Article

The Listsize Capacity of the Gaussian Channel with Decoder Assistance

Amos Lapidoth * [ID] **and Yiming Yan**

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

Abstract: The listsize capacity is computed for the Gaussian channel with a helper that—cognizant of the channel-noise sequence but not of the transmitted message—provides the decoder with a rate-limited description of said sequence. This capacity is shown to equal the sum of the cutoff rate of the Gaussian channel without help and the rate of help. In particular, zero-rate help raises the listsize capacity from zero to the cutoff rate. This is achieved by having the helper provide the decoder with a sufficiently fine quantization of the normalized squared Euclidean norm of the noise sequence.

Keywords: bit pipe; cutoff rate; decoder assistance; Gaussian channel; helper; listsize capacity

1. Introduction

The order-ρ listsize capacity $C_{\text{list}}^{(\rho)}$ of a channel is the supremum of the coding rates for which there exist codes guaranteeing the large-blocklength convergence to one of the ρ-th moment of the cardinality of the list of messages that, given the received output sequence, have positive a posteriori probability. It is zero for the Gaussian channel because, on this channel, no codeword is ruled out by any received sequence so said list contains all the messages. Here we derive this capacity for the Gaussian channel with a helper that observes the noise sequence and describes it to the decoder using a rate-limited noise-free bit pipe; see Figure 1.

Figure 1. Gaussian channel with decoder assistance.

We show that the listsize capacity $C_{\text{list}}^{(\rho)}(R_{\text{h}})$ is then the sum of bit-pipe's rate R_{h} and the order-ρ cutoff rate $R_{\text{cutoff}}(\rho)$ of the Gaussian channel without a helper

$$C_{\text{list}}^{(\rho)}(R_{\text{h}}) = R_{\text{cutoff}}(\rho) + R_{\text{h}}. \tag{1}$$

The latter's definition is similar to that of the listsize capacity, but with the list now comprising only those messages that are a posteriori at least as likely as the transmitted one. As we shall see, for the Gaussian channel with average power P, noise-variance N, and corresponding signal-to-noise ratio (SNR) A $\triangleq$ P/N,

$$R_{\text{cutoff}}(\rho) = R_0(\rho), \tag{2}$$

where

$$R_0(\rho) = \frac{1}{2}\ln\frac{1}{2}\left(1 + \frac{A}{1+\rho} + \sqrt{\left(1 - \frac{A}{1+\rho}\right)^2 + \frac{4A}{(1+\rho)^2}}\right)$$
$$+ \frac{1+\rho}{2\rho}\frac{1}{2}\left(1 + \frac{A}{1+\rho} - \sqrt{\left(1 - \frac{A}{1+\rho}\right)^2 + \frac{4A}{(1+\rho)^2}}\right)$$
$$+ \frac{1}{2\rho}\ln\frac{1}{2}\left(1 - \frac{A}{1+\rho} + \sqrt{\left(1 - \frac{A}{1+\rho}\right)^2 + \frac{4A}{(1+\rho)^2}}\right) \tag{3}$$

(in nats) is a function that plays a prominent role in the analysis of the Reliability Function of said channel (Section 7.4 in [1]), [2]. That analysis does not, however, carry over directly to our setting because it deals with error exponents and not lists.

It is interesting to note that (1) also holds when the help rate R_h is zero: the number of help bits required to increase the listsize capacity from zero to $R_{\text{cutoff}}(\rho)$ is sublinear in the blocklength. In fact, as we shall see, all it takes is a sufficiently fine quantization of the normalized squared Euclidean norm of the noise sequence.

The relation (1) is reminiscent of the analogous result on the erasures-only capacity $C_{\text{e-o}}(R_h)$ of the Gaussian channel with a rate-R_h helper (Remark 10 in [3]), namely, that

$$C_{\text{e-o}}(R_h) = C + R_h, \tag{4}$$

where C denotes the Shannon capacity of the Gaussian channel (without help) (Theorem 9.1.1 in [4]), and $C_{\text{e-o}}(R_h)$ is the erasures-only capacity, which is defined like $C_{\text{list}}^{(\rho)}(R_h)$ but with the requirement on the ρ-th moment of the list replaced by the requirement that the list be of size 1 with probability tending to one. (The Gaussian erasures-only capacity with a helper is given by the RHS of (4) irrespective of whether the assistance is provided to the encoder or decoder.) The latter result in turn is reminiscent of the analogous result on the Shannon capacity with a helper $C(R_h)$ [5–8]

$$C(R_h) = C + R_h. \tag{5}$$

In proving (1), we shall focus on the "direct part," i.e., that the right-hand side (RHS) of (1) is achievable. The "converse," that no rate exceeding the RHS of (1) is achievable, is omitted because it follows directly from (Remark 4 in [3]): There it is shown that this is true even if, given the received sequence and the provided help, the list contains only a *subset* of the messages that are of positive a posteriori probability, namely, those that are a posteriori at least as likely as the transmitted message.

The listsize capacity is relevant, for example, when the message set corresponds to tasks [9] and the transmitted message corresponds to one that must be performed by the decoder with absolute certainty. To ensure this, the decoder must perform all the tasks in the list of tasks that are not ruled out by the received sequence. (In addition to the transmitted task, other tasks need not but may be performed.) The ρ-th moment of the list's size then measures the receiver's average effort.

Results on the listsize capacity and the erasures-only capacity of general discrete memoryless channels (DMCs) in the absence of help are scarce. Noteworthy exceptions are the results of Pinsker and Sheverdjaev [10], Csiszár and Narayan [11], and Telatar [12], that provide sufficient conditions for the erasures-only capacity to equal the Shannon capacity and for the listsize capacity to equal the cutoff rate. Asymptotic results on the erasures-only capacity in the low-noise regime can be found in [13,14]. Once noiseless feedback is introduced, the problems become more tractable [15–17].

The rest of the paper is organized as follows. Section 2 describes our set-up and presents the main result. Section 3 contains some classical and some new observations

regarding Gallager's E_0 function and its modification. Section 4 derives the cutoff rate of the Gaussian channel without help and proves (2). Section 5 describes and analyzes a coding scheme that proves the direct part of (1).

2. The Main Result

A power-P blocklength-n encoder $f^{(n)}$ for a message set $\mathcal{M}$ is a mapping

$$f^{(n)} \colon \mathcal{M} \to \mathbb{R}^n \tag{6}$$

that maps each message $m \in \mathcal{M}$ to an n-tuple $f^{(n)}(m)$ whose Euclidean norm $\|f^{(n)}(m)\|$ satisfies

$$\|f^{(n)}(m)\|^2 \leq n\mathsf{P}, \quad m \in \mathcal{M}. \tag{7}$$

We sometimes use $\mathbf{x}_m$ to denote $f^{(n)}(m)$, and $x_{m,k}$ to denote the k-th component of $\mathbf{x}_m$, so

$$f^{(n)}(m) = \mathbf{x}_m = (x_{m,1}, \ldots, x_{m,n}). \tag{8}$$

The encoder is said to be of rate R if the cardinality of $\mathcal{M}$ is e^{nR}, in which case we often assume that $\mathcal{M} = \{1, \ldots, e^{nR}\}$. (We ignore the fact that e^{nR} need not be an integer; this issue washes out in the large-n asymptotics we study.)

When a message $m \in \mathcal{M}$ is sent over the discrete-time additive Gaussian noise channel using the encoder $f^{(n)}$, the channel produces the random vector $\mathbf{Y} \in \mathbb{R}^n$ whose k-th component Y_k is

$$Y_k = x_{m,k} + Z_k, \quad k = 1, \ldots, n, \tag{9}$$

where $\{Z_k\}$ are independent and identically distributed (IID) zero-mean Gaussians of variance N. We assume that N is positive and use $w(y|x)$ to denote the density of the channel's output when its input is x, i.e., the mean-x variance-N Gaussian density

$$w(y|x) = \frac{1}{\sqrt{2\pi\mathsf{N}}} e^{-\frac{(y-x)^2}{2\mathsf{N}}}, \quad x, y \in \mathbb{R}, \tag{10}$$

which we extend to n-tuples in a memoryless fashion:

$$w(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{n} w(y_k|x_k), \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{11}$$

For convenience, we define

$$\mathsf{A} = \frac{\mathsf{P}}{\mathsf{N}}. \tag{12}$$

Given an output sequence $\mathbf{y}$ and a message m, we define the "at-least-as-likely list"

$$\mathcal{L}(m, \mathbf{y}) = \{m' \in \mathcal{M} : w(\mathbf{y}|\mathbf{x}_{m'}) \geq w(\mathbf{y}|\mathbf{x}_m)\}. \tag{13}$$

Assuming, as we do, that the messages are a priori equally likely, this list comprises the messages that, given the output sequence $\mathbf{y}$, are a posteriori at least as likely as m.

If a message M, drawn equiprobably from $\mathcal{M}$, is transmitted over the channel with a resulting received sequence $\mathbf{Y}$, then the cardinality of the at-least-as-likely list is a random positive integer, and we denote its ρ-th moment $\mathbb{E}\big[|\mathcal{L}(M, \mathbf{Y})|^\rho\big]$:

$$\mathbb{E}\big[|\mathcal{L}(M, \mathbf{Y})|^\rho\big] = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \int w(\mathbf{y}|\mathbf{x}_m) |\mathcal{L}(m, \mathbf{y})|^\rho \, d\nu(\mathbf{y}), \tag{14}$$

where $\nu(\cdot)$ denotes the Lebesgue measure on $\mathbb{R}^n$.

For a given $\rho > 0$, we define the order-ρ cutoff rate $R_{\text{cutoff}}(\rho)$ as the supremum of the rates R for which there exists a sequence of rate-R power-P blocklength-n encoders $\{f^{(n)}\}$ satisfying

$$\lim_{n \to \infty} \mathbb{E}\big[|\mathcal{L}(M, \mathbf{Y})|^{\rho}\big] = 1. \tag{15}$$

Theorem 1. *The order-ρ cutoff rate $R_{\text{cutoff}}(\rho)$ of the additive Gaussian noise channel equals $R_0(\rho)$ of* (3).

Proof. See Section 4. $\square$

A $\mathcal{T}_n$-valued description of the noise sequence $\mathbf{Z} = (Z_1, \dots, Z_n)$ is a mapping

$$\phi^{(n)} \colon \mathbb{R}^n \to \mathcal{T}_n \tag{16}$$

with the understanding that $\phi^{(n)}(\mathbf{Z})$, which we denote T, is the description of $\mathbf{Z}$. We say that a sequence $\{\phi^{(n)}\}$ of descriptions is of rate R_{h} (nats) if

$$\lim_{n \to \infty} \frac{1}{n} \ln|\mathcal{T}_n| = R_{\text{h}}. \tag{17}$$

Suppose now that, in addition to the received sequence $\mathbf{Y}$, the receiver is also presented with the description $T = \phi^{(n)}(\mathbf{Z})$ of the noise, and that, based on the two, it forms the "remotely-plausible list" $\mathcal{L}(\mathbf{Y}, T)$ comprising the messages that have positive a posteriori probability given the two:

$$\mathcal{L}(\mathbf{y}, t) = \big\{m \in \mathcal{M} : \phi^{(n)}(\mathbf{y} - \mathbf{x}_m) = t\big\}. \tag{18}$$

Given $\rho > 0$, the listsize capacity $C_{\text{list}}^{(\rho)}(R_{\text{h}})$ with rate-R_{h} decoder assistance is the supremum of the rates R for which there exists a sequence of rate-R power-P blocklength-n encoders $\{f^{(n)}\}$ and a sequence $\{\phi^{(n)}\}$ of descriptions of rate R_{h} such that

$$\lim_{n \to \infty} \mathbb{E}\Big[\big|\mathcal{L}\big(\mathbf{Y}, \phi^{(n)}(\mathbf{Z})\big)\big|^{\rho}\Big] = 1. \tag{19}$$

Theorem 2. *On the Gaussian channel, the listsize capacity with rate-R_{h} decoder assistance $C_{\text{list}}^{(\rho)}(R_{\text{h}})$ is given by*

$$C_{\text{list}}^{(\rho)}(R_{\text{h}}) \quad = \quad R_{\text{cutoff}}(\rho) + R_{\text{h}} \tag{20}$$

where $R_{\text{cutoff}}(\rho)$ is the order-ρ cutoff rate of the channel (without assistance) as given in (2) *and* (3).

Proof. The "converse," that (19) cannot be achieved when the rate exceeds the RHS of (20), follows from (Remark 4 in [3]). The "direct part," describing a coding scheme that achieves (19) with rates approaching the RHS of (20), is proved in Section 5. $\square$

3. Preliminaries

Given $\rho \geq 0$ and any probability measure Q on $\mathbb{R}$, Gallager's E_0 function for our channel is defined as [1]

$$E_0(\rho, Q) = -\ln \int_{y \in \mathbb{R}} \left(\int_{x \in \mathbb{R}} w(y|x)^{\frac{1}{1+\rho}} \, dQ(x) \right)^{1+\rho} d\nu(y), \tag{21}$$

where $\nu(\cdot)$ is now the Lebesgue measure on $\mathbb{R}$. The result of maximizing $E_0(\rho, Q)$ over all Q under which $\mathbb{E}[X^2] \leq \text{P}$, is denoted $E_0^*(\rho)$:

$$E_0^*(\rho) = \sup_{Q \colon \int x^2 \, dQ(x) \leq \text{P}} E_0(\rho, Q). \tag{22}$$

The multi-letter extension of E_0 is

$$E_0^{(n)}(\rho, Q^{(n)}) = -\frac{1}{n} \ln \int_{\mathbf{y} \in \mathbb{R}^n} \left(\int_{\mathbf{x} \in \mathbb{R}^n} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \, dQ^{(n)}(\mathbf{x}) \right)^{1+\rho} d\nu(\mathbf{y}), \tag{23}$$

where $Q^{(n)}$ is a probability measure on $\mathbb{R}^n$; the integrals are over $\mathbb{R}^n$; the channel $w(\mathbf{y}|\mathbf{x})$ is defined in (11). Similarly,

$$E_0^{(n),*}(\rho) = \sup_{Q^{(n)} : \int \|\mathbf{x}\|^2 \, dQ^{(n)}(\mathbf{x}) \le n\mathsf{P}} E_0^{(n)}(\rho, Q^{(n)}). \tag{24}$$

Given probability measures $Q^{(m)}$ on $\mathbb{R}^m$ and $Q^{(n)}$ on $\mathbb{R}^n$ that satisfy the power constraints $\mathbb{E}[\|\mathbf{X}\|^2] \le m\mathsf{P}$ and $\mathbb{E}[\|\mathbf{X}\|^2] \le n\mathsf{P}$ respectively, the product measure $Q^{(m)} \times Q^{(n)}$ on $\mathbb{R}^{m+n}$ satisfies the power constraint $\mathbb{E}[\|\mathbf{X}\|^2] \le (m+n)\mathsf{P}$ and

$$(m+n) \, E_0^{(m+n)}(\rho, Q^{(m)} \times Q^{(n)}) = m \, E_0^{(m)}(\rho, Q^{(m)}) + n \, E_0^{(n)}(\rho, Q^{(n)}) \tag{25}$$

because

$$(m+n) \, E_0^{(m+n)}(\rho, Q^{(m)} \times Q^{(n)})$$

$$= -\ln \int_{\mathbf{y} \in \mathbb{R}^{m+n}} \left(\int_{\mathbf{x} \in \mathbb{R}^{m+n}} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \, d(Q^{(m)} \times Q^{(n)})(\mathbf{x}) \right)^{1+\rho} d\nu(\mathbf{y}) \tag{26}$$

$$= m \, E_0^{(m)}(\rho, Q^{(m)}) + n \, E_0^{(n)}(\rho, Q^{(n)}). \tag{27}$$

The sequence $\{ n \, E_0^{(n),*}(\rho) \}$ is thus superadditive, and Feket's Subadditive lemma implies that $\{ E_0^{(n),*}(\rho) \}$ converges to its supremum:

$$\lim_{n \to \infty} E_0^{(n),*}(\rho) = \sup_n E_0^{(n),*}(\rho). \tag{28}$$

We shall later see (cf. (55) ahead) that

$$\frac{1}{\rho} \sup_n E_0^{(n),*}(\rho) = R_0(\rho), \tag{29}$$

where $R_0(\rho)$ is defined in (3).

We shall also need Gallager's modified E_0 function. To highlight its relation to the unmodified function, which is quite general, we shall use $g(x)$ for x^2 and $g(\mathbf{x})$ for $\|\mathbf{x}\|^2$. We shall also replace P with Γ.

Given some $\rho \ge 0$, some probability distribution Q on $\mathbb{R}$ under which $\mathbb{E}[g(X)] \le \Gamma$, and some $r \ge 0$, the modified Gallager's E_0 function $E_{0,\mathrm{m}}(\rho, Q, r)$ is defined as

$$E_{0,\mathrm{m}}(\rho, Q, r) = -\ln \int_{y \in \mathbb{R}} \left(\int_{x \in \mathbb{R}} e^{r(g(x) - \Gamma)} w(y|x)^{\frac{1}{1+\rho}} \, dQ(x) \right)^{1+\rho} d\nu(y). \tag{30}$$

We shall also be interested in the maximum of $E_{0,\mathrm{m}}(\rho, Q, r)$ over both Q and r. We distinguish between two cases depending on whether $\mathbb{E}[g(X)] \le \Gamma$ holds strictly or not. In the former case we only allow r to be zero, whereas in the latter case it can be any non-negative number. We thus define

$$E_{0,\mathrm{m}}^*(\rho, Q) = \begin{cases} \sup_{r \ge 0} E_{0,\mathrm{m}}(\rho, Q, r), & \text{if } \int g(x) \, dQ(x) = \Gamma, \\ E_0(\rho, Q), & \text{if } \int g(x) \, dQ(x) < \Gamma, \end{cases} \tag{31}$$

and

$$E_{0,\mathrm{m}}^{**}(\rho) = \sup_{Q:\ \int g(x)\,\mathrm{d}Q(x)\leq\Gamma} E_{0,\mathrm{m}}^{*}(\rho,Q). \tag{32}$$

The next proposition provides a lower bound on $\lim E_0^{(n),*}(\rho)$.

Proposition 1. *Any probability distribution Q on $\mathbb{R}$ under which $g(X)$ is of finite second moment and of expectation Γ provides the lower bound*

$$\lim_{n\to\infty} E_0^{(n),*}(\rho) \geq E_{0,\mathrm{m}}^{*}(\rho,Q). \tag{33}$$

Proof. Let Q be any input distributions Q under which $g(X)$ is of finite second moment and $\mathbb{E}[g(X)] = \Gamma$. For each $n \in \mathbb{N}$, let $Q^{(n)}$ be the conditional distribution of the n-fold product distribution $Q^{\times n}$ given the event $\{\mathbf{X} \in \mathcal{A}_n\}$, where

$$\mathcal{A}_n = \left\{ \mathbf{x} \in \mathbb{R}^n : n\Gamma - \delta < g(\mathbf{x}) \leq n\Gamma \right\} \tag{34}$$

where $\delta > 0$ is some positive constant. Thus, for every Borel measurable subset $\mathcal{B}$ of $\mathbb{R}^n$,

$$Q^{(n)}(\mathcal{B}) = \frac{1}{\mu} Q^{\times n}(\mathcal{B} \cap \mathcal{A}_n), \quad \mathcal{B} \in \mathscr{B}(\mathbb{R}^n) \tag{35}$$

with

$$\mu = Q^{\times n}(\mathcal{A}_n). \tag{36}$$

For any $r \geq 0$, we can upper-bound the Radon–Nykodim derivative of $Q^{(n)}$ with respect to product distribution $Q^{\times n}$ as follows:

$$\frac{\mathrm{d}Q^{(n)}}{\mathrm{d}Q^{\times n}} = \frac{1}{\mu} \mathrm{I}\{\mathbf{x} \in \mathcal{A}_n\} \tag{37}$$

$$\leq \frac{1}{\mu} e^{r(g(\mathbf{x})-n\Gamma+\delta)} \tag{38}$$

$$= \frac{1}{\mu} e^{r\delta} e^{r(g(\mathbf{x})-n\Gamma)} \tag{39}$$

where $\mathrm{I}\{\text{statement}\}$ equals 1 if the statement is true and else 0. Using this bound on the Radon–Nykodim derivative we obtain:

$$E_0^{(n)}(\rho, Q^{(n)}) = -\frac{1}{n} \ln \int_{\mathbf{y}\in\mathbb{R}^n} \left(\int_{\mathbf{x}\in\mathbb{R}^n} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \, \mathrm{d}Q^{(n)}(\mathbf{x}) \right)^{1+\rho} \mathrm{d}\nu(\mathbf{y}) \tag{40}$$

$$\geq -\frac{1}{n} \ln \int_{\mathbf{y}\in\mathbb{R}^n} \left(\int_{\mathbf{x}\in\mathbb{R}^n} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \cdot \frac{1}{\mu} e^{r\delta} e^{r(g(\mathbf{x})-n\Gamma)} \, \mathrm{d}Q^{\times n}(\mathbf{x}) \right)^{1+\rho} \mathrm{d}\nu(\mathbf{y}) \tag{41}$$

$$= -\frac{1+\rho}{n} \ln \frac{e^{r\delta}}{\mu} - \ln \int_{y\in\mathbb{R}} \left(\int_{x\in\mathbb{R}} e^{r(g(x)-\Gamma)} w(y|x)^{\frac{1}{1+\rho}} \, \mathrm{d}Q(x) \right)^{1+\rho} \mathrm{d}\nu(y) \tag{42}$$

$$= -\frac{1+\rho}{n} \ln \frac{e^{r\delta}}{\mu} + E_{0,\mathrm{m}}(\rho, Q, r). \tag{43}$$

By the Central Limit Theorem, μ tends to $1/2$ as n tends to infinity, so (43) implies that

$$\liminf_{n\to\infty} \frac{1}{n} E_0^{(n)}(\rho, Q^{(n)}) \geq E_{0,\mathrm{m}}(\rho, Q, r). \tag{44}$$

Taking the supremum of the RHS over all $r \geq 0$, establishes that

$$\liminf_{n\to\infty} \frac{1}{n} E_0^{(n)}(\rho, Q^{(n)}) \geq E_{0,\mathrm{m}}^{*}(\rho, Q) \tag{45}$$

and hence, by (24), proves (33). □

We next turn to upper-bounding $\lim E_0^{(n),*}(\rho)$.

Proposition 2. *If the probability distribution $Q^{(n)}$ on $\mathbb{R}^n$ is such that $\mathbb{E}[g(\mathbf{X})] \leq n\Gamma$, and if f_{R} is any density on $\mathbb{R}$, then*

$$E_0^{(n)}(\rho, Q^{(n)}) \quad \leq \quad \sup_{P:\ \int g(x)\,\mathrm{d}P(x)\leq\Gamma} -(1+\rho)\int_{x\in\mathbb{R}} \ln\left(\int_{y\in\mathbb{R}} w(y|x)^{\frac{1}{1+\rho}} f_{\mathrm{R}}(y)^{\frac{\rho}{1+\rho}}\,\mathrm{d}y\right)\mathrm{d}P(x) \qquad (46)$$

and, consequently,

$$\lim_{n\to\infty} E_0^{(n),*}(\rho) \quad \leq \quad \sup_{P:\ \int g(x)\,\mathrm{d}P(x)\leq\Gamma} -(1+\rho)\int_{x\in\mathbb{R}} \ln\left(\int_{y\in\mathbb{R}} w(y|x)^{\frac{1}{1+\rho}} f_{\mathrm{R}}(y)^{\frac{\rho}{1+\rho}}\,\mathrm{d}y\right)\mathrm{d}P(x). \qquad (47)$$

Proof. The proof is based on Proposition 2 in [18], which implies that for every density $f_{\mathrm{R}}^{(n)}$ on $\mathbb{R}^n$ and any probability measure $Q^{(n)}$ on $\mathbb{R}^n$,

$$n\, E_0^{(n)}(\rho, Q^{(n)}) \leq -(1+\rho)\int_{\mathbf{x}\in\mathbb{R}^n} \ln\left(\int_{\mathbf{y}\in\mathbb{R}^n} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} f_{\mathrm{R}}^{(n)}(\mathbf{y})^{\frac{\rho}{1+\rho}}\,\mathrm{d}\mathbf{y}\right)\mathrm{d}Q^{(n)}(\mathbf{x}). \qquad (48)$$

Applying this inequality to the product density

$$f_{\mathrm{R}}^{(n)}(\mathbf{y}) = \prod_{i=1}^{n} f_{\mathrm{R}}(y_i), \qquad (49)$$

where f_{R} is a density on $\mathbb{R}$, and using the product form of the channel (11), we obtain that for any density f_{R} on $\mathbb{R}$

$$E_0^{(n)}(\rho, Q^{(n)}) \leq -\frac{1}{n}(1+\rho)\sum_{i=1}^{n}\int_{x_i\in\mathbb{R}} \ln\left(\int_{y\in\mathbb{R}} w(y|x_i)^{\frac{1}{1+\rho}} f_{\mathrm{R}}(y)^{\frac{\rho}{1+\rho}}\,\mathrm{d}y\right)\mathrm{d}Q_i^{(n)}(x_i) \qquad (50)$$

$$= -(1+\rho)\int_{x\in\mathbb{R}} \ln\left(\int_{y\in\mathbb{R}} w(y|x)^{\frac{1}{1+\rho}} f_{\mathrm{R}}(y)^{\frac{\rho}{1+\rho}}\,\mathrm{d}y\right)\mathrm{d}\bar{Q}(x), \qquad (51)$$

where $Q_i^{(n)}$ is the i-th marginal of $Q^{(n)}$, and $\bar{Q}$ is the probability measure on $\mathbb{R}$ defined by

$$\bar{Q} = \frac{1}{n}\sum_{i=1}^{n} Q_i^{(n)}. \qquad (52)$$

Observe that if $\mathbb{E}[g(\mathbf{X})] \leq n\Gamma$ under $Q^{(n)}$, then $\mathbb{E}[g(X)] \leq \Gamma$ under $\bar{Q}$. This observation and (51) establish (46). Since (46) holds for all n, (47) must also hold. □

4. The Cutoff Rate of the Gaussian Channel

In this section, we prove Theorem 1. Since scaling the output does not change the cutoff rate, we will assume WLOG that the noise variance is 1 and the transmit power is A; see (12). Thus,

$$w(y|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-x)^2}{2}}, \quad x, y \in \mathbb{R}, \qquad (53)$$

and each codeword $\mathbf{x}_m$ satisfies

$$\|\mathbf{x}_m\|^2 \leq n\mathsf{A}. \qquad (54)$$

4.1. Computing $\lim E_0^{(n),*}(\rho)$

Here we shall establish that on the Gaussian channel (53)

$$\lim_{n\to\infty} E_0^{(n),*}(\rho) = \rho R_0(\rho) = E_{0,\mathrm{m}}^*(\rho, Q_G), \tag{55}$$

where $R_0(\rho)$ is defined in (3), and Q_G is the zero-mean variance-A Gaussian distribution. To this end, we shall derive matching upper and lower bounds on the limit. We begin with the former.

4.1.1. Upper-Bounding $\lim E_0^{(n),*}(\rho)$

We show that on the channel (10)

$$\lim_{n\to\infty} E_0^{(n),*}(\rho) \le \rho R_0(\rho). \tag{56}$$

The proof is based on Proposition 2 with the density f_R corresponding to a centered Gaussian of variance σ^2, where

$$\sigma^2 = \frac{A}{(1+\rho)\beta} + 1 \tag{57}$$

and

$$\beta = \frac{1}{2}\left(1 - \frac{A}{1+\rho} + \sqrt{\left(1 - \frac{A}{1+\rho}\right)^2 + \frac{4A}{(1+\rho)^2}}\right). \tag{58}$$

Evaluating the RHS of (47) for this density, we obtain

$$\sup_{P:\,\mathbb{E}[X^2]\le A} -(1+\rho)\int_{x\in\mathbb{R}}\ln\int_{y\in\mathbb{R}} w(y|x)^{\frac{1}{1+\rho}} f_R(y)^{\frac{\rho}{1+\rho}}\,\mathrm{d}y\,\mathrm{d}P(x) \tag{59}$$

$$= \sup_{P:\,\mathbb{E}[X^2]\le A} -(1+\rho)\int_{x\in\mathbb{R}}\ln\int_{y\in\mathbb{R}} \frac{1}{(\sqrt{2\pi})^{\frac{1}{1+\rho}}} e^{-\frac{(y-x)^2}{2(1+\rho)}} \frac{1}{(\sqrt{2\pi\sigma^2})^{\frac{\rho}{1+\rho}}} e^{-\frac{y^2\rho}{2\sigma^2(1+\rho)}}\,\mathrm{d}y\,\mathrm{d}P(x) \tag{60}$$

$$= \sup_{P:\,\mathbb{E}[X^2]\le A} -(1+\rho)\int_{x\in\mathbb{R}}\ln\left(\sqrt{\frac{2\pi(1+\rho)^2}{\rho}}\sqrt{\sigma^2}^{\frac{1}{1+\rho}}\right)\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{x^2}{2\sigma_1^2}}\,\mathrm{d}P(x) \tag{61}$$

$$= \sup_{P:\,\mathbb{E}[X^2]\le A} -(1+\rho)\ln\left(\sqrt{\frac{(1+\rho)^2}{\rho}}\sqrt{\sigma^2}^{\frac{1}{1+\rho}}\frac{1}{\sqrt{\sigma_1^2}}\right) + (1+\rho)\int_{x\in\mathbb{R}}\frac{x^2}{2\sigma_1^2}\,\mathrm{d}P(x) \tag{62}$$

$$= -(1+\rho)\ln\left(\sqrt{\frac{(1+\rho)^2}{\rho}}\sqrt{\sigma^2}^{\frac{1}{1+\rho}}\frac{1}{\sqrt{\sigma_1^2}}\right) + (1+\rho)\frac{A}{2\sigma_1^2} \tag{63}$$

$$= \frac{1+\rho}{2}\frac{A}{\sigma_1^2} + \frac{1+\rho}{2}\ln\sigma_1^2 - \frac{1}{2}\ln\sigma^2 - \frac{1+\rho}{2}\ln\frac{(1+\rho)^2}{\rho} \tag{64}$$

where in (61) we defined

$$\sigma_1^2 \triangleq 1 + \rho + \frac{\sigma^2(1+\rho)}{\rho} \tag{65}$$

$$= \frac{A}{\rho\beta} + \frac{(1+\rho)^2}{\rho}. \tag{66}$$

To conclude the proof, it remains to show that the RHS of (64) coincides with $\rho R_0(\rho)$. To this end, observe that some basic algebra reveals that

$$\beta\left(\beta - 1 + \frac{A}{1+\rho}\right) = \frac{A}{(1+\rho)^2} \tag{67}$$

and

$$\left(\beta + \frac{A}{1+\rho}\right)(1-\beta) = \frac{A\rho}{(1+\rho)^2}. \tag{68}$$

Therefore, the first term in (64) can be rewritten as

$$\frac{1+\rho}{2}\frac{A}{\sigma_1^2} = \frac{1+\rho}{2}\frac{A}{\frac{A}{\rho\beta} + \frac{(1+\rho)^2}{\rho}} = \frac{1+\rho}{2}\frac{A\rho}{(1+\rho)^2}\frac{\beta}{\frac{A}{(1+\rho)^2} + \beta} \tag{69}$$

$$= \frac{A\rho}{2(1+\rho)}\frac{1}{\beta + \frac{A}{1+\rho}} = \frac{(1+\rho)(1-\beta)}{2}, \tag{70}$$

and the remaining terms rewritten as

$$\frac{1+\rho}{2}\ln\sigma_1^2 - \frac{1}{2}\ln\sigma^2 - \frac{1+\rho}{2}\ln\frac{(1+\rho)^2}{\rho}$$

$$= \frac{1+\rho}{2}\ln\left(\frac{A}{\rho\beta} + \frac{(1+\rho)^2}{\rho}\right) - \frac{1}{2}\ln\left(\frac{A}{(1+\rho)\beta} + 1\right) - \frac{1+\rho}{2}\ln\frac{(1+\rho)^2}{\rho} \tag{71}$$

$$= \frac{1+\rho}{2}\ln\left(\frac{A}{(1+\rho)^2\beta} + 1\right) - \frac{1}{2}\ln\left(\frac{A}{(1+\rho)\beta} + 1\right) \tag{72}$$

$$= \frac{\rho}{2}\ln\left(\beta + \frac{A}{1+\rho}\right) + \frac{1}{2}\ln\beta. \tag{73}$$

The sum equals to $\rho R_0(\rho)$.

4.1.2. Lower-Bounding $\lim E_0^{(n),*}(\rho)$

To lower-bound $\lim E_0^{(n),*}(\rho)$, we shall use Proposition 1 with Q chosen as a centered variance-A Gaussian distribution Q_{G}. For this probability distribution Gallager calculated $E_{0,\mathrm{m}}^*(\rho, Q_{\mathrm{G}})$ (Section 7.4 in [1]). He showed that for any $\rho > 0$,

$$E_{0,\mathrm{m}}^*(\rho, Q_{\mathrm{G}}) = \rho R_0(\rho), \tag{74}$$

where $R_0(\rho)$ is defined in (3). Using this result and Proposition 1 we obtain

$$\lim_{n\to\infty} E_0^{(n),*}(\rho) \geq E_{0,\mathrm{m}}^*(\rho, Q_{\mathrm{G}}) \tag{75}$$

$$= \rho R_0(\rho). \tag{76}$$

4.2. The Mapping $\rho \mapsto R_0(\rho)$ Is Monotonically Decreasing

For the purpose of proving the achievability of $R_0(\rho)$, we will need the fact that it is monotonically decreasing in ρ. In view of (55), it suffices to show that, for every $n \in \mathbb{N}$, the mapping $\rho \mapsto \rho^{-1}E_0^{(n),*}(\rho)$ is monotonically decreasing. In view of (24), the latter will follow once we establish the monotonicity of $\rho \mapsto \rho^{-1}E_0^{(n)}(\rho, Q^{(n)})$ for any fixed $Q^{(n)}$. Since $E_0^{(n)}(\rho, Q^{(n)})$ evaluates to zero at $\rho = 0$, this monotonicity can be established by showing that the mapping $\rho \mapsto E_0^{(n)}(\rho, Q^{(n)})$ is concave. This is established in (Appendix 5.B in [1]). (That appendix deals with finite alphabets, but the proof goes through also to our case.)

4.3. Achievability of $R_0(\rho)$

The achievability of $R_0(\rho)$ will be proved using a random-coding argument. Let Q be the zero-mean variance-A Gaussian distribution, let $\delta > 0$ be a positive constant, and let $Q^{(n)}$ be the distribution on $\mathbb{R}^n$ defined in (35) and (36). Draw the codewords $\{\mathbf{X}_m\}_{m=1,\dots,e^{nR}}$ of a blocklength-n random codebook independently, each according to $Q^{(n)}$, so $\|\mathbf{X}_m\|^2 \leq n\mathsf{A}$ with probability 1 for every $m \in \mathcal{M}$. By symmetry, $\mathbb{E}\left[|\mathcal{L}(m, \mathbf{Y})|^\rho\right]$ (where the expectation is

over the random choice of codebook and on the channel behavior) does not depend on m. Consequently,

$$\mathbb{E}\left[e^{-nR}\sum_{m\in\mathcal{M}}|\mathcal{L}(m,\mathbf{Y})|^{\rho}\right] = \mathbb{E}\left[|\mathcal{L}(1,\mathbf{Y})|^{\rho}\right], \tag{77}$$

and if we establish that $\mathbb{E}\left[|\mathcal{L}(1,\mathbf{Y})|^{\rho}\right]$ tends to 1, it will follow by the random-coding argument that there exists a codebook for which the LHS of (77)—with the expectation now over the channel behavior only—tends to 1.

Defining

$$B_m(\mathbf{x}_1,\mathbf{y}) = \mathbb{1}\{w(\mathbf{y}|\mathbf{X}_m) \geq w(\mathbf{y}|\mathbf{x}_1)\}, \quad \mathbf{x}_1, \mathbf{y} \in \mathbb{R}^n, \tag{78}$$

we can express the RHS of (77) as

$$\mathbb{E}\left[|\mathcal{L}(1,\mathbf{Y})|^{\rho}\right] = \mathbb{E}\left[\left(1 + \sum_{m\neq 1} B_m(\mathbf{X}_1,\mathbf{Y})\right)^{\rho}\right], \tag{79}$$

and we seek to show that

$$\lim_{n\to\infty}\mathbb{E}\left[\left(1 + \sum_{m\neq 1} B_m(\mathbf{X}_1,\mathbf{Y})\right)^{\rho}\right] = 1. \tag{80}$$

To this end, we shall need the following lemma.

Lemma 1. *Let $\{Z_n\}$ be a sequence of random variables taking values in $\mathbb{N}$, and let $\rho > 0$ be fixed. The following two conditions are then equivalent:*

(i) $\mathbb{E}[(1 + Z_n)^{\rho}] = 1 + o(1)$
(ii) $\mathbb{E}[Z_n^{\rho}] = o(1)$

where $o(1)$ tends to zero as n tends to infinity. Thus

$$\left(\lim_{n\to\infty}\mathbb{E}[(1 + Z_n)^{\rho}] = 1\right) \iff \left(\lim_{n\to\infty}\mathbb{E}[Z_n^{\rho}] = 0\right). \tag{81}$$

Proof. The implication (ii) $\Longrightarrow$ (i) follows by noting for any $z \in \mathbb{N}$ and $\rho > 0$

$$(1 + z)^{\rho} \leq 1 + 2^{\rho}z^{\rho}, \tag{82}$$

so

$$\mathbb{E}[(1 + Z_n)^{\rho}] \leq 1 + 2^{\rho}\,\mathbb{E}[Z_n^{\rho}]. \tag{83}$$

As for the implication (i) $\Longrightarrow$ (ii), note that any $y \in \mathbb{N}$ and $\rho > 0$

$$(1 + y)^{\rho} \geq y^{\rho} + \mathbb{1}\{y = 0\}, \tag{84}$$

so

$$\mathbb{E}[(1 + Z_n)^{\rho}] \geq \mathbb{E}[Z_n^{\rho}] + \Pr[Z_n = 0]. \tag{85}$$

The implication is now established by noting that (i) implies that $\Pr[Z_n = 0] \to 1$ because, by Markov's inequality (and the strict positivity of ρ),

$$\Pr[Z_n \neq 0] = \Pr[(1 + Z_n)^{\rho} - 1 \geq 2^{\rho} - 1] \tag{86}$$

$$\leq \frac{\mathbb{E}[(1 + Z_n)^{\rho}] - 1}{2^{\rho} - 1}. \tag{87}$$

□

In light of the above lemma, to establish (80) it suffices to show that

$$\lim_{n \to \infty} \mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{X}_1, \mathbf{Y}) \right)^{\rho} \right] = 0, \tag{88}$$

i.e., that

$$\lim_{n \to \infty} \mathbb{E}\left[\mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \right)^{\rho} \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right] = 0, \tag{89}$$

where the outer expectation is over $\mathbf{X}_1$ and $\mathbf{Y}$.

A related expectation—but one where it is the conditional expectation that is raised to the ρ-th power—is studied in the following lemma:

Lemma 2. *If $\rho > 0$ and $R < R_0(\rho)$, then*

$$\lim_{n \to \infty} \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]^{\rho} \right] = 0. \tag{90}$$

Proof. See Appendix A. □

To establish (88) using this lemma, we distinguish between two cases depending on whether $0 < \rho \leq 1$ or $\rho > 1$. In the former case $x \mapsto x^{\rho}$ is concave, so Jensen's inequality implies that

$$\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \right)^{\rho} \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right] \leq \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]^{\rho} \right], \tag{91}$$

which, together with Lemma 2, implies (88) whenever $R < R_0(\rho)$.

Suppose now that $\rho > 1$. Conditional on the transmitted codeword $\mathbf{x}_1$ and the output $\mathbf{y}$, the random variables $\{B_m\}_{m \neq 1}$ are IID Bernoulli, with B_m determined by $\mathbf{X}_m$. We can thus use Rosenthal's technique (Lemma 5.10 in [19]), [20] to obtain

$$\mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \right)^{\rho} \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]$$

$$\leq 2^{\rho^2} \max\left\{ \mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]^{\rho}, \mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right\} \tag{92}$$

$$\leq 2^{\rho^2} \left(\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]^{\rho} + \mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right). \tag{93}$$

Taking the expectation over $\mathbf{X}_1$ and $\mathbf{Y}$ yields

$$\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \right)^{\rho} \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right] \tag{94}$$

$$\leq 2^{\rho^2} \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{y}) \,\middle|\, \mathbf{Y} = \mathbf{y} \right]^{\rho} \right] + 2^{\rho^2} \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{y}) \,\middle|\, \mathbf{Y} = \mathbf{y} \right] \right] \tag{95}$$

$$\leq 2^{\rho^2} \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right]^{\rho} \right] + 2^{\rho^2} \mathbb{E}\left[\mathbb{E}\left[\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}) \,\middle|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y} \right] \right]. \tag{96}$$

The first term on the RHS can be treated using the lemma. The second—but for the 2^{ρ^2} constant—is the one encountered when ρ is 1. Since by Section 4.2, $R_0(\rho) \leq R_0(1)$ (because $\rho > 1$ for the case at hand), it too tends to zero when $R < R_0(\rho)$.

4.4. No Rate Exceeding $R_0(\rho)$ Is Achievable

To show the converse, we need Arıkan's lower bound on guessing [21].

Fix any sequence of rate-R blocklength-n codebooks $\{\mathcal{C}_n\}$ satisfying the cost constraint. For any $n \in \mathbb{N}$, let

$$Q^{(n)}(\mathbf{x}) = \begin{cases} \frac{1}{|\mathcal{C}_n|} & \text{if } \mathbf{x} \in \mathcal{C}_n, \\ 0 & \text{otherwise} \end{cases} \tag{97}$$

be the induced probability distribution on $\mathbb{R}^n$. Since the codebook satisfies the cost constraint, $\mathbb{E}[\|\mathbf{X}\|^2] \leq n\mathrm{A}$ under $Q^{(n)}$.

Given $\mathbf{y}$, list the messages $m \in \mathcal{M}$ in decreasing order of likelihood $w(\mathbf{y}|\mathbf{x}_m)$ (resolving ties arbitrarily, e.g., ranking low numerical values of m higher), and let $G(m|\mathbf{y})$ denote the ranking of the message m in this list. Note that

$$|\mathcal{L}(m, \mathbf{y})| \geq G(m|\mathbf{y}), \tag{98}$$

where the inequality can be strict because there may be messages that are in $\mathcal{L}(m, \mathbf{y})$ because they have the same likelihood as m, and that are yet ranked lower than m by $G(\cdot|\mathbf{y})$ because of the way ties are resolved. It follows from this inequality that the ρ-th moment of $|\mathcal{L}(M, \mathbf{Y})|$ cannot tend to one unless the ρ-th moment of $G(M|\mathbf{Y})$ does. By Arıkan's guessing inequality [21],

$$\mathbb{E}\left[G(M|\mathbf{Y})^\rho\right] \geq (1 + nR)^{-\rho} \cdot \exp\left(n\rho R - nE_0^{(n)}(\rho, Q^{(n)})\right), \tag{99}$$

so the ρ-th moment of $G(M|\mathbf{Y})$ can tend to one only if

$$\rho R \leq \liminf_{n \to \infty} E_0^{(n)}(\rho, Q^{(n)}). \tag{100}$$

From this, the converse now follows using (24) and (55) because

$$\liminf_{n \to \infty} E_0^{(n)}(\rho, Q^{(n)}) \leq \lim_{n \to \infty} E_0^{(n),*}(\rho) \tag{101}$$

$$= \rho R_0(\rho). \tag{102}$$

5. The Direct Part of Theorem 2

In this section we prove the direct part of Theorem 2: when the decoder can be provided with a rate-R_h description of the noise, the convergence (19) can be achieved at all transmission rates below $R_0(\rho) + R_\mathrm{h}$. As noted earlier, the converse follows directly from (Remark 4 in [3]).

Our proof treats the cases $R_\mathrm{h} = 0$ and $R_\mathrm{h} > 0$ separately. As in Section 4, we assume that the channel is normalized to having noise variance 1 and transmit power A.

5.1. Case 1: $R_\mathrm{h} = 0$

The analogous result for the modulo-additive channel was proved in [3] by having the helper provide the decoder with a lossless description of the type of the noise sequence. Since this type fully specifies the a posteriori probability of the transmitted message, the decoder's remotely-plausible-with-this-help list $\mathcal{L}(\mathbf{Y}, T)$ contains only messages whose a posteriori probability is equal to that of the correct message. It is therefore a subset of the at-least-as-likely list $\mathcal{L}(M, \mathbf{Y})$ (without help) and hence of smaller-or-equal ρ-th moment. Consequently, any rate that allows the latter to tend to one, also allows the former to tend to one.

On the Gaussian channel the likelihood $w(\mathbf{y}|\mathbf{x}_m)$ is specified by the normalized squared Euclidean norm of the noise sequence $\|\mathbf{z}\|^2/n$. The latter, however, cannot be described at zero rate with infinite precision. This motivates us to quantize it and have the quantized

version be the zero-rate help. The result will then follow by considering the high-resolution limit of the achievable rates. For this purpose, a uniform quantizer will do.

Given some large $M > 0$ (which determines the overload region) and some large K (corresponding to the number of quantization cells), we partition the interval $[0, M]$ into K subintervals, each of length $\Delta = M/K$. The helper, upon observing the noise sequence $\mathbf{Z}$, produces

$$T = \phi^{(n)}(\mathbf{Z}) = \begin{cases} \lfloor \|\mathbf{Z}\|^2/(n\Delta) \rfloor & \text{if } \|\mathbf{Z}\|^2/n < M, \\ K & \text{otherwise} \end{cases}. \tag{103}$$

The constant M, which does not depend on the blocklength n, is chosen large enough to guarantee that the large-deviation probability of overload $\Pr\left[\|\mathbf{Z}\|^2/n \geq M\right]$ decay sufficiently fast in n so that the contribution of the overload to the ρ-th moment of the list be negligible, even if an overload results in the list containing all e^{nR} codewords:

$$\lim_{n \to \infty} e^{n\rho R} \cdot \Pr\left[n^{-1}\|\mathbf{Z}\|^2 \geq M\right] = 0. \tag{104}$$

(Upper bounds on the tail of the χ^2 distribution show, for example, that for $R < R_0(\rho)$, the choice $M = \max\{2, 20\rho R_0(\rho)\}$ will do.) Since the help takes values in the finite set $\mathcal{T}_n = \{0, 1, \ldots, K\}$, where K does not depend on the blocklength, it is of zero rate.

As in Section 4.3, we consider a random codebook $\{\mathbf{X}_m\}_{m=1,\ldots,e^{nR}}$ whose codewords are drawn independently from the conditional Gaussian distribution, i.e., from $Q^{(n)}$ defined in (35) and (36) with Q being Q_G, the centered variance-A Gaussian distribution. Using the same symmetry arguments, we also assume that the transmitted message is $m = 1$ and study the ρ-th moment of the list under this assumption. Defining

$$V_m(\mathbf{x}_1, \mathbf{y}) = \mathbb{1}\{\phi^{(n)}(\mathbf{y} - \mathbf{X}_m) = \phi^{(n)}(\mathbf{y} - \mathbf{x}_1)\}, \quad \mathbf{x}_1, \mathbf{y} \in \mathbb{R}^n, \tag{105}$$

we can express the ρ-th moment of the remotely-plausible list when $m = 1$ as

$$\mathbb{E}\left[|\mathcal{L}(\mathbf{Y}, T)|^\rho\right] = \mathbb{E}\left[\left(1 + \sum_{m \neq 1} V_m(\mathbf{X}_1, \mathbf{Y})\right)^\rho\right]. \tag{106}$$

In view of Lemma 1, we need to prove that

$$\lim_{n \to \infty} \mathbb{E}\left[\left(\sum_{m \neq 1} V_m(\mathbf{X}_1, \mathbf{Y})\right)^\rho\right] = 0, \tag{107}$$

where the expectation is over both the random choice of the codebook and the channel behavior.

To analyze the LHS of (107), we define for every $\mathbf{x}_1, \mathbf{y} \in \mathbb{R}^n$ and every message $m \neq 1$ the binary random variable

$$B_m(\mathbf{x}_1, \mathbf{y}; \Delta) = \mathbb{1}\left\{w(\mathbf{y}|\mathbf{X}_m) \geq w(\mathbf{y}|\mathbf{x}_1) \cdot e^{-\frac{n\Delta}{2}}\right\}. \tag{108}$$

Our analysis of $V_m(\mathbf{x}_1, \mathbf{y})$ depends on whether $\phi^{(n)}(\mathbf{y} - \mathbf{x}_1)$ differs from K (no overload) or equals K (corresponding to quantizer overload). In the former case, the random variable $V_m(\mathbf{x}_1, \mathbf{y})$ can be upper bounded by $B_m(\mathbf{x}_1, \mathbf{y}; \Delta)$ because

$$V_m(\mathbf{x}_1, \mathbf{y}) = \mathbb{1}\left\{\phi^{(n)}(\mathbf{y} - \mathbf{X}_m) = \phi^{(n)}(\mathbf{y} - \mathbf{x}_1)\right\} \tag{109}$$

$$\leq \mathbb{1}\left\{\left|\|\mathbf{y} - \mathbf{X}_m\|^2 - \|\mathbf{y} - \mathbf{x}_1\|^2\right| < n\Delta\right\} \tag{110}$$

$$\leq \mathbb{1}\left\{\|\mathbf{y} - \mathbf{X}_m\|^2 \leq \|\mathbf{y} - \mathbf{x}_1\|^2 + n\Delta\right\} \tag{111}$$

$$= \mathbb{1}\left\{e^{-\frac{\|\mathbf{y} - \mathbf{X}_m\|^2}{2}} \geq e^{-\frac{\|\mathbf{y} - \mathbf{x}_1\|^2 + n\Delta}{2}}\right\} \tag{112}$$

$$= B_m(\mathbf{x}_1, \mathbf{y}; \Delta), \tag{113}$$

where (110) holds because, for the case at hand, the equality of helper's description implies that $\|\mathbf{y} - \mathbf{X}_m\|^2$ and $\|\mathbf{y} - \mathbf{x}_1\|^2$ lie in a same interval of length $n\Delta$. In the latter case—which is exponentially rare when M exceeds the noise variance—we simply upper bound $V_m(\mathbf{x}_1, \mathbf{y})$ by 1.

The ρ-th moment of the list can now be expressed using the law of total expectation as

$$\mathbb{E}\left[\left(\sum_{m \neq 1} V_m(\mathbf{X}_1, \mathbf{Y})\right)^{\rho}\right]$$

$$= \mathbb{E}\left[\left(\sum_{m \neq 1} V_m(\mathbf{X}_1, \mathbf{Y})\right)^{\rho} \,\middle|\, T \neq K\right] \Pr[T \neq K] + \mathbb{E}\left[\left(\sum_{m \neq 1} V_m(\mathbf{X}_1, \mathbf{Y})\right)^{\rho} \,\middle|\, T = K\right] \Pr[T = K] \tag{114}$$

$$\leq \mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{X}_1, \mathbf{Y}; \Delta)\right)^{\rho} \,\middle|\, T \neq K\right] \Pr[T \neq K] + e^{n\rho R} \Pr[T = K] \tag{115}$$

$$\leq \mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{X}_1, \mathbf{Y}; \Delta)\right)^{\rho}\right] + e^{n\rho R} \Pr[T = K]. \tag{116}$$

The second term on the RHS of (116) tends to zero by (104). The first term is studied in the following lemma:

Lemma 3. *If $\rho > 0$, $\Delta > 0$, and $R < R_0(\rho) - \Delta$, then*

$$\lim_{n \to \infty} \mathbb{E}\left[\left(\sum_{m \neq 1} B_m(\mathbf{X}_1, \mathbf{Y}; \Delta)\right)^{\rho}\right] = 0. \tag{117}$$

Proof. See Appendix B. □

For a given $R < R_0(\rho)$, achievability is thus established using this lemma and (116) by picking M sufficiently large for (104) to hold, and then picking K large enough to guarantee that $R < R_0(\rho) - M/K$ so that, by Lemma 3, the first term on the RHS of (116) will also tend to zero.

5.2. Case 2: $R_h > 0$

The key to proving the achievability of $R_{\text{cutoff}}(\rho) + R_h$ is in showing that rate-R_h help can be utilized to increase the data rate by R_h, and that this can be done losslessly, with arbitrarily small (positive) power, and in one channel use. To show how this can be done, we show that—by using the channel once to send a single input that is bounded by $\sqrt{A}$ (with A any prespecified positive number) and using help taking values in the set $\mathcal{T} = \{0, \ldots, \kappa - 1\}$—we can send error-free a message taking values in said set. To transmit $m \in \{0, \ldots, \kappa - 1\}$, the encoder sends

$$x = m \cdot \frac{\sqrt{A}}{\kappa}, \tag{118}$$

which is upper-bounded by $\sqrt{A}$. Upon observing the noise Z, the helper produces the description T by quantizing the normalized noise and taking modulo, i.e.,

$$T = \left\lfloor Z \cdot \frac{\kappa}{\sqrt{A}} \right\rfloor \quad \mathrm{mod}\ \kappa, \tag{119}$$

which is an element of $\{0, \ldots, \kappa - 1\}$. Based on Y and T, the decoder can calculate

$$\hat{m} = \left\lfloor Y \cdot \frac{\kappa}{\sqrt{A}} - T \right\rfloor \quad \mathrm{mod}\ \kappa, \tag{120}$$

which equals m, because

$$\hat{m} = \left\lfloor (x + Z) \cdot \frac{\kappa}{\sqrt{A}} - T \right\rfloor \quad \mathrm{mod}\ \kappa \tag{121}$$

$$= \left\lfloor m + Z \cdot \frac{\kappa}{\sqrt{A}} - T \right\rfloor \quad \mathrm{mod}\ \kappa \tag{122}$$

$$= \left(m + \left\lfloor Z \cdot \frac{\kappa}{\sqrt{A}} \right\rfloor - T \right) \quad \mathrm{mod}\ \kappa \tag{123}$$

$$= m, \tag{124}$$

where (123) holds because m and T are both integers.

Using this building-block, we can now prove the achievability of $R_{\mathrm{cutoff}}(\rho) + R_{\mathrm{h}}$ by employing two-phase time sharing. Specifically, we propose the following blocklength-$(n + 1)$ scheme. In the first n channel uses, the helper operates at rate zero as in Section 5.1. By the achievability result proved in Section 5.1, for any $R < R_0(\rho)$, there exists a sequence of blocklength-n rate-R codebooks $\{\mathbf{x}_m\}_{m=1,\ldots,e^{nR}}$, with $\|\mathbf{x}_m\|^2 \leq (n - 1)A$ for every m, and zero-rate helpers $\phi(Z^n)$, such that the remotely-plausible-list $\mathcal{L}(Y^n, \phi(Z^n))$ satisfies

$$\lim_{n \to \infty} \mathbb{E}\left[\left| \mathcal{L}(Y^n, \phi(Z^n)) \right|^\rho \right] = 1. \tag{125}$$

In the $(n + 1)$-th channel-use we use the aforementioned coding scheme with κ being $\lceil e^{nR_{\mathrm{h}}} \rceil$. Since that scheme is error-free, the overall remotely-plausible-list for the two phases has the same cardinality as that of the first phase, namely $\left| \mathcal{L}(Y^n, \phi(Z^n)) \right|$, and hence, its ρ-th moment tends to 1 by (125).

The achievability now follows by verifying that, the power of the transmitted input sequence $\mathbf{x}$ satisfies

$$\|\mathbf{x}\|^2 = \|x^n\|^2 + \|x_{n+1}\|^2 \leq nA + A = (n + 1)A; \tag{126}$$

the rate of the helper is

$$\frac{1}{n+1}\left(0 + nR_{\mathrm{h}}\right) \tag{127}$$

and the rate achieved by the scheme is

$$\frac{1}{n+1}\left(nR_0(\rho) + nR_{\mathrm{h}}\right) \tag{128}$$

which tend to R_{h} and $R_0(\rho) + R_{\mathrm{h}}$, respectively, as n tends to infinity.

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

Appendix A. Proof of Lemma 2

We shall establish that the expectation

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{m\neq 1} B_m(\mathbf{x}_1,\mathbf{y}) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^{\rho}\right] \tag{A1}$$

$$= \int_{\mathbf{y}\in\mathbb{R}^n} \int_{\mathbf{x}_1\in\mathbb{R}^n} \mathbb{E}\left[\sum_{m\neq 1} B_m(\mathbf{x}_1,\mathbf{y}) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^{\rho} w(\mathbf{y}|\mathbf{x}_1)\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A2}$$

tends to zero as n tends to infinity whenever $R < R_0(\rho)$.

First notice that conditional on the transmitted codeword $\mathbf{x}_1$ and the channel output $\mathbf{y}$, the random variables $\{B_m\}_{m\neq 1}$ are IID Bernoulli, with B_m determined by $\mathbf{X}_m$ and being of probability of success

$$p(\mathbf{x}_1,\mathbf{y}) = \Pr\left[w(\mathbf{y}|\mathbf{X}_m) \geq w(\mathbf{y}|\mathbf{x}_1)\right] \tag{A3}$$

$$= \Pr\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}} \geq w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\right] \tag{A4}$$

$$\leq w(\mathbf{y}|\mathbf{x}_1)^{-\frac{1}{1+\rho}}\, \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right], \tag{A5}$$

where the last inequality follows from Markov's inequality. Thus

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{m\neq 1} B_m(\mathbf{x}_1,\mathbf{y}) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^{\rho}\right]$$

$$= \int_{\mathbf{y}\in\mathbb{R}^n} \int_{\mathbf{x}_1\in\mathbb{R}^n} \mathbb{E}\left[\sum_{m\neq 1} B_m(\mathbf{x}_1,\mathbf{y}) \,\Big|\, \mathbf{X} = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^{\rho} w(\mathbf{y}|\mathbf{x}_1)\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A6}$$

$$\leq e^{n\rho R} \int_{\mathbf{y}\in\mathbb{R}^n} \int_{\mathbf{x}_1\in\mathbb{R}^n} p(\mathbf{x}_1,\mathbf{y})^{\rho}\, w(\mathbf{y}|\mathbf{x}_1)\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A7}$$

$$\leq e^{n\rho R} \int_{\mathbf{y}\in\mathbb{R}^n} \int_{\mathbf{x}_1\in\mathbb{R}^n} \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right]^{\rho} w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}). \tag{A8}$$

$$= e^{n\rho R} \int_{\mathbf{y}\in\mathbb{R}^n} \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right]^{\rho} \left(\int_{\mathbf{x}_1\in\mathbb{R}^n} w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\, dQ^{(n)}(\mathbf{x}_1)\right) d\nu(\mathbf{y}) \tag{A9}$$

$$= e^{n\rho R} \int_{\mathbf{y}\in\mathbb{R}^n} \left(\int_{\mathbf{x}\in\mathbb{R}^n} w(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}}\, dQ^{(n)}(\mathbf{x})\right)^{1+\rho} d\nu(\mathbf{y}) \tag{A10}$$

$$\leq \left(\frac{e^{r\delta}}{\mu}\right)^{1+\rho} e^{n\rho R} \left(\int_{y\in\mathbb{R}} \left(\int_{x\in\mathbb{R}} e^{r(x^2-A)} w(y|x)^{\frac{1}{1+\rho}}\, dQ_G(x)\right)^{1+\rho} d\nu(y)\right)^n \tag{A11}$$

$$= \left(\frac{e^{r\delta}}{\mu}\right)^{1+\rho} e^{n\rho R}\, e^{-nE_{0,m}(\rho,Q_G,r)}, \tag{A12}$$

where (A11) follows from the upper bound (39) on the Radon–Nykodim derivative and holds for every $r \geq 0$. Choosing r as $r^\star$ that achieves $E_{0,m}^*(\rho, Q_G)$ (cf. (31)), we obtain

$$\mathbb{E}\left[\mathbb{E}\left[\sum_{m\neq 1} B_m(\mathbf{x}_1,\mathbf{y}) \,\Big|\, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^{\rho}\right] \leq \left(\frac{e^{r^\star\delta}}{\mu}\right)^{1+\rho} e^{n\rho R}\, e^{-nE_{0,m}^*(\rho,Q_G)} \tag{A13}$$

$$= \left(\frac{e^{r^\star\delta}}{\mu}\right)^{1+\rho} e^{n\rho(R-R_0(\rho))}, \tag{A14}$$

where the last equality follows from (74).

The Central Limit Theorem guarantees that, as n tends to infinity, μ approaches $1/2$. Consequently, the RHS of (A14) tends to zero whenever $R < R_0(\rho)$.

Appendix B. Proof of Lemma 3

To prove the lemma, we shall establish that, whenever $R < R_0(\rho) - \Delta$,

$$\lim_{n \to \infty} \mathbb{E}\left[\mathbb{E}\left[\left.\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}; \Delta) \right| \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^\rho\right] = 0, \tag{A15}$$

where the outer expectation is over $\mathbf{X}_1$ and $\mathbf{Y}$. From this (117) will follow in much the same way that (88) followed from (90) in Section 4.3.

To establish (A15), first note that, conditional on the transmitted codeword $\mathbf{x}_1$ and the channel output $\mathbf{y}$, the random variables $\{B_m(\mathbf{x}_1, \mathbf{y}; \Delta)\}_{m \neq 1}$ are IID Bernoulli, with B_m determined by $\mathbf{X}_m$ and being of probability of success

$$p(\mathbf{x}_1, \mathbf{y}; \Delta) = \Pr\left[w(\mathbf{y}|\mathbf{X}_m) \geq w(\mathbf{y}|\mathbf{x}_1)\, e^{-\frac{n\Delta}{2}}\right] \tag{A16}$$

$$= \Pr\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}} \geq w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\, e^{-\frac{n\Delta}{2(1+\rho)}}\right] \tag{A17}$$

$$\leq w(\mathbf{y}|\mathbf{x}_1)^{-\frac{1}{1+\rho}}\, e^{\frac{n\Delta}{2(1+\rho)}}\, \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right], \tag{A18}$$

where the last inequality follows from Markov's inequality. Consequently,

$$\mathbb{E}\left[\mathbb{E}\left[\left.\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}; \Delta) \right| \mathbf{X}_1 = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^\rho\right]$$

$$= \int_{\mathbf{y} \in \mathbb{R}^n} \int_{\mathbf{x}_1 \in \mathbb{R}^n} \mathbb{E}\left[\left.\sum_{m \neq 1} B_m(\mathbf{x}_1, \mathbf{y}; \Delta) \right| \mathbf{X} = \mathbf{x}_1, \mathbf{Y} = \mathbf{y}\right]^\rho w(\mathbf{y}|\mathbf{x}_1)\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A19}$$

$$\leq e^{n\rho R} \int_{\mathbf{y} \in \mathbb{R}^n} \int_{\mathbf{x}_1 \in \mathbb{R}^n} p(\mathbf{x}_1, \mathbf{y}; \Delta)^\rho\, w(\mathbf{y}|\mathbf{x}_1)\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A20}$$

$$\leq e^{n\rho R} e^{\frac{n\rho\Delta}{2(1+\rho)}} \int_{\mathbf{y} \in \mathbb{R}^n} \int_{\mathbf{x}_1 \in \mathbb{R}^n} \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right]^\rho w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A21}$$

$$< e^{n\rho\Delta} e^{n\rho R} \int_{\mathbf{y} \in \mathbb{R}^n} \int_{\mathbf{x}_1 \in \mathbb{R}^n} \mathbb{E}\left[w(\mathbf{y}|\mathbf{X}_m)^{\frac{1}{1+\rho}}\right]^\rho w(\mathbf{y}|\mathbf{x}_1)^{\frac{1}{1+\rho}}\, dQ^{(n)}(\mathbf{x}_1)\, d\nu(\mathbf{y}) \tag{A22}$$

where (A22) holds because $\rho, \Delta > 0$ so $n\rho\Delta / (2(1+\rho)) < n\rho\Delta$.

Except for the $e^{n\rho\Delta}$ factor, the RHS of (A22) is identical to the RHS of (A8), which was shown to decay at least as fast as $e^{n\rho(R - R_0(\rho))}$; see (A14). It follows that the RHS of (A22) tends to zero whenever $R + \Delta < R_0(\rho)$.

References

1. Gallager, R.G. *Information Theory and Reliable Communication*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1968.
2. Verdú, S. Error exponents and α-mutual information. *Entropy* **2021**, *23*, 199. [CrossRef]
3. Lapidoth, A.; Marti, G.; Yan, Y. Other helper capacities. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Victoria, Australia, 12–20 July 2021; pp. 1272–1277. [CrossRef]
4. Cover, T.M. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.
5. Kim, Y. Capacity of a class of deterministic relay channels. *IEEE Trans. Inf. Theory* **2008**, *54*, 1328–1329. [CrossRef]
6. Bross, S.I.; Lapidoth, A.; Marti, G. Decoder-assisted communications over additive noise channels. *IEEE Trans. Commun.* **2020**, *68*, 4150–4161. [CrossRef]
7. Lapidoth, A.; Marti, G. Encoder-assisted communications over additive noise channels. *IEEE Trans. Inf. Theory* **2020**, *66*, 6607–6616. [CrossRef]
8. Merhav, N. On error exponents of encoder-assisted communication systems. *IEEE Trans. Inf. Theory* **2021**, *67*, 7019–7029. [CrossRef]
9. Bunte, C.; Lapidoth, A. Encoding tasks and Rényi entropy. *IEEE Trans. Inf. Theory* **2014**, *60*, 5065–5076. [CrossRef]
10. Pinsker, M.S.; Sheverdjaev, A.Y. Transmission capacity with zero error and erasure. *Probl. Peredachi Informatsii* **1970**, *6*, 20–24.
11. Csiszar, I.; Narayan, P. Channel capacity for a given decoding metric. *IEEE Trans. Inf. Theory* **1995**, *41*, 35–43. [CrossRef]
12. Telatar, I.E. Zero-error list capacities of discrete memoryless channels. *IEEE Trans. Inf. Theory* **1997**, *43*, 1977–1982. [CrossRef]
13. 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]

14. Bunte, C.; Lapidoth, A.; Samorodnitsky, A. The zero-undetected-error capacity approaches the Sperner capacity. *IEEE Trans. Inf. Theory* **2014**, *60*, 3825–3833. [CrossRef]

15. Nakiboğlu, B.; Zheng, L. Errors-and-erasures decoding for block codes with feedback. *IEEE Trans. Inf. Theory* **2012**, *58*, 24–49. [CrossRef]

16. Bunte, C.; Lapidoth, A. The zero-undetected-error capacity of discrete memoryless channels with feedback. In Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 1–5 October 2012, pp. 1838–1842. [CrossRef]

17. Bunte, C.; Lapidoth, A. On the listsize capacity with feedback. *IEEE Trans. Inf. Theory* **2014**, *60*, 6733–6748. [CrossRef]

18. Lapidoth, A.; Miliou, N. Duality bounds on the cutoff rate with applications to Ricean fading. *IEEE Trans. Inf. Theory* **2006**, *52*, 3003–3018. [CrossRef]

19. Pfister, C. On Rényi Information Measures and Their Applications. Ph.D. Thesis, ETH Zurich, Zurich, Switzerland, 2019.

20. Rosenthal, H.P. On the subspaces of L^p ($p > 2$) spanned by sequences of independent random variables. *Isr. J. Math.* **1970**, *8*, 273–303. [CrossRef]

21. Arıkan, E. An inequality on guessing and its application to sequential decoding. *IEEE Trans. Inf. Theory* **1996**, *42*, 99–105. [CrossRef]