal{M}$) and used (34); in (47) we used the trivial bound $|\mathcal{F}_j \setminus \mathcal{F}_i| \leq |\mathcal{F}_j|$; and in (48) we could simplify the expression because the dependence on i and j is no longer.

The above construction demonstrates that every transmission scheme that drives $\sum_{t\in\mathcal{T}} P_T(t) \max_{m\in\mathcal{M}} P_{Y^n|X^n,T}\big(\mathcal{Y}_{\text{er}}|f_m(t),t\big)$ to zero induces a zero missed-identification scheme that drives the false-alarm probability to zero. Since the former exists for all rates up to $C_{\text{e-o}}(R_{\text{h}})$, we conclude, by (37), that $C_{\text{ID},0}(R_{\text{h}}) \geq C_{\text{e-o}}(R_{\text{h}})$. This, in turn, implies that $C_{\text{ID},0}(R_{\text{h}}) \geq C_{\text{Sh}}(R_{\text{h}})$ and hence concludes the achievability proof for encoder-assistance because, on the modulo additive noise channel, $C_{\text{e-o}}(R_{\text{h}}) = C_{\text{Sh}}(R_{\text{h}})$. □

3.2. Decoder Assistance

When, rather than to the encoder, the assistance is to the decoder, the transmission codewords are n-tuples in $\mathcal{A}^n$, and we denote the transmission codebook $\mathcal{C} = \{x^n(m)\}_{m\in\mathcal{M}}$. For the induced identification scheme we use the same message subsets as before, with IM i being transmitted by choosing uniformly at random a message M from the subset $\mathcal{F}_i$ and transmitting the codeword $x^n(M)$. To avoid any missed detections, we set the acceptance region corresponding to IM i and decoder assistance t to be:

$$\mathcal{D}_i(t) = \Big\{ y^n \in \mathcal{A}^n : \exists m \in \mathcal{F}_i \text{ s.t. } P_{Y^n,T|X^n}\big(y^n, t|x^n(m)\big) > 0 \Big\} \tag{51}$$

The analysis of the false-alarm probability is nearly identical to that with encoder assistance and is omitted.

4. Converse Part: Help Provided to Both Transmitter and Receiver

In this section we establish the converse for all the cases of interest by proving that the inequality:

$$C_{\text{ID}}(R_{\text{h}}) \leq \log|\mathcal{A}| - \big\{H(P_Z) - R_{\text{h}}\big\}^+ \tag{52}$$

holds even when the help is provided to both encoder and decoder. The RHS of (52) is the helper Shannon capacity, irrespective of whether the help is provided to the encoder, to the decoder, or to both [3] (Section V).

There are two main steps to the proof. The first addresses the conditional probabilities of the two types of testing errors conditional on a given description $T = t$. It relates the two to the conditional entropy of the noise given the description, namely, $H(Z^n|T = t)$. Very roughly, this corresponds to proving the converse part of the ID-capacity theorem for the channel whose noise is distributed according to the conditional distribution of Z^n given $T = t$. The difficulty in this step is that, given $T = t$, the noise is not memoryless, and the channel may not even be stable. Classical type-based techniques for proving the converse part of the ID-capacity theorem—such as those employed in [7] (Theorem 12), [13] (Section III), or [14] (Section III)—are therefore not applicable. Instead, we extend to the helper setting Watanabe's technique [12], which is inspired by the partial channel resolvability method introduced by Steinberg [15].

The second step in the proof addresses the unconditional error probabilities. This step is needed because, in the definition of achievability (see (13) and (14)), the error probabilities are averaged over the noise description t. We will show that, when the identification rate exceeds the Shannon capacity, there exists an IM i^* for which the sum of the two types of errors is large whenever the description t is in a subset $\mathcal{T}^*$ of $\mathcal{T}$ whose probability is bounded away from zero. This will imply that, for this IM i^*, the sum of the averaged probabilities of error is bounded away from zero, thus contradicting the achievability.

4.1. Additional Notation

Given a PMF P_X and a conditional PMF $P_{Y|X}$, we write $P_X \circ P_{Y|X}$ for the joint PMF that assigns the pair (x, y) the probability $P_X(x)\, P_{Y|X}(y|x)$. We use $I_{P \circ P_{Y|X}}(X; Y)$ to denote the mutual information between X and Y under the joint distribution $P \circ P_{Y|X}$. The product PMF of marginals P_X and P_Y is denoted $P_X \times P_Y$; it assigns (x, y) the probability $P_X(x)\, P_Y(y)$.

For the hypothesis testing problem of guessing whether some observation X was drawn $\sim P_X$ (the "null hypothesis") or $\sim Q_X$ (the "alternative hypothesis"), we use $K(\cdot|X)$ to denote a generic randomized test that, after observing $X = x$, guesses the null hypothesis ($X \sim P_X$) with probability $K(0|X = x)$ and the alternative ($X \sim Q_X$) with probability $K(1|X = x)$. (Here $K(0|X = x) + K(1|X = x) = 1$ for every $x \in \mathcal{X}$.) The type I error probability associated with $K(\cdot|X)$ is:

$$\lambda_1[K] = \sum_{x \in \mathcal{X}} P_X(x)\, K(1|x) \tag{53}$$

and the type II:

$$\lambda_2[K] = \sum_{x \in \mathcal{X}} Q_X(x)\, K(0|x). \tag{54}$$

For a given $0 < \epsilon < 1$ we define:

$$\beta_\epsilon(P_X, Q_X) = \inf_{K:\ \lambda_1[K] \le \epsilon} \lambda_2[K] \tag{55}$$

to be the least type-II error probability that can be achieved under the constraint that the type-I error probability does not exceed ϵ.

4.2. Conditional Missed-Detection and False-Alarm Probabilities

The following lemma follows directly from Watanabe's work [12].

Lemma 2 ([12] Theorem 1 and Corollary 2)**.** *Let $P_{Y^n|X^n,T=t}$ be the n-letter conditional distribution of the channel output sequence given that the noise description is $T = t$ and the input is X^n. For any λ_1, $\lambda_2 > 0$ with $\lambda_1 + \lambda_2 < 1$, any $0 < \eta < 1 - \lambda_1 - \lambda_2$, and any fixed $t \in \mathcal{T}$, the condition:*

$$p^i_{\mathrm{MD}}(t) + p^i_{\mathrm{FA}}(t) < \lambda_1 + \lambda_2, \qquad \forall i \in \mathcal{N} \tag{56}$$

implies:

$$\begin{aligned} \log\log \mathsf{N} \le \sup_{P \in \mathcal{P}(\mathcal{X}^n)} \inf_{Q \in \mathcal{P}(\mathcal{Y}^n)} - \log \beta_{\lambda_1+\lambda_2+\eta}\big(P \circ P_{Y^n|X^n,T=t}, P \times Q\big) \\ + \log\log |\mathcal{A}|^n + 2\log\left(\frac{1}{\eta}\right) + 2 \end{aligned} \tag{57}$$

and hence:

$$\frac{1}{n}\log\log \mathsf{N} \le \frac{1}{n} \sup_{P \in \mathcal{P}(\mathcal{X}^n)} \inf_{Q \in \mathcal{P}(\mathcal{Y}^n)} - \log \beta_{\lambda_1+\lambda_2+\eta}\big(P \circ P_{Y^n|X^n,T=t}, P \times Q\big) + \psi_n(\eta), \tag{58}$$

where:

$$\psi_n(\eta) = \frac{\log n}{n} + \frac{\log\log |\mathcal{A}|}{n} - \frac{2}{n}\log \eta + \frac{2}{n} \tag{59}$$

which—for any fixed $\eta > 0$—tends to 0 as n tends to ∞.

Substituting $P_{Y^n|X^n,T=t}$ for $P_{Y|X}$ in the following theorem will allow us to link the RHS of (57) with the conditional mutual information between X^n and Y^n given $t \in \mathcal{T}$. The theorem's proof was inspired by the proof of [16] (Theorem 8). See also [17] (Lemma 1).

Theorem 2. *Given any* $0 < \epsilon < 1$ *and any conditional PMF* $P_{Y|X}$*,*

$$\sup_{P\in\mathcal{P}(\mathcal{X})} \inf_{Q\in\mathcal{P}(\mathcal{Y})} -\log\beta_\epsilon\big(P\circ P_{Y|X}, P\times Q\big) \le \frac{\sup_{P\in\mathcal{P}(\mathcal{X})} I_{P\circ P_{Y|X}}(X;Y)+h(\epsilon)}{1-\epsilon} \tag{60}$$

where $h(\epsilon) \triangleq -\epsilon\log(\epsilon)-(1-\epsilon)\log(1-\epsilon)$ *is the binary entropy function.*

Proof. Applying the data-processing inequality for relative entropy to the binary hypothesis testing setting (see, e.g., [18] (Thm. 30.12.5)) we conclude that for any randomized test $K(\cdot|X)$,

$$D_{\text{bin}}\big(1-\lambda_1[K]\,\big\|\,\lambda_2[K]\big) \le D\big(P\circ P_{Y|X}\,\big\|\,P\times Q\big) \tag{61}$$

where:

$$D_{\text{bin}}(\alpha\,\|\,\beta) \triangleq \alpha\log\frac{\alpha}{\beta}+(1-\alpha)\log\frac{1-\alpha}{1-\beta} \tag{62}$$

denotes the binary divergence function. Since there exists a randomized test $K^*(\cdot|X)$ for which $\big(\lambda_1[K^\star],\lambda_2[K^\star]\big) = \big(\epsilon,\beta_\epsilon(P\circ P_{Y|X},P\times Q)\big)$ (see, e.g., [18] (Lemma 30.5.4 and Proposition 30.8.1) we can apply (61) to $K^\star(\cdot|X)$ to conclude that:

$$D_{\text{bin}}\big(1-\epsilon\,\big\|\,\beta_\epsilon(P\circ P_{Y|X},P\times Q)\big) \le D\big(P\circ P_{Y|X}\,\big\|\,P\times Q\big). \tag{63}$$

(The above existence also holds when $\beta_\epsilon(P\circ P_{Y|X},P\times Q))$ is zero, but for this case we can verify (63) directly by noting that in this case, since $\epsilon<1$, the RHS of (63) is $+\infty$.) The LHS of (63) can be lower bounded by lower-bounding the binary divergence function as:

$$D_{\text{bin}}\big(1-\epsilon\,\big\|\,\beta_\epsilon(P\circ P_{Y|X},P\times Q)\big) \ge -h(\epsilon)-(1-\epsilon)\log\beta_\epsilon(P\circ P_{Y|X},P\times Q). \tag{64}$$

It follows from (63) and (64) that:

$$-\log\beta_\epsilon(P\circ P_{Y|X},P\times Q) \le \frac{D\big(P\circ P_{Y|X}\,\big\|\,P\times Q\big)+h(\epsilon)}{1-\epsilon} \tag{65}$$

so the infimum over Q of the LHS is upper bounded by the infimum over Q on the RHS. The latter (for fixed $P\in\mathcal{P}(\mathcal{X})$) is achieved when Q is the Y-marginal of $P\circ P_{Y|X}$, a marginal that we denote P_Y:

$$\inf_Q D\big(P\circ P_{Y|X}\,\big\|\,P\times Q\big) = I_{P\circ P_{Y|X}}(X;Y). \tag{66}$$

This is a special case of a more general result on Rényi divergence [19] (Theorem II.2). Here we give a simple proof for K-L divergence:

$$\begin{aligned}
&D\big(P\circ P_{Y|X}\,\big\|\,P\times Q\big)\\
&\quad= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P\circ P_{Y|X}(x,y)\log\left(\frac{P\circ P_{Y|X}(x,y)}{P(x)Q(y)}\right) &(67)\\
&\quad= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P\circ P_{Y|X}(x,y)\log\left(\frac{P\circ P_{Y|X}(x,y)}{P(x)P_Y(y)}\frac{P_Y(y)}{Q(y)}\right) &(68)\\
&\quad= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P\circ P_{Y|X}(x,y)\log\left(\frac{P\circ P_{Y|X}(x,y)}{P(x)P_Y(y)}\right)+\sum_y P_Y(y)\log\frac{P_Y(y)}{Q(y)} &(69)\\
&\quad\ge I_{P\circ P_{Y|X}}(X;Y)+0 &(70)
\end{aligned}$$

with equality if and only if Q equals P_Y.

From (63), (64), and (66) we obtain:

$$\begin{aligned}
&\sup_{P\in\mathcal{P}(\mathcal{X})}\inf_{Q\in\mathcal{P}(\mathcal{Y})} -\log\beta_\epsilon(P\circ P_{Y|X},P\times Q)\\
&\quad\le \sup_{P\in\mathcal{P}(\mathcal{X})}\inf_{Q\in\mathcal{P}(\mathcal{Y})}\frac{D_{\text{bin}}(1-\epsilon\,\|\,\beta_\epsilon(P\circ P_{Y|X},P\times Q))+h(\epsilon)}{1-\epsilon} && (71)\\
&\quad\le \sup_{P\in\mathcal{P}(\mathcal{X})}\inf_{Q\in\mathcal{P}(\mathcal{Y})}\frac{D(P\circ P_{Y|X}\,\|\,P\times Q)+h(\epsilon)}{1-\epsilon} && (72)\\
&\quad= \sup_{P\in\mathcal{P}(\mathcal{X})}\frac{I_{P\circ P_{Y|X}}(X;Y)+h(\epsilon)}{1-\epsilon}. && (73)
\end{aligned}$$

□

Applying Lemma 2 and Theorem 2 to our channel when its law is conditioned on $T=t$ yields the following corollary.

Corollary 1. *On the* MMANC, *for any* $\lambda_1,\lambda_2>0$ *with* $\lambda_1+\lambda_2<1$, *any* $0<\eta<1-\lambda_1-\lambda_2$, *and any fixed* $t\in\mathcal{T}$, *the condition:*

$$p^i_{\text{MD}}(t)+p^i_{\text{FA}}(t)<\lambda_1+\lambda_2,\qquad\forall i\in\mathcal{N} \tag{74}$$

implies:

$$\frac{1}{n}\log\log\mathsf{N}\le\frac{\log|\mathcal{A}|-H(Z^n|T=t)/n}{1-\epsilon'}+\psi_n(\eta), \tag{75}$$

where $\epsilon'=\lambda_1+\lambda_2+\eta$.

Proof. Substituting $\mathcal{X}^n$ for $\mathcal{X}$, $\mathcal{Y}^n$ for $\mathcal{Y}$, $P_{Y^n|X^n,T=t}$ for $P_{Y|X}$, and ϵ' for ϵ in Theorem 2, we obtain:

$$\begin{aligned}
&\sup_{P\in\mathcal{P}(\mathcal{X}^n)}\inf_{Q\in\mathcal{P}(\mathcal{Y}^n)} -\log\beta_{\epsilon'}(P\circ P_{Y^n|X^n,T=t},P\times Q)\\
&\quad\le\frac{\sup_{P\in\mathcal{P}(\mathcal{X}^n)} I_{P\circ P_{Y^n|X^n,T=t}}(X^n;Y^n)+h(\epsilon')}{1-\epsilon'}. && (76)
\end{aligned}$$

Given $P\in\mathcal{P}(\mathcal{X}^n)$ and $P_{Y^n|X^n,T=t}$, the mutual information term in (76) can be upper-bounded as follows:

$$\begin{aligned}
I_{P\circ P_{Y^n|X^n,T=t}}(X^n;Y^n)&\le n\log|\mathcal{A}|-H_{P\circ P_{Y^n|X^n,T=t}}(Y^n|X^n,T=t) && (77)\\
&=n\log|\mathcal{A}|-\sum_{x^n}P(x^n|T=t)\,H(Y^n|X^n=x^n,T=t) && (78)\\
&=n\log|\mathcal{A}|-\sum_{x^n}P(x^n|T=t)\,H(Z^n|X^n=x^n,T=t) && (79)\\
&=n\log|\mathcal{A}|-H(Z^n|T=t). && (80)
\end{aligned}$$

Applying (76) and (80) to (58) in Lemma 2 establishes Corollary 1. □

4.3. Averaging over T

Corollary 1 deals with identification for a given fixed $T=t$, but our definition of achievability in (13) and (14) entails averaging over t, which we must thus study. We begin by lower-bounding the conditional entropy of the noise sequence Z^n given the assistance T:

$$\begin{aligned}
H(Z^n|T)&=H(Z^n,T)-H(T) && (81)\\
&\ge\{H(Z^n)-nR_{\text{h}}\}^+ && (82)
\end{aligned}$$

$$= n\left\{H(P_Z) - R_{\mathrm{h}}\right\}^{+}. \tag{83}$$

We next define, for every $\delta > 0$, the subset of descriptions:

$$\mathcal{T}^{*}(\delta) = \Big\{t \in \mathcal{T} \colon H(Z^n|T = t) \geq n\{H(P_Z) - R_{\mathrm{h}} - \delta\}^{+}\Big\}. \tag{84}$$

These are poor noise descriptions in the sense that, after they are revealed, the remaining uncertainty about the noise is still large. Key is that their probability is bounded away from zero. In fact, as we next argue:

$$P_T(\mathcal{T}^{*}(\delta)) \geq \begin{cases} \frac{\delta}{\log|\mathcal{A}| - H(P_Z) + R_{\mathrm{h}} + \delta} & \text{if } R_{\mathrm{h}} < H(P_Z) - \delta \\ 1 & \text{if } R_{\mathrm{h}} \geq H(P_Z) - \delta \end{cases} \tag{85}$$

where in the second case the probability is 1 because when $R_{\mathrm{h}} \geq H(P_Z) - \delta$ the condition appearing in the definition of $\mathcal{T}^{*}(\delta)$ in (84) translates to $H(Z^n|T = t) \geq 0$. As to the first case, we begin with (83) to obtain:

$$\begin{aligned} n(H(P_Z) - R_{\mathrm{h}}) &\leq H(Z^n|T) && (86)\\ &= \sum_{t \notin \mathcal{T}^{*}(\delta)} P_T(t)\, H(Z^n|T = t) + \sum_{t \in \mathcal{T}^{*}(\delta)} P_T(t)\, H(Z^n|T = t) && (87)\\ &\leq \Big(1 - P_T\big(\mathcal{T}^{*}(\delta)\big)\Big) \cdot n\big(H(P_Z) - R_{\mathrm{h}} - \delta\big) + P_T\big(\mathcal{T}^{*}(\delta)\big) \cdot n \log|\mathcal{A}| && (88) \end{aligned}$$

from which the first case of the bound in (85) follows. Here (87) follows from expressing $\mathcal{T}$ as the disjoint union of $\mathcal{T}^{*}(\delta)$ and $\mathcal{T} \setminus \mathcal{T}^{*}(\delta)$, and (88) follows from the definition of $\mathcal{T}^{*}(\delta)$ and the bound $H(Z^n|T = t) \leq n \log|\mathcal{A}|$.

Inequality (85) establishes that the probability of a poor description is lower bounded by a positive constant that does not depend on n. Using Corollary 1 for such t's will be the key to the converse.

Henceforth, we fix some sequence of identification codes of rate R exceeding $C_{\mathrm{Sh}}(R_{\mathrm{h}})$, i.e., satisfying $R > \log|\mathcal{A}| - \{H(P_Z) - R_{\mathrm{h}}\}^{+}$, and show that $p_{\mathrm{MD,max}} + p_{\mathrm{FA,max}}$ cannot tend to 0 as n tends to ∞. For such a rate R, there exist $R', \delta > 0$; a pair $\lambda_1, \lambda_2 > 0$ with $\lambda_1 + \lambda_2 < 1$; and some $\eta \in (0, 1 - \lambda_1 - \lambda_2)$ such that:

$$R > R' > \frac{\log|\mathcal{A}| - \{H(P_Z) - R_{\mathrm{h}} - \delta\}^{+}}{1 - \epsilon'} \tag{89}$$

where $\epsilon' \triangleq \lambda_1 + \lambda_2 + \eta < 1$. Fix such $R', \delta, \lambda_1, \lambda_2, \eta$, and ϵ'.

Since the inequality on R' in (89) is strict, and since $\psi_n(\eta)$ tends to zero with n, it follows that the inequality continues to hold also when we add $\psi_n(\eta)$ to the RHS provided that n is sufficiently large, i.e., that there exists some $n_0(\eta)$ such that:

$$R' > \frac{\log|\mathcal{A}| - \{H(P_Z) - R_{\mathrm{h}} - \delta\}^{+}}{1 - \epsilon'} + \psi_n(\eta), \qquad n \geq n_0(\eta). \tag{90}$$

It then follows from (90) and the definition of $\mathcal{T}^{*}(\delta)$ in (84) that, whenever $n \geq n_0(\eta)$, R' exceeds the RHS of (75):

$$R' > \frac{\log|\mathcal{A}| - n^{-1} H(Z^n|T = t)}{1 - \epsilon'}, \qquad \forall t \in \mathcal{T}^{*}(\delta). \tag{91}$$

Corollary 1 thus implies that, for $n > n_0(\eta)$:

$$\Big(\frac{1}{n} \log\log \mathsf{N} > R'\Big) \implies \Big(\forall t \in \mathcal{T}^{\star}\ \exists i(t) \in \mathcal{N} \text{ s.t. } p^{i}_{\mathrm{MD}}(t) + p^{i}_{\mathrm{FA}}(t) \geq \lambda_1 + \lambda_2\Big). \tag{92}$$

However, we need a stronger statement because, in the above, the IM i for which $p^i_{\text{MD}}(t) + p^i_{\text{FA}}(t) \geq \lambda_1 + \lambda_2$ depends on t, whereas in our definition of achievability we are averaging over T for fixed IM. The stronger result we will establish is that the condition on the LHS of (92) implies that, for all sufficiently large n, there exists some IM i^* (that does not depend on t) which performs poorly for *every* t in $\mathcal{T}^*(\delta)$, i.e., for which:

$$p^{i^*}_{\text{MD}}(t) + p^{i^*}_{\text{FA}}(t) \geq \lambda_1 + \lambda_2, \qquad \forall t \in \mathcal{T}^*(\delta). \tag{93}$$

That is, we will show that for sufficiently large n:

$$\left(\frac{1}{n}\log\log \mathsf{N} > R'\right) \implies \left(\exists i \in \mathcal{N} \text{ s.t } \min_{t \in \mathcal{T}^*(\delta)} \left\{p^i_{\text{MD}}(t) + p^i_{\text{FA}}(t)\right\} \geq \lambda_1 + \lambda_2\right). \tag{94}$$

To this end, define for each $t \in \mathcal{T}^*(\delta)$:

$$\mathcal{N}(t) = \left\{i \in \mathcal{N} \colon p^i_{\text{MD}}(t) + p^i_{\text{FA}}(t) < \lambda_1 + \lambda_2\right\} \tag{95}$$

and consider the identification code that results when we restrict our code to the IMs in $\mathcal{N}(t)$ (while keeping the same acceptance regions). Applying Corollary 1 to this restricted code using (91), we obtain that:

$$\frac{1}{n}\log\log|\mathcal{N}(t)| < R', \qquad \forall t \in \mathcal{T}^*(\delta). \tag{96}$$

Consequently,

$$\left|\bigcup_{t \in \mathcal{T}^*(\delta)} \mathcal{N}(t)\right| \leq \sum_{t \in \mathcal{T}^*(\delta)} |\mathcal{N}(t)| \leq 2^{nR_{\text{h}}}\, 2^{2^{nR'}} \tag{97}$$

where the second inequality holds by (96) and the fact that $\mathcal{T}^*(\delta)$ is contained in $\mathcal{T}$, and the latter's cardinality is $2^{nR_{\text{h}}}$.

Since $R' < R$ (89), there exists some $n_1(R, R', R_{\text{h}})$ such that:

$$2^{nR_{\text{h}}} 2^{2^{nR'}} < 2^{2^{nR}}, \qquad n \geq n_1(R, R', R_{\text{h}}). \tag{98}$$

We can use this to upper-bound the RHS of (97) to obtain that, for $n \geq \max\{n_0(\eta), n_1(R, R', R_{\text{h}})\}$:

$$\left|\bigcup_{t \in \mathcal{T}^*(\delta)} \mathcal{N}(t)\right| < \mathsf{N}. \tag{99}$$

The complement (in $\mathcal{N}$) of the union on the LHS of (99) is thus not empty, which proves the existence of some $i^\star \in \mathcal{N}$ for which (93) holds.

With $i^\star$ in hand, the converse follows from the fact that the probability that T is in $\mathcal{T}^*(\delta)$ is bounded away from zero (85), because for every $n \geq \max\{n_0(\eta), n_1(R, R', R_{\text{h}})\}$:

$$\begin{aligned}
p_{\text{MD,max}} + p_{\text{FA,max}} &= \max_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} P_T(t)\, p^i_{\text{MD}}(t) + \max_{i \in \mathcal{N}} \sum_{t \in \mathcal{T}} P_T(t)\, p^i_{\text{FA}}(t) && (100)\\
&\geq \sum_{t \in \mathcal{T}} P_T(t) \left(p^{i^*}_{\text{MD}}(t) + p^{i^*}_{\text{FA}}(t)\right) && (101)\\
&\geq \sum_{t \in \mathcal{T}^*(\delta)} P_T(t) \left(p^{i^*}_{\text{MD}}(t) + p^{i^*}_{\text{FA}}(t)\right) && (102)\\
&\geq \sum_{t \in \mathcal{T}^*(\delta)} P_T(t) \cdot (\lambda_1 + \lambda_2) && (103)\\
&= P_T(\mathcal{T}^*(\delta)) \cdot (\lambda_1 + \lambda_2) && (104)
\end{aligned}$$

where (100) follows from the definitions in (13) and (14); in (101) we replaced the maximum with the IM i^*; and (103) follows from (93). Thus, any code of rate $R > \log |\mathcal{A}| - \{H(P_Z) - R_\text{h}\}^+$ with large enough n must have $p_{\text{MD,max}} + p_{\text{FA,max}} \geq P_T(\mathcal{T}^*) \cdot (\lambda_1 + \lambda_2)$, and the latter is bounded away from zero. This concludes the proof of the converse part.

Author Contributions: Writing—original draft preparation, A.L. and B.N.; writing—review and editing, A.L. and B.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Christian Deppe and Johannes Rosenberger who read a preprint of this paper and provided them with helpful comments. They also thank the guest editor and the anonymous reviewers for their insightful comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MMANC	Memoryless modulo-additive noise channel
IID	Identical independent distribution
IM	Identification message
LHS	Left hand side
RHS	Right hand side

Appendix A. Proof of Proposition 1

Proof. As in [4], the code construction entails time-sharing between two schemes: a "zero-rate help scheme" corresponding to help of zero rate, and a "high-rate help scheme" corresponding to help of a rate exceeding the noise entropy. In the former, the help comprises one bit, indicating whether or not the noise is typical. We denote this help $T^{(\text{z})}$ and assume that it takes values in the set $\mathcal{H} = \{\tau, \alpha\}$, with $T^{(\text{z})} = \tau$ indicating that the noise is typical and $T^{(\text{z})} = \alpha$ that it is atypical (when the help is to the encoder, the helper additionally provides the encoder with the description of one noise sample in order to enable the encoder to convey $T^{(\text{z})}$ to the decoder error free).

When the help is of high rate, we denote it $T^{(\text{h})}$. It has two parts, that we denote $T^{(\text{h})}_{\text{t/a}}$ and $T^{(\text{h})}_{\text{d}}$, so $T^{(\text{h})} = (T^{(\text{h})}_{\text{t/a}}, T^{(\text{h})}_{\text{d}})$. The first part, $T^{(\text{h})}_{\text{t/a}}$, indicates whether or not the noise is typical and hence takes values in $\mathcal{H}$. The second part, $T^{(\text{h})}_{\text{d}}$, describes the noise (perfectly) when the latter is typical, and is null otherwise (as above, when the help is to the encoder, the helper additionally provides the encoder with the description of one noise sample in order to enable the encoder to convey $T^{(\text{h})}_{\text{t/a}}$ to the decoder error free). The help in the time-sharing scheme, which we denote T, comprises the help in the zero-rate part and the help in the high-rate part:

$$T = \left(T^{(\text{z})}, T^{(\text{h})}\right) \tag{A1}$$

The duty cycle is chosen so that the rate of T be R_h (or the entropy of the noise, if the latter is smaller than R_h). We assume throughout that $R < \log |\mathcal{A}|$.

The transmission code derived in [4] has two salient properties:

- In the high-rate scheme, conditional on $T^{(\text{h})}_{\text{t/a}} = \tau$ (i.e., on the noise being typical and that it can therefore be perfectly described by $T^{(\text{h})}_{\text{d}}$), no erasures are declared.
- In the zero-rate scheme, conditional on $T^{(\text{z})} = \tau$ (i.e., on the noise being typical), the maximal (over the messages) probability of erasure is upper bounded by some ϵ_n tending to zero.

(To guarantee the second property, the code is constructed—as in [4]—using random coding and we then expurgate half the codewords to obtain a code whose maximal probability of erasure is smaller than ϵ. The asserted property then follows by bounding, for each message, the conditional probability of erasure given that the noise is typical by the ratio of the unconditional probability of erasure to the probability that the noise is typical.)

We next analyze the time-sharing scheme. We focus on the case where $0 < R_{\mathrm{h}} \leq H(P_Z)$. The remaining cases, where $R_{\mathrm{h}} = 0$ or $R_{\mathrm{h}} > H(P_Z)$ are very similar, except that they require no time sharing. We use the superscript (h) for quantities occurring in the high-rate help phase, and the superscript (z) for those in the zero-rate phase. For example, $m^{(\mathrm{h})}, \mathbf{X}^{(\mathrm{h})}, \mathbf{Y}^{(\mathrm{h})}, T^{(\mathrm{h})}$ are the message, input sequence, output sequence, and help in the high-rate help phase; and the set of output sequences causing an erasure in this phase is denoted $\mathcal{Y}_{\mathrm{er}}^{(\mathrm{h})}$. The set of outputs causing an erasure in the time-sharing scheme is:

$$\mathcal{Y}_{\mathrm{er}} = \left\{ y^n \colon \mathbf{y}^{(\mathrm{h})} \in \mathcal{Y}_{\mathrm{er}}^{(\mathrm{h})} \text{ or } \mathbf{y}^{(\mathrm{z})} \in \mathcal{Y}_{\mathrm{er}}^{(\mathrm{z})} \right\}. \tag{A2}$$

For the time-sharing scheme we now have:

$$\begin{aligned}
&\sum_{t \in \mathcal{T}} P_T(t) \max_{m \in \mathcal{M}} P_{Y^n|X^n,T}\left(\mathcal{Y}_{\mathrm{er}} \middle| f_m(t), t\right) \\
&\quad \leq \sum_{t \in \mathcal{T}} P_T(t) \max_{m \in \mathcal{M}} \Bigg[P_{\mathbf{Y}^{(\mathrm{h})}|\mathbf{Y}^{(\mathrm{h})},T^{(\mathrm{h})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{h})} \middle| f_{m^{(\mathrm{h})}}^{(\mathrm{h})}(t^{(\mathrm{h})}), t^{(\mathrm{h})}\right) \\
&\qquad\qquad\qquad\qquad + P_{\mathbf{Y}^{(\mathrm{z})}|\mathbf{X}^{(\mathrm{z})},T^{(\mathrm{z})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{z})} \middle| f_{m^{(\mathrm{z})}}^{(\mathrm{z})}(t^{(\mathrm{z})}), t^{(\mathrm{z})}\right) \Bigg] && \text{(A3)} \\
&\quad \leq \sum_{t \in \mathcal{T}} P_T(t) \max_{m^{(\mathrm{h})} \in \mathcal{M}^{(\mathrm{h})}} P_{\mathbf{Y}^{(\mathrm{h})}|\mathbf{X}^{(\mathrm{h})},T^{(\mathrm{h})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{h})} \middle| f_{m^{(\mathrm{h})}}^{(\mathrm{h})}(t^{(\mathrm{h})}), t^{(\mathrm{h})}\right) \\
&\qquad + \sum_{t \in \mathcal{T}} P_T(t) \max_{m^{(\mathrm{z})} \in \mathcal{M}^{(\mathrm{z})}} P_{\mathbf{Y}^{(\mathrm{z})}|\mathbf{X}^{(\mathrm{z})},T^{(\mathrm{z})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{z})} \middle| f_{m^{(\mathrm{z})}}^{(\mathrm{z})}(t^{(\mathrm{z})}), t^{(\mathrm{z})}\right) && \text{(A4)} \\
&\quad = \sum_{t^{(\mathrm{h})})} P_{T^{(\mathrm{h})}}(t^{(\mathrm{h})}) \max_{m^{(\mathrm{h})} \in \mathcal{M}^{(\mathrm{h})}} P_{\mathbf{Y}^{(\mathrm{h})}|\mathbf{X}^{(\mathrm{h})},T^{(\mathrm{h})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{h})} \middle| f_{m^{(\mathrm{h})}}^{(\mathrm{h})}(t^{(\mathrm{h})}), t^{(\mathrm{h})}\right) \\
&\qquad + \sum_{t^{(\mathrm{z})}} P_{T^{(\mathrm{z})}}(t^{(\mathrm{z})}) \max_{m^{(\mathrm{z})} \in \mathcal{M}^{(\mathrm{z})}} P_{\mathbf{Y}^{(\mathrm{z})}|\mathbf{X}^{(\mathrm{z})},T^{(\mathrm{z})}}\left(\mathcal{Y}_{\mathrm{er}}^{(\mathrm{z})} \middle| f_{m^{(\mathrm{z})}}^{(\mathrm{z})}(t^{(\mathrm{z})}), t^{(\mathrm{z})}\right) && \text{(A5)} \\
&\quad \leq P\big(T_{\mathrm{t/a}}^{(\mathrm{h})} = \alpha\big) + P\big(T^{(\mathrm{z})} = \alpha\big) \cdot 1 + P\big(T^{(\mathrm{z})} = \tau\big) \cdot \epsilon_n && \text{(A6)} \\
&\quad \leq P\big(T_{\mathrm{t/a}}^{(\mathrm{h})} = \alpha\big) + P\big(T^{(\mathrm{z})} = \alpha\big) + \epsilon_n && \text{(A7)}
\end{aligned}$$

which establishes the proposition, because the RHS tends to zero. Here (A3) follows from (A2) and the union-of-events bound; (A4) holds (in this case with equality) because the maximum of a sum is upper bounded by the sum of the maxima; and (A6) holds by the aforementioned salient properties of the code construction. □

References

1. Kim, Y.H. Capacity of a class of deterministic relay channels. *IEEE Trans. Inf. Theory* **2008**, *54*, 1328–1329. [CrossRef]
2. Bross, S.I.; Lapidoth, A.; Marti, G. Decoder-assisted communications over additive noise channels. *IEEE Trans. Commun.* **2020**, *68*, 4150–4161. [CrossRef]
3. Lapidoth, A.; Marti, G. Encoder-assisted communications over additive noise channels. *IEEE Trans. Inf. Theory* **2020**, *66*, 6607–6616. [CrossRef]
4. Lapidoth, A.; Marti, G.; Yan, Y. Other helper capacities. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Virtual, 12–20 July 2021; pp. 1272–1277. [CrossRef]
5. Lapidoth, A.; Yan, Y. The listsize capacity of the Gaussian channel with decoder assistance. *Entropy* **2022**, *24*, 29. [CrossRef] [PubMed]
6. Merhav, N. On error exponents of encoder-assisted communication systems. *IEEE Trans. Inf. Theory* **2021**, *67*, 7019–7029. [CrossRef]
7. Ahlswede, R.; Dueck, G. Identification via channels. *IEEE Trans. Inf. Theory* **1989**, *35*, 15–29. [CrossRef]

8. Ahlswede, R.; Cai, N.; Zhang, Z. Erasure, list, and detection zero-error capacities for low noise and a relation to identification. *IEEE Trans. Inf. Theory* **1996**, *42*, 55–62. [CrossRef]
9. Steinberg, Y.; Merhav, N. Identification in the presence of side information with application to watermarking. *IEEE Trans. Inf. Theory* **2001**, *47*, 1410–1422. [CrossRef]
10. Ahlswede, R.; Dueck, G. Identification in the presence of feedback—A discovery of new capacity formulas. *IEEE Trans. Inf. Theory* **1989**, *35*, 30–36. [CrossRef]
11. Wiese, M.; Labidi, W.; Deppe, C.; Boche, H. Identification over Additive Noise Channels in the Presence of Feedback. *IEEE Trans. Inf. Theory* **2022**, 1. [CrossRef]
12. Watanabe, S. Minimax converse for identification via channels. *IEEE Trans. Inf. Theory* **2022**, *68*, 25–34. [CrossRef]
13. Han, T.; Verdú, S. New results in the theory of identification via channels. *IEEE Trans. Inf. Theory* **1992**, *38*, 14–25. [CrossRef]
14. Bracher, A.; Lapidoth, A. Identification via the broadcast channel. *IEEE Trans. Inf. Theory* **2017**, *63*, 3480–3501. [CrossRef]
15. Steinberg, Y. New converses in the theory of identification via channels. *IEEE Trans. Inf. Theory* **1998**, *44*, 984–998. [CrossRef]
16. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]
17. Rosenberger, J.; Ibrahim, A.; Bash, B.A.; Deppe, C.; Ferrara, R.; Pereg, U. Capacity Bounds for Identification with Effective Secrecy. *arXiv* **2023**, arXiv:2306.14792.
18. Lapidoth, A. *A Foundation in Digital Communication*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2017.
19. Aishwarya, G.; Madiman, M. Remarks on Rényi versions of conditional entropy and mutual information. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; pp. 1117–1121. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.