

On the Listsize Capacity With Feedback

Christoph Bunte and Amos Lapidoth, *Fellow, IEEE*

Abstract—The listsize capacity of a discrete memoryless channel is the largest transmission rate for which the expectation—or, more generally, the ρ -th moment—of the number of messages that could have produced the output of the channel approaches one as the blocklength tends to infinity. We show that for channels with feedback, this rate is upper bounded by the maximum of Gallager’s E_0 function divided by ρ , and that equality holds when the zero-error capacity of the channel is positive. To establish this inequality, we prove that feedback does not increase the cutoff rate. Relationships to other notions of channel capacity are explored.

Index Terms—Cutoff rate, feedback, listsize capacity, zero-error capacity, zero-undetected-error capacity.

I. INTRODUCTION AND RESULTS

THE main focus of this paper is the *listsize capacity* of discrete memoryless channels (DMCs) with feedback. We begin by recalling the definition of the listsize capacity and, to put things into perspective, some other notions of channel capacity.

A. Various Notions of Capacity

If a code for a DMC is to be decoded without errors, then for every sequence of output letters there can be at most one message that, when fed to the encoder, can produce it. The *zero-error capacity* C_0 of a DMC is the largest rate of codes with this property. Determining C_0 for arbitrary DMCs is one of the longest standing open problems in Information Theory [1]. If we only require that the correct message be decodable with probability approaching one as the blocklength tends to infinity, then suddenly the problem becomes tractable. Indeed, the largest rate achievable in this sense is the *Shannon capacity* C .

The zero-error capacity is a purely combinatoric quantity: it depends only on the zeros of the channel matrix. The Shannon capacity, in contrast, is a continuous function of the channel matrix. Two notions of channel capacity that lie between these two extremes are the listsize capacity and the *zero-undetected-error (z.u.e.) capacity*; they may be defined as follows.

- 1) Consider a decoder that outputs the list of all the messages that could have produced the given output of the channel. The listsize capacity is the largest rate

achievable in the sense that the ρ -th moment of the length of this list approaches one as the blocklength tends to infinity [2], [3]. It is denoted by $C_\ell(\rho)$. In this paper, ρ can be any number greater than zero.

- 2) Consider a decoder that either outputs the correct message (when there is a unique message that could have produced the given output) or declares an erasure (otherwise). The z.u.e. capacity is the largest rate achievable in the sense that the probability of erasure approaches zero as the blocklength tends to infinity [2], [4]. It is denoted by C_{0-u} .

For any given channel,

$$C_0 \leq C_\ell(\rho) \leq C_{0-u} \leq C. \quad (1)$$

The first and third inequalities are obvious, and the second inequality is proved in Proposition 1.9 ahead. The listsize and z.u.e. capacities are not purely combinatoric quantities, nor are they continuous functions of the channel matrix. But like C_0 , determining $C_\ell(\rho)$ or C_{0-u} for arbitrary DMCs is, to the best of our knowledge, an open problem.

B. Feedback and New Results

The picture changes when there is a noiseless feedback link from the output of the channel to the encoder. Indeed, for channels with feedback, the zero-error capacity was proved by Shannon [5] to be equal to the single-letter expression (36) ahead. The z.u.e. capacity with feedback was found in [6] and [7] and can be expressed as in (61) ahead.

Encouraged by these results, we focus here on the listsize capacity with feedback $C_{\ell,\text{fb}}(\rho)$. For channels with positive zero-error capacity we prove that $C_{\ell,\text{fb}}(\rho)$ equals the maximum over all input distributions of the ratio of Gallager’s E_0 function ([8, p. 138] or (190) in Appendix D) to ρ . Moreover, this maximum is always an upper bound on $C_{\ell,\text{fb}}(\rho)$:

Theorem 1.1: For any $\rho > 0$,

$$C_{\ell,\text{fb}}(\rho) \leq \max_P \frac{E_0(\rho, P)}{\rho}, \quad (2)$$

with equality if $C_0 > 0$.

A lower bound on $C_{\ell,\text{fb}}(\rho)$ when $C_0 = 0$ is provided in Section IV (Theorem 4.1). We can use Theorem 1.1 to show:

Proposition 1.2: Irrespective of $\rho > 0$, feedback can increase the listsize capacity.

Proof: The channel in Figure 1 has positive zero-error capacity, and $\max_P E_0(\rho, P)/\rho$ approaches $\log 3$ as ε tends to zero. Consequently, by Theorem 1.1, $C_{\ell,\text{fb}}(\rho)$ approaches $\log 3$ as ε tends to zero. But according to Proposition 1.7 ahead we may combine the output symbols 0 and 1 without

Manuscript received November 1, 2013; revised August 11, 2014; accepted August 19, 2014. Date of publication September 8, 2014; date of current version October 16, 2014. This paper was presented in part at 2013 IEEE Information Theory Workshop.

The authors are with the Signal and Information Processing Laboratory, ETH Zurich, Zurich 8092, Switzerland (e-mail: bunte@isi.ee.ethz.ch; lapidoth@isi.ee.ethz.ch).

Communicated by A. Wagner, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2014.2355815

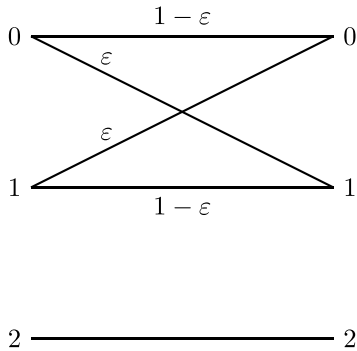


Fig. 1. A channel with $C_{\ell,fb}(\rho) > C_{\ell}(\rho)$.

altering $C_{\ell}(\rho)$, so $C_{\ell}(\rho) \leq \log 2$ because the resulting output alphabet is binary. \square

We note that also $C_{0-u,fb} > C_{0-u}$ for the channel in Figure 1 [7].

The direct part of Theorem 1.1 is proved in Section III, where we also show that the inequality (2) need not be tight if $C_0 = 0$. In order to derive (2), we recall [3] the following operational meaning of the right-hand side of (2). Consider the list of all the messages that under a uniform prior are at least as likely as the correct one given the output of the channel. The *cutoff rate* $R_{\text{cutoff}}(\rho)$ is the largest rate of codes for which the ρ -th moment of the length of this list approaches one as the blocklength tends to infinity. Since the list of messages that could have produced the output contains those that are at least as likely as the correct one,

$$C_{\ell}(\rho) \leq R_{\text{cutoff}}(\rho), \quad (3)$$

and, for channels with feedback,

$$C_{\ell,fb}(\rho) \leq R_{\text{cutoff,fb}}(\rho). \quad (4)$$

En route to the converse part of Theorem 1.1 we prove:

Theorem 1.3: For any $\rho > 0$,

$$R_{\text{cutoff,fb}}(\rho) = \max_P \frac{E_0(\rho, P)}{\rho}. \quad (5)$$

Inequality (2) follows directly from (4) and Theorem 1.3. The converse part of Theorem 1.3 is proved in Section II. The achievability part follows from the well-known result (see [3])

$$R_{\text{cutoff}}(\rho) = \max_P \frac{E_0(\rho, P)}{\rho} \quad (6)$$

combined with the trivial fact that $R_{\text{cutoff,fb}}(\rho) \geq R_{\text{cutoff}}(\rho)$. To keep this paper self-contained, we prove the achievability part of (6) in Appendix B.¹

As a corollary to Theorem 1.3, we obtain that feedback does not increase the cutoff-rate:

Corollary 1.4: For any $\rho > 0$,

$$R_{\text{cutoff,fb}}(\rho) = R_{\text{cutoff}}(\rho). \quad (7)$$

This paper also contains the following other contributions:

- 1) A generalization of Forney's [10] lower bound on $C_{\ell}(1)$ to $C_{\ell}(\rho)$ for all $\rho > 0$ and a proof that the n -letter version of this bound becomes tight as $n \rightarrow \infty$ even

¹The case where $\rho = 1$ follows essentially from Gallager's derivation of the random coding error exponent [8, Sec. 5.6]. The general case, however, requires a bit more work.

when the input distributions (PMFs) are restricted to be uniform over their support; see Proposition 1.5 and Section V.

- 2) Sufficient conditions for equality in $C_{\ell}(\rho) \leq R_{\text{cutoff}}(\rho)$; see Proposition 1.6.
- 3) A simple method to tighten the upper bounds in $C_{\ell}(\rho) \leq R_{\text{cutoff}}(\rho)$ and $C_{0-u} \leq C$; see Proposition 1.7.
- 4) A proof that $\lim_{\rho \rightarrow 0} C_{\ell}(\rho) = C_{0-u}$; see Proposition 1.9.
- 5) The limit of $C_{\ell,fb}(\rho)$ as $\epsilon \rightarrow 0$ for a class of " ϵ -noise" channels; see Proposition 1.11.

C. Notation and Definitions

The cardinality of a finite set \mathcal{X} is denoted by $|\mathcal{X}|$. We use boldface letters to denote n -tuples, e.g., $\mathbf{x} = (x_1, \dots, x_n)$, and uppercase boldface letters for random n -tuples, e.g., $\mathbf{Y} = (Y_1, \dots, Y_n)$. Sometimes we use x^i as shorthand for (x_1, \dots, x_i) when $0 \leq i \leq n$, where x^0 is the empty tuple. All logarithms are natural logarithms. We adopt the convention that $a \log(b/c)$ equals zero if $a = 0$; equals $+\infty$ if $a > 0$, $b > 0$, and $c = 0$; and equals $-\infty$ if $a > 0$, $b = 0$, and $c > 0$. For information-theoretic quantities like entropy and relative entropy we follow the notation in [9]. In some of the proofs we use basic results about types, all of which can be found in [9, Ch. 2]. In particular, the set of all sequences of type P is denoted by T_P . The set of all sequences whose conditional type is V given \mathbf{x} , i.e., the V -shell of \mathbf{x} , is denoted by $T_V(\mathbf{x})$. Throughout $(\delta_n)_{n \geq 1}$ is used to denote sequences of nonnegative numbers that tend to zero. We write δ'_n, δ''_n , etc., if we want to emphasize that different such sequences are being used. The indicator function is denoted by $1\{\cdot\}$.

A *discrete memoryless channel (DMC)* is specified by its transition law (channel matrix) $W(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, where \mathcal{X} and \mathcal{Y} are finite input and output alphabets. If P is a probability mass function (PMF) on \mathcal{X} , then PW denotes the distribution induced on \mathcal{Y} by P and the transition law W

$$(PW)(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x), \quad y \in \mathcal{Y}. \quad (8)$$

We write P^n for the product PMF on \mathcal{X}^n

$$P^n(\mathbf{x}) = \prod_{i=1}^n P(x_i), \quad \mathbf{x} \in \mathcal{X}^n. \quad (9)$$

The support of a PMF P is denoted by $\text{supp}(P)$, i.e., $\text{supp}(P) = \{x \in \mathcal{X} : P(x) > 0\}$. If $\mathcal{A} \subseteq \mathcal{X}$, we write $P(\mathcal{A})$ instead of $\sum_{x \in \mathcal{A}} P(x)$. Similarly, if $\mathcal{B} \subseteq \mathcal{Y}$, we write $W(\mathcal{B}|x)$ instead of $\sum_{y \in \mathcal{B}} W(y|x)$.

In the absence of feedback, a blocklength- n rate- R encoder is a mapping²

$$f: \{1, \dots, e^{nR}\} \rightarrow \mathcal{X}^n. \quad (10)$$

The domain of f is the *message set* and the (not necessarily distinct) *codewords* $f(1), \dots, f(e^{nR})$ constitute the *codebook*. We sometimes write \mathbf{x}_m instead of $f(m)$ for the codeword to

²More precise would be the integer part of e^{nR} , but for typographical reasons we write e^{nR} instead of $\lfloor e^{nR} \rfloor$.

which the encoder maps the m -th message. Sending the m -th message induces on \mathcal{Y}^n the distribution

$$W^n(\mathbf{y}|f(m)), \quad \mathbf{y} \in \mathcal{Y}^n, \quad (11)$$

where

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n W(y_i|x_i), \quad \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n. \quad (12)$$

We often use the notation

$$\mathcal{X}(y) = \{x \in \mathcal{X} : W(y|x) > 0\}, \quad (13)$$

and

$$\mathcal{X}^n(\mathbf{y}) = \{\mathbf{x} \in \mathcal{X}^n : W^n(\mathbf{y}|\mathbf{x}) > 0\}. \quad (14)$$

Given an encoder f as in (10), we define the lists³

$$\mathcal{L}(\mathbf{y}) = \{m : W^n(\mathbf{y}|f(m)) > 0\}, \quad \mathbf{y} \in \mathcal{Y}^n, \quad (15)$$

and

$$\mathcal{L}(m, \mathbf{y}) = \{\tilde{m} : W^n(\mathbf{y}|f(\tilde{m})) \geq W^n(\mathbf{y}|f(m))\}. \quad (16)$$

Stated differently, $\mathcal{L}(\mathbf{y})$ is the list of all messages that can produce the output sequence \mathbf{y} , and $\mathcal{L}(m, \mathbf{y})$ is the list of all messages that under the uniform prior are at least as likely as the m -th message given that \mathbf{y} is observed at the output.

We can now give precise definitions of $C_\ell(\rho)$, C_{0-u} , and $R_{\text{cutoff}}(\rho)$.

- 1) $C_\ell(\rho)$ is the supremum of all rates R for which there exists a sequence of blocklength- n rate- R encoders $(f_n)_{n \geq 1}$ such that

$$\lim_{n \rightarrow \infty} e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(\mathbf{y})|^\rho = 1. \quad (17)$$

- 2) C_{0-u} is the supremum of all rates R for which there exists a sequence of blocklength- n rate- R encoders $(f_n)_{n \geq 1}$ such that

$$\lim_{n \rightarrow \infty} e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y}: |\mathcal{L}(\mathbf{y})| \geq 2} W^n(\mathbf{y}|f_n(m)) = 0. \quad (18)$$

- 3) $R_{\text{cutoff}}(\rho)$ is the supremum of all rates R for which there exists a sequence of blocklength- n rate- R encoders $(f_n)_{n \geq 1}$ such that

$$\lim_{n \rightarrow \infty} e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(m, \mathbf{y})|^\rho = 1. \quad (19)$$

It follows from Gallager's derivation of the Channel Coding Theorem [8, Ch. 5] that the Shannon capacity C can be achieved by strict ML-decoding, i.e., by a decoder that either produces the unique message of maximum likelihood (if there is one) or erases (otherwise). Consequently, we may define C in terms of the list $\mathcal{L}(m, \mathbf{y})$ as follows.

³We use the word "list" in the sense of a set.

4. C is the supremum of all rates R for which there exists a sequence of blocklength- n rate- R encoders $(f_n)_{n \geq 1}$ such that

$$\lim_{n \rightarrow \infty} e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y}: |\mathcal{L}(m, \mathbf{y})| \geq 2} W^n(\mathbf{y}|f_n(m)) = 0. \quad (20)$$

(The above definitions remain unchanged when the average over the messages is replaced with the maximum. This follows from a standard expurgation argument.)

To extend the above definitions to channels with feedback, we replace f with an n -tuple $(f^{(1)}, \dots, f^{(n)})$, where

$$f^{(i)}: \{1, \dots, e^{nR}\} \times \mathcal{Y}^{i-1} \rightarrow \mathcal{X}, \quad i = 1, \dots, n. \quad (21)$$

(By convention, \mathcal{Y}^0 contains only the empty tuple.) In this case, sending the m -th message induces on \mathcal{Y}^n the distribution

$$\prod_{i=1}^n W(y_i|f^{(i)}(m, \mathbf{y}^{i-1})), \quad \mathbf{y} \in \mathcal{Y}^n. \quad (22)$$

The definitions of $\mathcal{L}(\mathbf{y})$, $\mathcal{L}(m, \mathbf{y})$, $C_{\ell, \text{fb}}(\rho)$, $C_{0-u, \text{fb}}$, and $R_{\text{cutoff, fb}}(\rho)$ are analogous to their no-feedback counterparts.

D. Bounds—Old and New

We begin with some known lower bounds on $C_\ell(\rho)$ and C_{0-u} . Forney [10] showed that

$$C_{0-u} \geq \max_P - \sum_{\mathbf{y} \in \mathcal{Y}} (PW)(\mathbf{y}) \log P(\mathcal{X}(\mathbf{y})) \quad (23)$$

and

$$C_\ell(1) \geq \max_P - \log \sum_{\mathbf{y} \in \mathcal{Y}} (PW)(\mathbf{y}) P(\mathcal{X}(\mathbf{y})), \quad (24)$$

where the maxima are over all PMFs on \mathcal{X} . Forney's bounds can be derived using standard random coding where each component of each codeword is drawn independently according to a PMF P . In Section V we prove the following generalization of (24) (also using standard random coding).

Proposition 1.5: For any $\rho > 0$,

$$C_\ell(\rho) \geq \max_P - \rho^{-1} \log \sum_{\mathbf{y} \in \mathcal{Y}} (PW)(\mathbf{y}) P(\mathcal{X}(\mathbf{y}))^\rho. \quad (25)$$

Neither (23) nor (25) is tight in general.⁴ Tighter bounds can be derived using random coding over constant composition codes [2]–[4], [11]: The corresponding bound on C_{0-u} is

$$C_{0-u} \geq \max_P \min_{\substack{V \ll W \\ PV = PW}} I(P, V), \quad (26)$$

where the minimization is over all auxiliary channels $V(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, such that $V(y|x) = 0$ whenever $W(y|x) = 0$ (i.e., $V \ll W$) and such that the induced output distribution under P is the same as under the true channel W (i.e., $PV = PW$). The corresponding bound on $C_\ell(\rho)$ is

$$C_\ell(\rho) \geq \max_P \min_{\substack{V, V' \\ V \ll W \\ PV = PV'}} I(P, V) + \rho^{-1} D(V' || W | P). \quad (27)$$

⁴An example where they are not tight is the Z-channel; see [11, Example 4.1].

It is shown in [2] and [11] that (26) is at least as tight as (23). Appendix C contains a proof that (27) is at least as tight as (25). (This result may not have appeared in print before.) However, the weaker bounds are simpler because no minimization over auxiliary channels is required.

We can tighten any of the above lower bounds by applying them to the channel $W^n(\mathbf{y}|\mathbf{x})$, $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y} \in \mathcal{Y}^n$ and normalizing the result by $1/n$. Indeed, any blocklength- ν rate- R code for the channel W^n is a blocklength- $n\nu$ rate- R/n code for the channel W .⁵

To give a concrete example, the n -letter version of (25) is

$$C_\ell(\rho) \geq \frac{1}{n} \max_P -\rho^{-1} \log \sum_{\mathbf{y} \in \mathcal{Y}^n} (PW^n)(\mathbf{y}) P(\mathcal{X}^n(\mathbf{y}))^\rho, \quad (28)$$

where the maximum is over all PMFs on \mathcal{X}^n . A numerical evaluation in [2] of the one and two-letter versions of (26) for a specific channel suggests that a strict improvement is possible, and thus that (26) is not always tight. In [2] and [3] it is shown that the n -letter versions of (26) and (27) become tight as $n \rightarrow \infty$. In [12] it is shown that also the n -letter version of the weaker bound (23) becomes tight as $n \rightarrow \infty$, and that this is true even when the input PMFs are restricted to be uniform over their support. In Section V we prove a similar statement for the n -letter version of (25).

The aforementioned limits are not computable in general, but they can be useful nonetheless. For example, in [12] the multiletter version of (23) is used to derive an upper bound on C_{0-u} for the class of ε -noise channels (see below).

We now discuss upper bounds on $C_\ell(\rho)$ and C_{0-u} . Specifically, recall (3) and the rightmost inequality in (1):

$$C_{0-u} \leq C \quad \text{and} \quad C_\ell(\rho) \leq R_{\text{cutoff}}(\rho). \quad (29)$$

For a large class of channels the bounds in (29) are tight:

Proposition 1.6: *The inequalities in (29) hold with equality if there exist functions $A: \mathcal{X} \rightarrow (0, \infty)$ and $B: \mathcal{Y} \rightarrow (0, \infty)$ such that*

$$W(y|x) = A(x)B(y), \quad \text{if } W(y|x) > 0. \quad (30)$$

Proof: The hypothesis implies that the lists $\mathcal{L}(\mathbf{y})$ and $\mathcal{L}(m, \mathbf{y})$ coincide whenever $W^n(\mathbf{y}|f(m)) > 0$ and constant composition codes are used.⁶ Indeed, observe that if \mathbf{x} and \mathbf{x}' are codewords of the same type, then

$$\begin{aligned} W^n(\mathbf{y}|\mathbf{x}) &= \left(\prod_{i=1}^n A(x_i) \right) \left(\prod_{j=1}^n B(y_j) \right) \\ &= \left(\prod_{i=1}^n A(x'_i) \right) \left(\prod_{j=1}^n B(y_j) \right) \\ &= W^n(\mathbf{y}|\mathbf{x}'), \end{aligned} \quad (31)$$

⁵For blocklengths that are not divisible by n , we can interpolate as follows. Suppose the blocklength is $n\nu + \ell$ where $1 \leq \ell < n$. Then we use a good code for W^n of blocklength ν and rate R , and we extend it to a blocklength- $(n\nu + \ell)$ code for W by padding ℓ dummy symbols. Accordingly, the last ℓ output symbols are ignored at the receiver. The rate of the resulting code for W is $\nu R / (n\nu + \ell)$, and this approaches R/n as $\nu \rightarrow \infty$.

⁶Constant composition codes comprise codewords of the same type [9, p. 144].

where the first and last equality hold provided that $W^n(\mathbf{y}|\mathbf{x}) > 0$ and $W^n(\mathbf{y}|\mathbf{x}') > 0$. Thus, all codewords with positive likelihood have the same likelihood.

Since every code has a constant composition subcode of exponentially the same size (there are only polynomially many types), the proposition follows by comparing (17) and (19), and (18) and (20). \square

Proposition 1.6 is essentially due to Csiszár and Narayan [4] (they considered only C_{0-u}), who also observed that all channels with acyclic channel graphs⁷ can be factorized as in (30). This important special case had been proved earlier by Pinsker and Sheverdyaev [13] for C_{0-u} , and by Telatar [3] for $C_\ell(\rho)$. (An intermediate result was obtained by Telatar in [11, Sec. 4.3].) Notable examples of channels with acyclic channel graphs are the Z-channel and the binary erasure channel. In [4] it is conjectured that a necessary condition for $C_{0-u} = C$ is that a factorization of the channel law in the sense of (30) hold on some capacity-achieving subset of inputs (which is clearly also sufficient).

We can sometimes tighten the bounds in (29) by judiciously combining output symbols:

Proposition 1.7: *If $y, y' \in \mathcal{Y}$ are such that for every $x \in \mathcal{X}$, $W(y|x) > 0$ if, and only if, $W(y'|x) > 0$, then C_{0-u} and $C_\ell(\rho)$ are unaltered when y and y' are combined into a single output symbol distinct from all other output symbols.*

Proof: The set $\mathcal{L}(\mathbf{y})$ remains unchanged when any occurrence of y in \mathbf{y} is replaced with y' , or vice versa. \square

Using Proposition 1.7 we can also reduce the size of the output alphabet to at most $2^{|\mathcal{X}|} - 1$ symbols while preserving C_{0-u} and $C_\ell(\rho)$ (there are $2^{|\mathcal{X}|} - 1$ nonempty subsets of inputs). In particular, every binary-input DMC can be reduced to an asymmetric binary erasure channel (possibly with some transition probabilities equal to zero). And since the channel graph of the latter is acyclic, we can apply Proposition 1.6 to it. In this way we can determine C_{0-u} and $C_\ell(\rho)$ for any binary-input channel.

E. Relationships and Analogies

There is a remarkable similarity between the way $R_{\text{cutoff}}(\rho)$ relates to C and the way $C_\ell(\rho)$ relates to C_{0-u} . The following two propositions illustrate this. The first is well-known [8].

Proposition 1.8: *For every $\rho > 0$,*

- 1) $R_{\text{cutoff}}(\rho) \leq C$;
- 2) $R_{\text{cutoff}}(\rho) > 0 \iff C > 0 \iff$ *there exist x, x', y such that $W(y|x) \neq W(y|x')$;*
- 3) $\lim_{\rho \rightarrow 0} R_{\text{cutoff}}(\rho) = C$;
- 4) *and* $\lim_{\rho \rightarrow \infty} R_{\text{cutoff}}(\rho) = -\log \pi_0$, *where*

$$\pi_0 = \min_P \max_{y \in \mathcal{Y}} P(\mathcal{X}(y)). \quad (32)$$

⁷The channel graph of a DMC W is the undirected bipartite graph whose two independent sets are \mathcal{X} and \mathcal{Y} , and where there is an edge between x and y if $W(y|x) > 0$. It is customary to draw the inputs on the left and the outputs on the right, and to label the edges with the transition probabilities. Acyclic means that we cannot find distinct inputs x_1, \dots, x_n and distinct outputs y_1, \dots, y_n such that $W(y_i|x_i) > 0$ and $W(y_i|x_{i+1}) > 0$ for all $i \in \{1, \dots, n\}$ where $n \geq 2$ and $x_{n+1} = x_1$.

Proof: All assertions follow from (6), the fact that $C = \max_P I(P, W)$, and the properties of mutual information and Gallager's E_0 function (see [8, Th. 5.6.3] and Appendix D). \square

Proposition 1.8 remains (almost) true when $R_{\text{cutoff}}(\rho)$ is replaced with $C_\ell(\rho)$, and C is replaced with C_{0-u} .

Proposition 1.9: For every $\rho > 0$,

- 1) $C_\ell(\rho) \leq C_{0-u}$;
- 2) $C_\ell(\rho) > 0 \iff C_{0-u} > 0 \iff$ there exist x, x', y such that $W(y|x) > W(y|x') = 0$;
- 3) $\lim_{\rho \rightarrow 0} C_\ell(\rho) = C_{0-u}$;
- 4) and $\lim_{\rho \rightarrow \infty} C_\ell(\rho) = -\log \pi_0$.

Proof: Part 1 follows from Markov's inequality:

$$\begin{aligned} \Pr(|\mathcal{L}(\mathbf{Y})| \geq 2) &= \Pr(|\mathcal{L}(\mathbf{Y})|^\rho - 1 \geq 2^\rho - 1) \\ &\leq \frac{\mathbb{E}[|\mathcal{L}(\mathbf{Y})|^\rho] - 1}{2^\rho - 1}, \end{aligned} \quad (33)$$

and the right-hand side of (33) tends to zero if $\mathbb{E}[|\mathcal{L}(\mathbf{Y})|^\rho]$ tends to one.

To prove Part 2, assume that for every $y \in \mathcal{Y}$, $W(y|x) > 0$ for some $x \in \mathcal{X}$ implies $W(y|x') > 0$ for all $x' \in \mathcal{X}$. Then $|\mathcal{L}(\mathbf{y})| = e^{nR}$ for all sequences $\mathbf{y} \in \mathcal{Y}^n$ that can be produced by some (and hence all) messages. Thus, $C_{0-u} = C_\ell(\rho) = 0$. Conversely, if there exist x, x', y for which $W(y|x) > W(y|x') = 0$, then combine all outputs other than y into a single output distinct from y , and use only the inputs x and x' . This reduces the channel to a Z-channel with crossover probability $1 - W(y|x)$. For the Z-channel we have by Proposition 1.6 that $C_{0-u} = C$ and $C_\ell(\rho) = R_{\text{cutoff}}(\rho)$, where both C and $R_{\text{cutoff}}(\rho)$ are positive by Proposition 1.8 Part 2.

As to Part 3, since $C_\ell(\rho)$ is clearly nonincreasing in ρ , the limit exists and is upper-bounded by C_{0-u} on account of Part 1. On the other hand, it follows from the proof of [2, Th. 1] that for every rate $R < C_{0-u}$ the probability that $|\mathcal{L}(\mathbf{Y})|$ exceeds one can be driven to zero exponentially in the blocklength, i.e., we can find a sequence of blocklength- n rate- R encoders $(f_n)_{n \geq 1}$ for which this probability is bounded by $e^{-n\delta}$ for some (possibly very small) $\delta > 0$. For this sequence of encoders,

$$\begin{aligned} &\frac{1}{e^{nR}} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(\mathbf{y})|^\rho \\ &\leq 1 + e^{n\rho R} \frac{1}{e^{nR}} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y}: |\mathcal{L}(\mathbf{y})| \geq 2} W^n(\mathbf{y}|f_n(m)) \end{aligned} \quad (34)$$

$$\leq 1 + e^{n\rho R} e^{-n\delta}, \quad (35)$$

where (34) follows by splitting the sum over $\mathbf{y} \in \mathcal{Y}^n$ into a sum over all \mathbf{y} for which $|\mathcal{L}(\mathbf{y})| = 1$ and a sum over all other \mathbf{y} , and by using $|\mathcal{L}(\mathbf{y})| \leq e^{nR}$ to bound the latter. Part 3 follows by noting that the right-hand side of (35) tends to one as n tends to infinity if $\rho < \delta/R$.

As to Part 4, since $C_\ell(\rho) \leq R_{\text{cutoff}}(\rho)$, we have $\lim_{\rho \rightarrow \infty} C_\ell(\rho) \leq -\log \pi_0$ by Proposition 1.8 Part 4. On the other hand, it follows from (25) by replacing the average over $y \in \mathcal{Y}$ with the maximum that $C_\ell(\rho) \geq -\log \pi_0$ for all $\rho > 0$. \square

Proposition 1.10: Propositions 1.8 and 1.9 are true also for channels with feedback. In particular, if $C_\ell(\rho)$, $C_{\ell,\text{fb}}(\rho)$, C_{0-u} , or $C_{0-u,\text{fb}}$ is positive, then they all are.

Proof: In the case of Proposition 1.8 this follows from the fact that feedback does not increase the Shannon capacity or the cutoff rate (Corollary 1.4). In the case of Proposition 1.9 the original proof goes through except for Part 3. This part, however, is contained in Corollary 4.2 ahead. \square

The quantity $-\log \pi_0$ appearing in Propositions 1.8 and 1.9 has the following operational significance. Shannon [5] proved that the zero-error capacity with feedback $C_{0,\text{fb}}$ can be expressed as

$$C_{0,\text{fb}} = \begin{cases} -\log \pi_0 & \text{if } C_0 > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (36)$$

He further conjectured that

$$-\log \pi_0 = \min_{V \ll W} C(V), \quad (37)$$

where $C(V)$ denotes the Shannon capacity of the channel $V(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and where, as above, $V \ll W$ means that $V(y|x) = 0$ whenever $W(y|x) = 0$. Ahlswede proved this conjecture in [14]. Using the multiletter version of (27), Telatar [3] showed that

$$\lim_{\rho \rightarrow \infty} C_\ell(\rho) = \min_{V \ll W} C(V). \quad (38)$$

Combining (38) with Part 4 of Proposition 1.9 furnishes an alternative proof of (37).

F. Sperner Capacity and ε -Noise Channels

There is an interesting relationship between the listsize capacity, the z.u.e. capacity, and the Sperner capacity of directed graphs [12]. We say that a DMC is ε -noise if $\mathcal{X} \subseteq \mathcal{Y}$ and

$$W(x|x) \geq 1 - \varepsilon, \quad \text{for all } x \in \mathcal{X}. \quad (39)$$

A natural way to associate a directed graph G with an ε -noise channel W is to take \mathcal{X} as the vertex set and to introduce an edge from x to y if $x \neq y$ and $W(y|x) > 0$. It can be shown that [2], [12]

$$\lim_{\varepsilon \rightarrow 0} C_{0-u} = \lim_{\varepsilon \rightarrow 0} C_\ell(\rho) = \Sigma(G), \quad (40)$$

where $\Sigma(G)$ denotes the Sperner capacity of G . (The limits are to be understood in a uniform sense with respect to all ε -noise channels with given graph G .)

As a corollary to Theorem 4.1 ahead, we can show:

Proposition 1.11: For any ε -noise channel with $\mathcal{X} = \mathcal{Y}$ and $C_\ell(\rho) > 0$,

$$\lim_{\varepsilon \rightarrow 0} C_{\ell,\text{fb}}(\rho) = \log |\mathcal{X}|. \quad (41)$$

The proof of Proposition 1.11 is postponed until Section IV.

G. A Dual Source-Coding Problem

A source coding analog to the listsize capacity has recently been studied in [15]. There, the encoder uses nR bits to describe a sequence of length n emitted by an IID source P_X . Based on this description, the decoder produces a list of sequences that is guaranteed to contain the one emitted by the source. It is shown that the smallest rate R achievable in the sense that the ρ -th moment of the length of this list tends to one as n tends to infinity is given by the Rényi entropy of order $1/(1+\rho)$

$$H_{\frac{1}{1+\rho}}(X) = \frac{1}{\rho} \log \left(\sum_{x \in \mathcal{X}} P_X(x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (42)$$

It is also shown that if the source produces pairs (X, Y) and the Y -sequence is known as side-information at the encoder and decoder, then the smallest achievable rate is given by a conditional version of Rényi entropy

$$H_{\frac{1}{1+\rho}}(X|Y) = \frac{1}{\rho} \log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_{X,Y}(x, y)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (43)$$

This definition of conditional Rényi entropy was proposed by Arimoto [16], who showed that

$$\max_P \frac{E_0(\rho, P)}{\rho} = \max_P H_{\frac{1}{1+\rho}}(X) - H_{\frac{1}{1+\rho}}(X|Y), \quad (44)$$

where $(X, Y) \sim P(x)W(y|x)$. Thus, at least for channels whose channel law factorizes in the Csiszár-Narayan sense (30), Rényi entropy plays a role in channel and source coding with lists that is reminiscent of the role played by Shannon entropy in channel and source coding with the usual probability of error criteria.

II. THE CONVERSE PART OF THEOREM 1.3

In this section we prove the converse part of Theorem 1.3, i.e., we prove

$$R_{\text{cutoff,fb}}(\rho) \leq \max_P \frac{E_0(\rho, P)}{\rho}. \quad (45)$$

We need the following lemmas.

Lemma 2.1 ([17, Th. 1]): If the pair $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ (where \mathcal{X} and \mathcal{Y} are finite sets) has PMF $P_{X,Y}$, and if the function $G: \mathcal{X} \times \mathcal{Y} \rightarrow \{1, \dots, |\mathcal{X}|\}$ is one-to-one as a function of $x \in \mathcal{X}$ for every $y \in \mathcal{Y}$, then

$$E[G(X, Y)^\rho] \geq \frac{1}{(1 + \log|\mathcal{X}|)^\rho} \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P_{X,Y}(x, y)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (46)$$

Lemma 2.2 ([8, Th. 5.6.5]): A Necessary and sufficient condition for a PMF P to minimize

$$\sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x)W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \quad (47)$$

(and hence maximize $E_0(\rho, P)$) is

$$\sum_{y \in \mathcal{Y}} W(y|x)^{\frac{1}{1+\rho}} \alpha_y(P)^\rho \geq \sum_{y \in \mathcal{Y}} \alpha_y(P)^{1+\rho}, \quad (48)$$

for all $x \in \mathcal{X}$, with equality if $P(x) > 0$. Here,

$$\alpha_y(P) = \sum_{x \in \mathcal{X}} P(x)W(y|x)^{\frac{1}{1+\rho}}. \quad (49)$$

Equipped with these lemmas, we can now prove (45). Fix a sequence of rate- R blocklength- n encoders as in (21). For each $\mathbf{y} \in \mathcal{Y}^n$ list the messages in decreasing order of their likelihood (resolving ties arbitrarily)

$$\prod_{i=1}^n W(y_i | f_n^{(i)}(m, y^{i-1})), \quad 1 \leq m \leq e^{nR}. \quad (50)$$

Let $G(m, \mathbf{y})$ denote the position of the m -th message in this list. Then $G(\cdot, \mathbf{y})$ is one-to-one for every $\mathbf{y} \in \mathcal{Y}^n$, and

$$G(m, \mathbf{y}) \leq |\mathcal{L}(m, \mathbf{y})|, \quad 1 \leq m \leq e^{nR}. \quad (51)$$

(Equality holds in (51) if no message other than m has the same likelihood as m .) By Lemma 2.1,

$$\begin{aligned} & \frac{(1+nR)^\rho}{e^{nR}} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} G(m, \mathbf{y})^\rho \prod_{i=1}^n W(y_i | f_n^{(i)}(m, y^{i-1})) \\ & \geq \sum_{\mathbf{y} \in \mathcal{Y}^n} \left(\sum_{m=1}^{e^{nR}} \left(\frac{1}{e^{nR}} \prod_{i=1}^n W(y_i | f_n^{(i)}(m, y^{i-1})) \right)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\ & = e^{n\rho R} \sum_{\mathbf{y} \in \mathcal{Y}^n} \left(\sum_{\mathbf{f} \in \mathcal{F}} \tilde{P}(\mathbf{f}) \tilde{W}_n(\mathbf{y}|\mathbf{f})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \end{aligned} \quad (52)$$

where \tilde{W}_n is the channel whose input alphabet \mathcal{F} is the set of all n -tuples $\mathbf{f} = (f^{(1)}, \dots, f^{(n)})$ of functions of the form

$$f^{(i)}: \mathcal{Y}^{i-1} \rightarrow \mathcal{X}, \quad i = 1, \dots, n; \quad (53)$$

whose output alphabet is \mathcal{Y}^n ; and whose transition law is

$$\tilde{W}_n(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^n W(y_i | f^{(i)}(y^{i-1})), \quad \mathbf{y} \in \mathcal{Y}^n, \mathbf{f} \in \mathcal{F}, \quad (54)$$

and where \tilde{P} is the PMF on \mathcal{F} induced by uniform messages and the encoding functions:

$$\tilde{P}(\mathbf{f}) = \frac{|\{m : (f_n^{(1)}(m), \dots, f_n^{(n)}(m, \cdot)) = \mathbf{f}\}|}{e^{nR}}. \quad (55)$$

The proof is complete once we establish that

$$\sum_{\mathbf{y} \in \mathcal{Y}^n} \left(\sum_{\mathbf{f} \in \mathcal{F}} \tilde{P}(\mathbf{f}) \tilde{W}_n(\mathbf{y}|\mathbf{f})^{\frac{1}{1+\rho}} \right)^{1+\rho} \geq e^{-n \max_P E_0(\rho, P)}, \quad (56)$$

because it will then follow using (51) and (52) that the ρ -th moment of $|\mathcal{L}(M, \mathbf{Y})|$ cannot tend to one unless $R \leq \max_P E_0(\rho, P)/\rho$. To establish (56), let P^* be a PMF on \mathcal{X} that minimizes

$$\sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x)W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \quad (57)$$

and hence achieves the maximum of $E_0(\rho, P)$. We use Lemma 2.2 (applied to the channel \tilde{W}_n) to show that the PMF \tilde{P}^* on \mathcal{F} given by

$$\tilde{P}^*(\mathbf{f}) = \begin{cases} \prod_{i=1}^n P^*(x_i) & f^{(1)} \equiv x_1, \dots, f^{(n)} \equiv x_n, \\ 0 & \text{otherwise,} \end{cases} \quad (58)$$

minimizes the left-hand side of (56) over all PMFs on \mathcal{F} . The notation $f^{(i)} \equiv x_i$ means that $f^{(i)}(y^{i-1}) = x_i$ for all $y^{i-1} \in \mathcal{Y}^{i-1}$. To verify that \tilde{P}^* satisfies the conditions of Lemma 2.2 for the channel \tilde{W}_n , observe that

$$\begin{aligned} & \sum_{\mathbf{y} \in \mathcal{Y}^n} \tilde{W}_n(\mathbf{y}|\mathbf{f})^{\frac{1}{1+\rho}} \left(\sum_{\mathbf{f}' \in \mathcal{F}} \tilde{P}^*(\mathbf{f}') \tilde{W}_n(\mathbf{y}|\mathbf{f}')^{\frac{1}{1+\rho}} \right)^\rho \\ &= \sum_{\mathbf{y} \in \mathcal{Y}^n} \tilde{W}_n(\mathbf{y}|\mathbf{f})^{\frac{1}{1+\rho}} \left(\sum_{\mathbf{x} \in \mathcal{X}^n} (P^*)^n(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right)^\rho \\ &= \sum_{y_1} W(y_1|f^{(1)})^{\frac{1}{1+\rho}} \alpha_{y_1}(P^*)^\rho \\ & \quad \times \cdots \times \sum_{y_n} W(y_n|f^{(n)}(y^{n-1}))^{\frac{1}{1+\rho}} \alpha_{y_n}(P^*)^\rho. \end{aligned} \quad (59)$$

Applying (48) (with P replaced by P^*) to the innermost of the nested sums on the right-hand side of (59) (the sum over y_n), then to the second innermost (the sum over y_{n-1}), and so on, we obtain

$$\begin{aligned} & \sum_{\mathbf{y} \in \mathcal{Y}^n} \tilde{W}_n(\mathbf{y}|\mathbf{f})^{\frac{1}{1+\rho}} \left(\sum_{\mathbf{f}' \in \mathcal{F}} \tilde{P}^*(\mathbf{f}') \tilde{W}_n(\mathbf{y}|\mathbf{f}')^{\frac{1}{1+\rho}} \right)^\rho \\ & \geq \left(\sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P^*(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \right)^n \\ &= \sum_{\mathbf{y} \in \mathcal{Y}^n} \left(\sum_{\mathbf{x} \in \mathcal{X}^n} (P^*)^n(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x})^{\frac{1}{1+\rho}} \right)^{1+\rho} \\ &= \sum_{\mathbf{y} \in \mathcal{Y}^n} \left(\sum_{\mathbf{f}' \in \mathcal{F}} \tilde{P}^*(\mathbf{f}') \tilde{W}_n(\mathbf{y}|\mathbf{f}')^{\frac{1}{1+\rho}} \right)^{1+\rho}, \end{aligned} \quad (60)$$

with equality if $f^{(1)} \equiv x_1, \dots, f^{(n)} \equiv x_n$ and $P^*(x_i) > 0$ for all $i \in \{1, \dots, n\}$, i.e., with equality if $\tilde{P}^*(\mathbf{f}) > 0$. The PMF \tilde{P}^* thus satisfies the conditions of Lemma 2.2 (for the channel \tilde{W}_n) for minimizing the left-hand side of (56), and the value of this minimum is equal to the right-hand side of (56). \square

III. THE DIRECT PART OF THEOREM 1.1

Before presenting the proof of the direct part of Theorem 1.1, we comment on the necessity of the assumption $C_0 > 0$. Since the z.u.e. capacity with feedback is given by [6], [7]

$$C_{0-\text{u,fb}} = \begin{cases} C & \text{if } C_{0-\text{u}} > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (61)$$

one might suspect that for equality in (2) it suffices that $C_\ell(\rho)$ be positive (and not necessarily C_0). This, however, is not true:

Proposition 3.1: A positive value of $C_\ell(\rho)$ does not guarantee equality in (2).

Proof: A counterexample is the channel in Figure 2. For this channel $C_0 = 0$, $C_\ell(\rho) > 0$, and $\max_P E_0(\rho, P)/\rho$ is at least close to $\log 2$ for small ε . But even with feedback, if the received sequence contains only the symbols 0 and 1, then the decoder cannot rule out any of the messages and the list it produces is of size e^{nR} . And regardless of what is fed to the channel, the probability of observing only the symbols 0 and 1

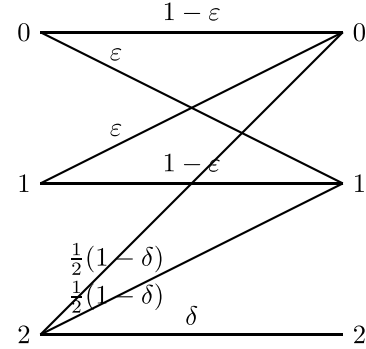


Fig. 2. A channel with $0 < C_{\ell,\text{fb}}(\rho) < \max_P E_0(\rho, P)/\rho$.

at the output is at least $(1-\delta)^n$. Consequently, the ρ -th moment of the length of the list produced by the decoder is at least

$$e^{n\rho(R+\rho^{-1}\log(1-\delta))}, \quad (62)$$

and $C_{\ell,\text{fb}}(\rho)$ must thus be bounded by $-\rho^{-1}\log(1-\delta)$, which is close to zero for very small $\delta > 0$ and hence smaller than $\max_P E_0(\rho, P)/\rho$ if $\varepsilon > 0$ is sufficiently small. \square

To prove the direct part of Theorem 1.1, we propose the following coding scheme. Let P^* be a PMF on \mathcal{X} that achieves the maximum of $E_0(\rho, P)$. Select a sequence of types $(P_n)_{n \geq 1}$ with $P_n \rightarrow P^*$ as $n \rightarrow \infty$, where each P_n is a type in \mathcal{X}^n .⁸ In the first phase, we send one of e^{nR} messages using the length- n type- P_n codewords $\mathbf{x}_1, \dots, \mathbf{x}_{e^{nR}}$. (We will generate the codebook at random later on). In the second phase, after the output sequence $\mathbf{y} \in \mathcal{Y}^n$ has been observed through the feedback link, we use a zero-error code (of rate at least $\log 2$) to describe the conditional type V of \mathbf{y} given the codeword.⁹ Since the number of conditional types is polynomial in n , this requires at most $o(n)$ additional channel uses. Let $\mathcal{M}(\mathbf{y}, V) \subseteq \{1, \dots, e^{nR}\}$ denote the set of all messages that are mapped to codewords given which \mathbf{y} has conditional type V , i.e.,

$$\mathcal{M}(\mathbf{y}, V) = \{1 \leq m \leq e^{nR} : \mathbf{y} \in T_V(\mathbf{x}_m)\}. \quad (63)$$

At the end of the second phase both the encoder and the decoder know $\mathcal{M}(\mathbf{y}, V)$ and the decoder knows that the transmitted message is an element of it. We fix some (small) $\alpha > 0$ and partition $\mathcal{M}(\mathbf{y}, V)$ into $e^{n\alpha}$ lists of lengths at most

$$\lceil e^{-n\alpha} |\mathcal{M}(\mathbf{y}, V)| \rceil. \quad (64)$$

In the third phase, we send the index of the list containing the correct message using a zero-error code (of rate at least $\log 2$). This requires at most $\lceil n\alpha/\log 2 \rceil$ additional channel uses. Note that the length of this list is determined by the codeword and the first n channel outputs. We can upper-bound its ρ -th moment by

$$e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|\mathbf{x}_m) \lceil e^{-n\alpha} |\mathcal{M}(\mathbf{y}, P_{\mathbf{y}|\mathbf{x}_m})| \rceil^\rho, \quad (65)$$

⁸This is possible because the set of PMFs with rational components is dense in the set of all PMFs.

⁹To avoid uniqueness issues, we define the conditional type $V(y|x)$ only for $x \in \mathcal{X}$ with $P_n(x) > 0$. Also, when the zero-error capacity is positive, then it is at least $\log 2$.

where $P_{\mathbf{y}|\mathbf{x}_m}$ denotes the conditional type of \mathbf{y} given \mathbf{x}_m . Using the inequality

$$[\zeta]^\rho < 1 + 2^\rho \zeta^\rho, \quad \zeta \geq 0, \quad (66)$$

we can upper-bound (65) by

$$1 + 2^\rho e^{-n(R+\rho\alpha)} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|\mathbf{x}_m) |\mathcal{M}(\mathbf{y}, P_{\mathbf{y}|\mathbf{x}_m})|^\rho. \quad (67)$$

Changing the order of summation, we can rewrite (67) as

$$1 + 2^\rho e^{-n(R+\rho\alpha)} \times \sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_V \sum_{m \in \mathcal{M}(\mathbf{y}, V)} W^n(\mathbf{y}|\mathbf{x}_m) |\mathcal{M}(\mathbf{y}, V)|^\rho, \quad (68)$$

where the middle sum extends over conditional types V . Using the identity

$$W^n(\mathbf{y}|\mathbf{x}_m) = e^{-n(D(V||W|P_n)+H(V|P_n))}, \quad \text{if } m \in \mathcal{M}(\mathbf{y}, V),$$

and the fact that $\mathcal{M}(\mathbf{y}, V)$ can be nonempty only if \mathbf{y} has type $P_n V$, we can rewrite (68) as

$$1 + 2^\rho e^{-n(R+\rho\alpha)} \times \sum_V \sum_{\mathbf{y} \in T_{P_n V}} e^{-n(D(V||W|P_n)+H(V|P_n))} |\mathcal{M}(\mathbf{y}, V)|^{1+\rho}. \quad (69)$$

Next, we average the upper bound (69) over all realizations of a random codebook $\mathbf{X}_1, \dots, \mathbf{X}_{e^{nR}}$ in which each codeword is drawn independently and uniformly from T_{P_n} . This average is

$$1 + 2^\rho e^{-n(R+\rho\alpha)} \times \sum_V \sum_{\mathbf{y} \in T_{P_n V}} e^{-n(D(V||W|P_n)+H(V|P_n))} \mathbb{E}[|\mathcal{M}(\mathbf{y}, V)|^{1+\rho}]. \quad (70)$$

We now upper-bound the $(1+\rho)$ -th moment of $|\mathcal{M}(\mathbf{y}, V)|$. Under the given distribution of the codebook,

$$|\mathcal{M}(\mathbf{y}, V)| = \sum_{m=1}^{e^{nR}} \mathbb{1}\{\mathbf{y} \in T_V(\mathbf{X}_m)\} \quad (71)$$

is a sum of IID Bernoulli random variables (RVs). To compute the probability of the event $\{\mathbf{y} \in T_V(\mathbf{X}_m)\}$ observe that if $\mathbf{x}_m \in T_{P_n}$ and $\mathbf{y} \in T_{P_n V}$, then \mathbf{y} is in the V -shell of \mathbf{x}_m if, and only if, \mathbf{x}_m is in the \tilde{V} -shell of \mathbf{y} , where

$$\tilde{V}(x|y) = \frac{V(y|x)P_n(x)}{(P_n V)(y)}, \quad x \in \mathcal{X}, \quad y \in \text{supp}(P_n V). \quad (72)$$

Consequently, if $\mathbf{y} \in T_{P_n V}$, then

$$\begin{aligned} \Pr(\mathbf{y} \in T_V(\mathbf{X}_m)) &= \Pr(\mathbf{X}_m \in T_{\tilde{V}}(\mathbf{y})) \\ &= \frac{|T_{\tilde{V}}(\mathbf{y})|}{|T_{P_n}|} \end{aligned} \quad (73)$$

$$\leq e^{-n(H(P_n)-H(\tilde{V}|P_n V)-\delta_n)} \quad (74)$$

$$= e^{-n(I(P_n, V)-\delta_n)}, \quad (75)$$

where (73) follows because $T_{\tilde{V}}(\mathbf{y}) \subseteq T_{P_n}$ when $\mathbf{y} \in T_{P_n V}$, and because \mathbf{X}_m is drawn uniformly at random from T_{P_n} ; where (74) follows because $|T_{\tilde{V}}(\mathbf{y})| \leq e^{nH(\tilde{V}|P_n V)}$ when

$\mathbf{y} \in T_{P_n V}$, and because $|T_{P_n}| \geq e^{n(H(P_n)-\delta_n)}$; and where (75) follows by noting that

$$\begin{aligned} H(P_n) - H(\tilde{V}|P_n V) &= H(P_n V) - H(V|P_n) \\ &= I(P_n, V). \end{aligned} \quad (76)$$

It is important to note that the δ_n appearing in (75) does not depend on V . In fact, it can be taken as

$$\delta_n = \frac{|\mathcal{X}| \log(n+1)}{n}. \quad (77)$$

To bound the $(1+\rho)$ -th moment of a binomial RV with exponential parameters, we use Lemma 1.1 (Appendix A), specifically (126). This yields for every $\mathbf{y} \in T_{P_n V}$

$$\begin{aligned} \mathbb{E}[|\mathcal{M}(\mathbf{y}, V)|^{1+\rho}] \\ \leq \gamma e^{n(R-I(P_n, V)+\delta_n)} + \gamma e^{n(1+\rho)(R-I(P_n, V)+\delta_n)}. \end{aligned} \quad (78)$$

Using (78), the fact that $|T_{P_n V}| \leq e^{nH(P_n V)}$, and (76), we can upper-bound (70) by

$$\begin{aligned} 1 + \gamma 2^\rho \sum_V e^{-n(\rho\alpha+D(V||W|P_n)-\delta_n)} \\ + \gamma 2^\rho \sum_V e^{-n(\rho\alpha-\rho R+D(V||W|P_n)+\rho I(P_n, V)-(1+\rho)\delta_n)}. \end{aligned} \quad (79)$$

Since $D(V||W|P_n)$ is nonnegative, and since the number of conditional types V is polynomial in n , we can upper-bound (79) by

$$\begin{aligned} 1 + \gamma 2^\rho e^{-n(\rho\alpha-\delta'_n)} \\ + \gamma 2^\rho e^{-n(\rho\alpha-\rho R+\min_V D(V||W|P_n)+\rho I(P_n, V)-\delta'_n)}, \end{aligned} \quad (80)$$

where the minimum is over all channels $V(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. The first exponential term tends to zero as n tends to infinity if $\alpha > 0$. The second exponential term tends to zero if

$$R < \alpha + \lim_{n \rightarrow \infty} \min_V I(P_n, V) + \rho^{-1} D(V||W|P_n). \quad (81)$$

The rate of the coding scheme approaches $R/(1+\alpha/\log 2)$ as n tends to infinity. Letting α tend to zero, it thus follows that any rate below

$$\lim_{n \rightarrow \infty} \min_V I(P_n, V) + \rho^{-1} D(V||W|P_n) \quad (82)$$

is achievable. And since by [9, Exercise 10.24]

$$I(P_n, V) + \rho^{-1} D(V||W|P_n) \geq \frac{E_0(\rho, P_n)}{\rho}, \quad (83)$$

it follows from the continuity of $E_0(\rho, P)$ in P that all rates below $E_0(\rho, P^*)/\rho$ are achievable. \square

IV. A LOWER BOUND ON THE LISTSIZE CAPACITY WITH FEEDBACK

The direct part of Theorem 1.1 is useless when $C_0 = 0$. With this case in mind, we propose

Theorem 4.1: If $C_\ell(\rho) > 0$, then

$$C_{\ell, \text{fb}}(\rho) \geq \frac{R^*(\rho)}{1 + \frac{\rho R^*(\rho)}{\log \frac{1}{1-q^x}}}, \quad (84)$$

where

$$R^*(\rho) = \sup_{\xi > 0} \max_P \frac{E_0(\xi, P)}{\xi + \rho}, \quad (85)$$

and where q^* is the maximum of $W(\mathcal{Y}_0|x_1)$ taken over all $x_1 \in \mathcal{X}$ and over all the subsets $\mathcal{Y}_0 \subset \mathcal{Y}$ for which there exists some $x_0 \in \mathcal{X}$ with $W(\mathcal{Y}_0|x_0) = 0$. If $C_0 > 0$, i.e., if the zero-error capacity is positive, then $q^* = 1$, and we interpret the right-hand side of (84) as $R^*(\rho)$.

Note that the assumption $C_\ell(\rho) > 0$ implies $q^* > 0$. Indeed, if $C_\ell(\rho) > 0$, then, by Proposition 1.9 Part 2, we can find x_0, x_1, y_0 such that $W(y_0|x_0) = 0$ and $W(y_0|x_1) > 0$. Taking $\mathcal{Y}_0 = \{y_0\}$ thus shows that $q^* \geq W(\mathcal{Y}_0|x_1) > 0$. Also note that, in view of Theorem 1.1 and Proposition 1.6, the lower bound in (84) is interesting only when $C_0 = 0$ and the channel law does not factorize in the Csiszár-Narayan sense (30).

Before presenting a proof of Theorem 4.1, we use it to provide a proof of Proposition 1.11, and we give another corollary to Theorem 4.1, Corollary 4.2, which contains the earlier result (61).

As to the proof of Proposition 1.11, in the notation of Theorem 4.1 we have $q^* \geq 1 - \varepsilon$ if W is ε -noise and $C_\ell(\rho) > 0$. Indeed, $C_\ell(\rho) > 0$ implies that there exist x_0 and y_0 such that $W(y_0|x_0) = 0$ (Proposition 1.9 Part 2), and the ε -noise property implies that $W(y_0|y_0) \geq 1 - \varepsilon$. Consequently, $\mathcal{Y}_0 = \{y_0\}$, $x_1 = y_0$ is a feasible choice in the definition of q^* . Moreover, if P is the uniform PMF on \mathcal{X} , then

$$\begin{aligned} E_0(\xi, P) &= (1 + \xi) \log |\mathcal{X}| - \log \sum_{y \in \mathcal{X}} \left(\sum_{x \in \mathcal{X}} W(y|x)^{\frac{1}{1+\xi}} \right)^{1+\xi} \\ &\geq (1 + \xi) \log |\mathcal{X}| - \log \sum_{y \in \mathcal{X}} \left(1 + (|\mathcal{X}| - 1) \varepsilon^{\frac{1}{1+\xi}} \right)^{1+\xi} \\ &= \xi \log |\mathcal{X}| - (1 + \xi) \log \left(1 + (|\mathcal{X}| - 1) \varepsilon^{\frac{1}{1+\xi}} \right). \end{aligned} \quad (86)$$

Now fix $\delta > 0$ and choose $\xi > 0$ large enough such that $\xi/(\xi + \rho) > 1 - \delta$. Then from (85) and (86) it follows that

$$R^*(\rho) \geq (1 - \delta) \log |\mathcal{X}| - \frac{1 + \xi}{\xi + \rho} \log \left(1 + (|\mathcal{X}| - 1) \varepsilon^{\frac{1}{1+\xi}} \right), \quad (87)$$

and since the second term on the right-hand side tends to zero as $\varepsilon \rightarrow 0$, it follows from (84) that

$$\liminf_{\varepsilon \rightarrow 0} C_{\ell, \text{fb}}(\rho) \geq (1 - \delta) \log |\mathcal{X}|. \quad (88)$$

Letting $\delta \rightarrow 0$ thus proves (41). \square

Corollary 4.2:

$$\lim_{\rho \rightarrow 0} C_{\ell, \text{fb}}(\rho) = C_{0-\text{u}, \text{fb}} = \begin{cases} C & \text{if } C_{0-\text{u}} > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (89)$$

Proof: If $C_{0-\text{u}} = 0$, then by Proposition 1.10 also $C_{0-\text{u}, \text{fb}} = 0$ and $C_{\ell, \text{fb}}(\rho) = 0$ for all $\rho > 0$. If $C_{0-\text{u}} > 0$, then $C_\ell(\rho) > 0$ and (84) holds for all $\rho > 0$. Moreover,

from (85) we have

$$\begin{aligned} \lim_{\rho \rightarrow 0} R^*(\rho) &= \sup_{\rho > 0} \sup_{\xi > 0} \max_P \frac{E_0(\xi, P)}{\xi + \rho} \\ &= \max_P \sup_{\xi > 0} \frac{E_0(\xi, P)}{\xi} \\ &= \max_P I(P, W) \end{aligned} \quad (90)$$

$$= C, \quad (91)$$

where (90) follows because $E_0(\xi, P)/\xi$ is nonincreasing in $\xi > 0$, $E_0(0, P) = 0$, and $\partial E_0(\xi, P)/\partial \xi|_{\xi=0} = I(P, W)$ (see [8, Th. 5.6.3]). Consequently, by (84), $\lim_{\rho \rightarrow 0} C_{\ell, \text{fb}}(\rho) \geq C$. And since $C_{\ell, \text{fb}}(\rho) \leq C_{0-\text{u}, \text{fb}} \leq C$, it follows that $\lim_{\rho \rightarrow 0} C_{\ell, \text{fb}}(\rho) = C_{0-\text{u}, \text{fb}} = C$. \square

To prove Theorem 4.1, we propose the following coding scheme. Select a positive integer ℓ and let x_0, x_1, \mathcal{Y}_0 achieve q^* . In the first phase, we use a blocklength- n rate- R encoder paired with a decoder that produces a list of the ℓ most likely messages given the output of the channel (resolving ties arbitrarily). As shown in [8, Exercise 5.20], for every PMF P on \mathcal{X} we can find a sequence of such encoders (indexed by the blocklength n) such that the probability of the correct message not being on the list is at most $e^{-n(E_0(\xi, P) - \xi R)}$ for every $0 \leq \xi \leq \ell$.

Thanks to the feedback, the transmitter knows which messages are on the decoder's list, and in the second phase it tries to tell the receiver whether the correct message is among them. To indicate that the correct message is on the list, it sends n' times the symbol x_1 ; otherwise it sends n' times the symbol x_0 . Accordingly, if the receiver observes at least one symbol in \mathcal{Y}_0 during the second phase, it knows with certainty that the correct message is on the list (because $W(\mathcal{Y}_0|x_0) = 0$); otherwise it assumes that the correct message is not on the list, it ignores the third phase, and it produces a final list containing all e^{nR} messages.

If the first two phases are successful, i.e., if the list contains the correct message and the receiver is aware of it, then the third phase is used to transmit the position of the correct message in the list. To this end, we construct ℓ auxiliary codewords $\mathbf{x}_1, \dots, \mathbf{x}_\ell$ of length $k\ell$, where k is a fixed positive integer, as follows. The components $(i - 1)k + 1, \dots, ik$ of \mathbf{x}_i equal x_1 and all its other components equal x_0 . The receiver can identify the correct auxiliary codeword, and thus produce the correct message, if at least one symbol in \mathcal{Y}_0 is observed at the output during the third phase (because $W(\mathcal{Y}_0|x_0) = 0$ and the x_1 -patterns are disjoint). If no symbol in \mathcal{Y}_0 is observed during the third phase, it produces the list of size ℓ (which is guaranteed to contain the correct message). If the first or the second phase is unsuccessful, then it does not matter what the transmitter does in the third phase. For concreteness, it sends $k\ell$ times the symbol x_0 .

To analyze the performance of this coding scheme, define the events

$$\begin{aligned} E_1 &= \{\text{correct message not on the list after 1st phase}\}, \\ E_2 &= \{\text{no symbol in } \mathcal{Y}_0 \text{ is observed in 2nd phase}\}, \\ E_3 &= \{\text{no symbol in } \mathcal{Y}_0 \text{ is observed in 3rd phase}\}. \end{aligned}$$

Let L be the length of the list produced by the receiver. The ρ -th moment of L is upper-bounded by

$$1 + \mathbb{E}[L^\rho | E_1] \Pr(E_1) + \mathbb{E}[L^\rho | E_1^c \cap E_2] \Pr(E_1^c \cap E_2) \\ + \mathbb{E}[L^\rho | E_1^c \cap E_2^c \cap E_3] \Pr(E_1^c \cap E_2^c \cap E_3). \quad (92)$$

We upper-bound the right-hand side of (92) term by term, beginning with

$$\mathbb{E}[L^\rho | E_1] \Pr(E_1) \leq e^{n\rho R} e^{-n(E_0(\xi, P) - \xi R)} \\ = e^{-n(\xi + \rho) \left(\frac{E_0(\xi, P)}{\xi + \rho} - R \right)}, \quad 0 \leq \xi \leq \ell. \quad (93)$$

The right-hand side of (93) approaches zero as n tends to infinity provided that $R < R^*(\rho)$ and ℓ is large enough so that we can pick a ξ in the interval $[0, \ell]$ and a PMF P that achieve a value of $E_0(\xi, P)/(\xi + \rho)$ close enough to $R^*(\rho)$. The next term on the right-hand side of (92) can be upper-bounded as follows.

$$\mathbb{E}[L^\rho | E_1^c \cap E_2] \Pr(E_1^c \cap E_2) \leq e^{n\rho R} (1 - q^*)^{n'} \\ = e^{n \left(\rho R - \frac{n'}{n} \log \frac{1}{1 - q^*} \right)}. \quad (94)$$

The right-hand side of (94) approaches zero as n tends to infinity if we choose

$$n' = n(1 + \delta) \frac{\rho R}{\log \frac{1}{1 - q^*}} \quad (95)$$

for an arbitrarily small $\delta > 0$. (If $C_0 > 0$, and hence $q^* = 1$, then we may take $n' = 1$.) Finally,

$$\mathbb{E}[L^\rho | E_1^c \cap E_2^c \cap E_3] \Pr(E_1^c \cap E_2^c \cap E_3) \leq \ell(1 - q^*)^k, \quad (96)$$

and the right-hand side can be made arbitrarily small by choosing k sufficiently large. (If $C_0 > 0$, we may take $k = 1$.) The rate of the coding scheme is

$$\frac{R}{1 + \frac{n'}{n} + \frac{k\ell}{n}}. \quad (97)$$

Choosing first ℓ sufficiently large, then R close to $R^*(\rho)$, then n' as in (95) with δ sufficiently small, then k sufficiently large, and finally n sufficiently large shows that that all rates strictly less than the right-hand side of (84) are achievable. \square

V. A PROOF OF PROPOSITION 1.5 AND THE ASYMPTOTIC TIGHTNESS OF (28)

In this section we derive the lower bound (25) and show that its n -letter version (28) becomes tight as n tends to infinity even when P is restricted to be uniform over its support.

We begin with a proof of (25). Given a blocklength- n rate- R codebook $\mathbf{x}_1, \dots, \mathbf{x}_{e^{nR}}$, we can write the ρ -th moment of $|\mathcal{L}(\mathbf{Y})|$ as

$$\frac{1}{e^{nR}} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y} | \mathbf{x}_m) \left(1 + \sum_{m' \neq m} Z_{m'}(\mathbf{y}) \right)^\rho, \quad (98)$$

where we define

$$Z_m(\mathbf{y}) = 1\{W^n(\mathbf{y} | \mathbf{x}_m) > 0\}, \quad 1 \leq m \leq e^{nR}. \quad (99)$$

If the codebook is generated at random by drawing each component of each codeword independently according to a PMF P on \mathcal{X} , then the expectation of (98) (over the codebook) is

$$\sum_{\mathbf{y} \in \mathcal{Y}^n} (PW)^n(\mathbf{y}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} Z_m(\mathbf{y}) \right)^\rho \right], \quad (100)$$

where for every $\mathbf{y} \in \mathcal{Y}^n$ the RVs $Z_1(\mathbf{y}), \dots, Z_{e^{nR}}(\mathbf{y})$ are IID Bernoulli. Note that

$$\Pr(Z_m(\mathbf{y}) = 1) = \prod_{i=1}^n P(\mathcal{X}(y_i)) \\ = \prod_{y \in \mathcal{Y}} P(\mathcal{X}(y))^{nP_y(y)} \\ = e^{n \sum_{y \in \mathcal{Y}} P_y(y) \log P(\mathcal{X}(y))} \\ = e^{-nF(P_y)}, \quad (101)$$

where P_y is the type of \mathbf{y} , and where we define

$$F(Q) = - \sum_{y \in \mathcal{Y}} Q(y) \log P(\mathcal{X}(y)). \quad (102)$$

To prove (25) it suffices to show that (100) tends to one as n tends to infinity whenever

$$R < -\rho^{-1} \log \sum_{y \in \mathcal{Y}} (PW)(y) P(\mathcal{X}(y))^\rho. \quad (103)$$

We first show that (103) is equivalent to

$$R < \min_Q F(Q) + \rho^{-1} D(Q || PW). \quad (104)$$

where the minimum is over all PMFs Q on \mathcal{Y} . Indeed, observe that

$$F(Q) + \rho^{-1} D(Q || PW) \\ = -\rho^{-1} \sum_{y \in \mathcal{Y}} Q(y) \log \frac{(PW)(y) P(\mathcal{X}(y))^\rho}{Q(y)} \\ \geq -\rho^{-1} \log \sum_{y \in \mathcal{Y}} (PW)(y) P(\mathcal{X}(y))^\rho, \quad (105)$$

where (105) follows from Jensen's Inequality. The choice

$$Q(y) = \frac{(PW)(y) P(\mathcal{X}(y))^\rho}{\sum_{y' \in \mathcal{Y}} (PW)(y') P(\mathcal{X}(y'))^\rho}, \quad y \in \mathcal{Y}, \quad (106)$$

achieves equality in (105).

Using Lemma 1.1 (Appendix A), specifically (125), we can upper-bound (100) by

$$1 + \gamma \sum_{\mathbf{y} \in \mathcal{Y}^n} (PW)^n(\mathbf{y}) \left(e^{n(R - F(P_y))} 1\{R \leq F(P_y)\} \right. \\ \left. + e^{n\rho(R - F(P_y))} 1\{R > F(P_y)\} \right). \quad (107)$$

Since $(PW)^n(T_Q) \leq e^{-nD(Q || PW)}$, we can upper-bound the sum in (107) by

$$\sum_{Q: R \leq F(Q)} e^{n(R - F(Q) - D(Q || PW))} \\ + \sum_{Q: R > F(Q)} e^{n\rho(R - F(Q) - \rho^{-1} D(Q || PW))}, \quad (108)$$

where Q runs over all types in \mathcal{Y}^n . Next, we show that if the rate R satisfies (104), then (108) tends to zero as n tends to infinity. Assume therefore that (104) holds and define the positive number

$$\delta = \min_Q F(Q) + \rho^{-1} D(Q||PW) - R. \quad (109)$$

The second sum in (108) tends to zero as n tends to infinity because the summand is upper-bounded by $e^{-n\rho\delta}$ and the number of different types is polynomial in n . To show that the first sum in (108) tends to zero, we consider separately the cases $\rho \geq 1$ and $\rho < 1$. In the former case, the summand is upper-bounded by $e^{-n\delta}$ because $D(Q||PW) \geq \rho^{-1} D(Q||PW)$. In the latter case, the summand is upper-bounded by $e^{-n\rho\delta}$ because $R - F(Q) \leq \rho(R - F(Q))$ when $R \leq F(Q)$. We conclude that (108) tends to zero as n tends to infinity for all rates R satisfying (104). In view of the equivalence of (103) and (104), this completes the proof of (25).

To prove that (28) is asymptotically tight even when P is restricted to be uniform over its support, we define

$$J_n(\rho, P) = -\frac{1}{n\rho} \log \sum_{\mathbf{y} \in \mathcal{Y}^n} (PW^n)(\mathbf{y}) P(\mathcal{X}^n(\mathbf{y}))^\rho. \quad (110)$$

Since (28) holds for every n , and since restricting the feasible set cannot help,

$$C_\ell(\rho) \geq \overline{\lim}_{n \rightarrow \infty} \max_{P \in \mathcal{U}_n} J_n(\rho, P), \quad (111)$$

where \mathcal{U}_n denotes the set of PMFs on \mathcal{X}^n that are uniform over their support. It remains to show that

$$C_\ell(\rho) \leq \underline{\lim}_{n \rightarrow \infty} \max_{P \in \mathcal{U}_n} J_n(\rho, P). \quad (112)$$

To this end, fix a sequence of rate- R blocklength- n encoders $(f_n)_{n \geq 1}$ with

$$e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(\mathbf{y})|^\rho \leq 1 + \varepsilon_n, \quad (113)$$

where $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$. We first argue that the number of codewords to which only one message is mapped by f_n is at least $e^{n(R-\delta_n)}$. Indeed, if $m \neq m'$ and $f_n(m) = f_n(m')$, then $|\mathcal{L}(\mathbf{y})| \geq 2$ whenever $W^n(\mathbf{y}|f_n(m)) > 0$ (because then also $W^n(\mathbf{y}|f_n(m')) > 0$), and hence

$$\sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(\mathbf{y})|^\rho \geq 2^\rho. \quad (114)$$

If we define

$$\mathcal{M}_n = \{1 \leq m \leq e^{nR} : f_n(m') \neq f_n(m) \text{ for all } m' \neq m\},$$

then it follows from (113) and (114) that

$$\begin{aligned} e^{-nR} |\mathcal{M}_n^c| 2^\rho &\leq e^{-nR} \sum_{m=1}^{e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\mathcal{L}(\mathbf{y})|^\rho \\ &\leq 1 + \varepsilon_n, \end{aligned} \quad (115)$$

where \mathcal{M}_n^c denotes the set complement of \mathcal{M}_n in $\{1, \dots, e^{nR}\}$. Rearranging (115) gives

$$|\mathcal{M}_n^c| \leq e^{nR} 2^{-\rho} (1 + \varepsilon_n). \quad (116)$$

Since $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$, there exists n_0 such that $2^{-\rho} (1 + \varepsilon_n) < 1$ for all $n \geq n_0$. Henceforth assume that $n \geq n_0$. Since $|\mathcal{M}_n| + |\mathcal{M}_n^c| = e^{nR}$, it follows from (116) that

$$\begin{aligned} |\mathcal{M}_n| &\geq e^{nR} (1 - 2^{-\rho} (1 + \varepsilon_n)) \\ &= e^{n(R-\delta_n)}. \end{aligned} \quad (117)$$

Since $1 + \varepsilon_n < 2^\rho$, restricting the message set to \mathcal{M}_n can only decrease the ρ -th moment of the length of the list, so

$$\frac{1}{|\mathcal{M}_n|} \sum_{m \in \mathcal{M}_n} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y}|f_n(m)) |\tilde{\mathcal{L}}(\mathbf{y})|^\rho \leq 1 + \varepsilon_n, \quad (118)$$

where

$$\tilde{\mathcal{L}}(\mathbf{y}) = \{m \in \mathcal{M}_n : W^n(\mathbf{y}|f_n(m)) > 0\}. \quad (119)$$

Let P_n be the uniform PMF on the set $\{f_n(m) : m \in \mathcal{M}_n\}$. Then $P_n \in \mathcal{U}_n$ and

$$\begin{aligned} P_n(\mathcal{X}^n(\mathbf{y})) &= \frac{|\tilde{\mathcal{L}}(\mathbf{y})|}{|\mathcal{M}_n|} \\ &\leq e^{-n(R-\delta_n)} |\tilde{\mathcal{L}}(\mathbf{y})|, \end{aligned} \quad (120)$$

where (120) follows from (117). Rearranging (120) gives

$$|\tilde{\mathcal{L}}(\mathbf{y})| \geq e^{n(R-\delta_n)} P_n(\mathcal{X}^n(\mathbf{y})). \quad (121)$$

Combining (121) and (118), and taking logarithms, we obtain

$$\log(1 + \varepsilon_n) \geq n\rho(R - \delta_n) - n\rho J_n(\rho, P_n). \quad (122)$$

Dividing by $n\rho$ and letting $n \rightarrow \infty$ shows that

$$R \leq \underline{\lim}_{n \rightarrow \infty} J_n(\rho, P_n). \quad (123)$$

The right-hand side of (123) is upper-bounded by the right-hand side of (112) because $P_n \in \mathcal{U}_n$. \square

APPENDIX A EXPONENTIAL UPPER BOUNDS ON THE ρ -TH MOMENT OF BINOMIAL RVs

Lemma 1.1: Let $X_1, \dots, X_{e^{n\alpha}}$ be IID Bernoulli RVs with success probability

$$p_n = \Pr(X_i = 1) = 1 - \Pr(X_i = 0) \leq e^{-n\beta}, \quad (124)$$

where $n \in \mathbb{N}$, $\alpha > 0$ and $\beta \geq 0$. Let $\rho > 0$. Then

$$\mathbb{E} \left[\left(1 + \sum_{i=1}^{e^{n\alpha}} X_i \right)^\rho \right] \leq \begin{cases} 1 + \gamma e^{n(\alpha-\beta)} & \text{if } \beta \geq \alpha, \\ \gamma e^{n\rho(\alpha-\beta)} & \text{if } \beta < \alpha, \end{cases} \quad (125)$$

and

$$\mathbb{E} \left[\left(\sum_{i=1}^{e^{n\alpha}} X_i \right)^\rho \right] \leq \begin{cases} \gamma e^{n(\alpha-\beta)} & \text{if } \beta \geq \alpha, \\ \gamma e^{n\rho(\alpha-\beta)} & \text{if } \beta < \alpha, \end{cases} \quad (126)$$

where

$$\gamma = \max\{e^{e^\rho - 1}, (\lceil \rho \rceil!)^2 \lceil \rho \rceil\}. \quad (127)$$

Proof: We use the inequalities

$$\zeta < 1 + \zeta \leq e^\zeta, \quad \zeta \in \mathbb{R}, \quad (128)$$

and

$$e^{\eta\zeta} \leq 1 + \zeta(e^\eta - 1), \quad 0 \leq \zeta \leq 1, \quad \eta > 0. \quad (129)$$

(The inequality (129) is a consequence of the convexity of the function $\zeta \mapsto e^{\eta\zeta}$.)

We begin with a proof of (125). Consider first the case $\beta \geq \alpha$ and observe that

$$\mathbb{E}\left[\left(1 + \sum_{i=1}^{e^{n\alpha}} X_i\right)^\rho\right] \leq \mathbb{E}\left[\exp\left(\rho \sum_{i=1}^{e^{n\alpha}} X_i\right)\right] \quad (130)$$

$$= \mathbb{E}[e^{\rho X_1}]^{e^{n\alpha}} \quad (131)$$

$$= (1 + p_n(e^\rho - 1))^{e^{n\alpha}} \quad (132)$$

$$\leq \exp(p_n e^{n\alpha} (e^\rho - 1)) \quad (133)$$

$$\leq \exp(e^{n(\alpha-\beta)} (e^\rho - 1)) \quad (134)$$

$$\leq 1 + \gamma e^{n(\alpha-\beta)}, \quad (135)$$

where (130) and (132) follow from (128); where (131) follows because the X_i 's are IID; where (133) follows from (124); where (134) follows from (129) with $\eta = e^\rho - 1$ and $\zeta = e^{n(\alpha-\beta)}$; and where (135) follows from (127).

Now consider the case $\beta < \alpha$ and observe that

$$\begin{aligned} & \mathbb{E}\left[\left(1 + \sum_{i=1}^{e^{n\alpha}} X_i\right)^\rho\right] \\ &= e^{n\rho(\alpha-\beta)} \mathbb{E}\left[\left(e^{-n(\alpha-\beta)} + e^{-n(\alpha-\beta)} \sum_{i=1}^{e^{n\alpha}} X_i\right)^\rho\right] \end{aligned} \quad (136)$$

$$\leq e^{n\rho(\alpha-\beta)} \mathbb{E}\left[\left(1 + e^{-n(\alpha-\beta)} \sum_{i=1}^{e^{n\alpha}} X_i\right)^\rho\right] \quad (137)$$

$$\leq e^{n\rho(\alpha-\beta)} \mathbb{E}\left[\exp\left(\rho e^{-n(\alpha-\beta)} \sum_{i=1}^{e^{n\alpha}} X_i\right)\right] \quad (138)$$

$$= e^{n\rho(\alpha-\beta)} \mathbb{E}[\exp(\rho e^{-n(\alpha-\beta)} X_1)]^{e^{n\alpha}} \quad (139)$$

$$= e^{n\rho(\alpha-\beta)} (1 + p_n(\exp(\rho e^{-n(\alpha-\beta)}) - 1))^{e^{n\alpha}} \quad (140)$$

$$\leq e^{n\rho(\alpha-\beta)} \exp(p_n e^{n\alpha} (\exp(\rho e^{-n(\alpha-\beta)}) - 1)) \quad (141)$$

$$\leq e^{n\rho(\alpha-\beta)} e^{\rho - 1} \quad (142)$$

$$\leq \gamma e^{n\rho(\alpha-\beta)}, \quad (143)$$

where (136) follows because $e^{-n(\alpha-\beta)} \leq 1$; where (137) and (138) follow from (128); where (139) follows from (124); where (140) follows from (129) with $\eta = \rho$ and $\zeta = e^{-n(\alpha-\beta)}$; and where (141) follows from (127).

We now prove (126). The case $\beta < \alpha$ is implied by (125), and we only need to treat the case $\beta \geq \alpha$. We first show that (126) holds when ρ is an arbitrary positive integer, which we denote by k . For any such k ,

$$\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^k = \sum_{(k_1, \dots, k_{e^{n\alpha}})} \prod_{i=1}^{e^{n\alpha}} X_i^{k_i}, \quad (144)$$

where the sum on the right-hand side extends over all possible choices of nonnegative integers $k_1, \dots, k_{e^{n\alpha}}$ that sum up to k .

Taking the expectation on both sides of (142) yields

$$\mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^k\right] = \sum_{(k_1, \dots, k_{e^{n\alpha}})} \prod_{i=1}^{e^{n\alpha}} \mathbb{E}[X_i^{k_i}], \quad (145)$$

where we used the independence of the X_i 's. Since the X_i 's are 0–1 valued, we have $X_i^{k_i} = X_i$ if $k_i \geq 1$, and $X_i^{k_i} = 1$ if $k_i = 0$. Since the X_i 's have identical distributions, we thus have

$$\prod_{i=1}^{e^{n\alpha}} \mathbb{E}[X_i^{k_i}] = \mathbb{E}[X_1]^{\sum_{i: k_i \geq 1} 1}. \quad (146)$$

Using the trivial upper bound

$$\binom{k}{k_1, \dots, k_{e^{n\alpha}}} \leq k!, \quad (147)$$

and substituting (146) into (143), we obtain

$$\mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^k\right] \leq k! \sum_{k_1 + \dots + k_{e^{n\alpha}} = k} \mathbb{E}[X_1]^{\sum_{i: k_i \geq 1} 1}. \quad (148)$$

For any choice of nonnegative integers $k_1, \dots, k_{e^{n\alpha}}$ that sum up to k , the number of indices i for which $k_i \geq 1$ must be between 1 and k , so we may rewrite (146) as

$$\mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^k\right] \leq k! \sum_{\ell=1}^k \binom{e^{n\alpha}}{\ell} \binom{k-1}{\ell-1} \mathbb{E}[X_1]^\ell, \quad (149)$$

where the first binomial coefficient accounts for the number of ways we can choose exactly ℓ of the $e^{n\alpha}$ integers $k_1, \dots, k_{e^{n\alpha}}$ to be positive, and where the second binomial coefficient accounts for the number of ways we can choose the values of ℓ positive integers that sum up to k . Upper-bounding $\binom{k-1}{\ell-1}$ by $k!$ and upper-bounding $\binom{e^{n\alpha}}{\ell}$ by $e^{n\ell\alpha}$, (147) becomes

$$\begin{aligned} \mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^k\right] &\leq (k!)^2 \sum_{\ell=1}^k e^{n\ell\alpha} \mathbb{E}[X_1]^\ell \\ &\leq (k!)^2 \sum_{\ell=1}^k e^{n\ell(\alpha-\beta)} \\ &\leq e^{n(\alpha-\beta)} (k!)^2 k. \end{aligned} \quad (150)$$

This proves (126) for $\beta \geq \alpha$ and all nonnegative integer values of ρ . If $\beta \geq \alpha$ but ρ is not an integer, then

$$\mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^\rho\right] \leq \mathbb{E}\left[\left(\sum_{i=1}^{e^{n\alpha}} X_i\right)^{\lceil \rho \rceil}\right] \quad (151)$$

$$\leq e^{n(\alpha-\beta)} (\lceil \rho \rceil!)^2 \lceil \rho \rceil, \quad (152)$$

$$\leq \gamma e^{n(\alpha-\beta)}, \quad (153)$$

where (149) follows from (148), and where (150) follows from (127). This completes the proof of (126). \square

APPENDIX B

A PROOF OF THE DIRECT PART OF (6)

Here we prove the achievability part of (6), i.e., we prove that for all $\rho > 0$,

$$R_{\text{cutoff}}(\rho) \geq \max_P \frac{E_0(\rho, P)}{\rho}. \quad (151)$$

Fix $\rho > 0$ and a PMF P on \mathcal{X} . Generate a random blocklength- n rate- R codebook $\mathbf{X}_1, \dots, \mathbf{X}_{e^{nR}}$ by drawing each component of each codeword independently according to P . It suffices to show that the expectation of

$$\frac{1}{e^{nR}} \sum_{1 \leq m \leq e^{nR}} \sum_{\mathbf{y} \in \mathcal{Y}^n} W^n(\mathbf{y} | \mathbf{X}_m) |\mathcal{L}(m, \mathbf{y})|^\rho \quad (152)$$

(with respect to the distribution of the codebook) tends to one as n tends to infinity when $R < E_0(\rho, P)/\rho$. This expectation can be expressed as

$$\sum_{\mathbf{y} \in \mathcal{Y}^n} \sum_{\mathbf{x}_1 \in \mathcal{X}^n} W^n(\mathbf{y} | \mathbf{x}_1) P^n(\mathbf{x}_1) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} B_m(\mathbf{y}, \mathbf{x}_1) \right)^\rho \right], \quad (153)$$

where we define the RVs

$$B_m(\mathbf{y}, \mathbf{x}) = 1 \{ W^n(\mathbf{y} | \mathbf{X}_m) \geq W^n(\mathbf{y} | \mathbf{x}) \}. \quad (154)$$

Note that the distribution of $B_m(\mathbf{y}, \mathbf{x})$ depends on \mathbf{x} and \mathbf{y} only via their joint type. Moreover, if $\mathbf{x} \in T_Q$ and $\mathbf{y} \in T_V(\mathbf{x})$, then

$$W^n(\mathbf{y} | \mathbf{x}) = e^{-n(D(V||W|Q) + H(V|Q))}. \quad (155)$$

Thus, by introducing for every type Q , every conditional type V , and every $m \in \{1, \dots, e^{nR}\}$ the RV

$$\tilde{B}_m(Q, V) = 1 \{ W^n(\mathbf{y}_{QV} | \mathbf{X}_m) \geq e^{-n(D(V||W|Q) + H(V|Q))} \}, \quad (156)$$

where \mathbf{y}_{QV} is an arbitrary sequence in \mathcal{Y}^n of type QV , we can rewrite (153) as

$$\sum_{Q, V} (P \circ W)^n(T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right], \quad (157)$$

where the sum extends over all types Q and all conditional types V , and where $P \circ W$ denotes the distribution on $\mathcal{X} \times \mathcal{Y}$ induced by P and W

$$(P \circ W)(x, y) = P(x)W(y|x), \quad x \in \mathcal{X}, y \in \mathcal{Y}. \quad (158)$$

Next, we derive an upper-bound on (157). To this end, note that for fixed Q and V the RVs

$$\tilde{B}_1(Q, V), \dots, \tilde{B}_{e^{nR}}(Q, V) \quad (159)$$

are IID Bernoulli. We can upper-bound their probability of success as follows.

$$\begin{aligned} \Pr(\tilde{B}_m(Q, V) = 1) &= \Pr(W^n(\mathbf{y}_{QV} | \mathbf{X}_m) \geq e^{-n(D(V||W|Q) + H(V|Q))}) \\ &= \Pr(W^n(\mathbf{y}_{QV} | \mathbf{X}_m)^{\frac{1}{1+\rho}} \geq e^{-\frac{n}{1+\rho}(D(V||W|Q) + H(V|Q))}) \\ &\leq e^{\frac{n}{1+\rho}(D(V||W|Q) + H(V|Q))} \mathbb{E} \left[W^n(\mathbf{y}_{QV} | \mathbf{X}_m)^{\frac{1}{1+\rho}} \right], \end{aligned} \quad (160)$$

where (160) follows from Markov's inequality. As to the expectation on the right-hand side of (160),

$$\begin{aligned} \mathbb{E} \left[W^n(\mathbf{y}_{QV} | \mathbf{X}_m)^{\frac{1}{1+\rho}} \right] &= \prod_{i=1}^n \mathbb{E} \left[W(y_{QV,i} | X_{m,i})^{\frac{1}{1+\rho}} \right] \\ &= \prod_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{n(QV)(y)} \end{aligned} \quad (161)$$

$$\begin{aligned} &= e^{n \sum_{y \in \mathcal{Y}} (QV)(y) \log \sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}}} \\ &= e^{-nK(QV)}, \end{aligned} \quad (162)$$

where (161) follows from the independence of the components of the codewords, and where we define for every PMF \tilde{P} on \mathcal{Y}

$$K(\tilde{P}) = - \sum_{y \in \mathcal{Y}} \tilde{P}(y) \log \sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}}. \quad (163)$$

Substituting (162) into (160),

$$\Pr(\tilde{B}_m(Q, V) = 1) \leq e^{-n(K(QV) - \frac{D(V||W|Q) + H(V|Q)}{1+\rho})}. \quad (164)$$

Having bounded the probability of success of $\tilde{B}(Q, V)$, we next use Lemma 1.1 (Appendix A), specifically (125), to conclude that the ρ -th moment in (157) is bounded by

$$1 + \gamma e^{n(R - K(QV) + \frac{D(V||W|Q) + H(V|Q)}{1+\rho})} \quad (165)$$

if $(Q, V) \in \mathcal{G}(R)$, where

$$\mathcal{G}(R) = \left\{ (Q, V) : K(QV) - \frac{D(V||W|Q) + H(V|Q)}{1+\rho} \geq R \right\},$$

and otherwise is bounded by

$$\gamma e^{n\rho(R - K(QV) + \frac{D(V||W|Q) + H(V|Q)}{1+\rho})}. \quad (166)$$

The other term in (157) can be bounded as

$$\begin{aligned} (P \circ W)^n(T_{Q \circ V}) &\leq e^{-nD(Q \circ V || P \circ W)} \\ &= e^{-n(D(Q||P) + D(V||W|Q))}. \end{aligned} \quad (167)$$

Using (165), (166) and (167), we can bound the summand in (157). We treat separately the cases $(Q, V) \notin \mathcal{G}(R)$ and $(Q, V) \in \mathcal{G}(R)$. In the former case, (166) and (167) give

$$\begin{aligned} (P \circ W)^n(T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] &\leq \gamma e^{n\rho(R - K(QV) - \rho^{-1}D(Q||P) + \frac{H(V|Q) - \rho^{-1}D(V||W|Q)}{1+\rho})}, \\ &\quad (Q, V) \notin \mathcal{G}(R). \end{aligned} \quad (168)$$

We upper-bound the right-hand side of (168) in terms of R , n , ρ , and $E_0(\rho, P)$ by showing that

$$\min_{Q, V} \left\{ K(QV) + \rho^{-1}D(Q||P) - \frac{H(V|Q) - \rho^{-1}D(V||W|Q)}{1+\rho} \right\} = \frac{E_0(\rho, P)}{\rho}, \quad (169)$$

where the minimum is over all PMFs Q on \mathcal{X} and all auxiliary channels $V(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$. To establish (169), define

$$\alpha(y) = \left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^\rho, \quad (170)$$

and observe that

$$\begin{aligned} & K(QV) + \rho^{-1} D(Q||P) - \frac{H(V|Q) - \rho^{-1} D(V||W|Q)}{1 + \rho} \\ &= -\frac{1}{\rho} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} Q(x) V(y|x) \log \frac{P(x) W(y|x)^{\frac{1}{1+\rho}} \alpha(y)}{Q(x) V(y|x)} \\ &\geq -\frac{1}{\rho} \log \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \alpha(y) \quad (171) \\ &= -\frac{1}{\rho} \log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\ &= \frac{E_0(\rho, P)}{\rho}, \quad (172) \end{aligned}$$

where (171) follows from Jensen's Inequality. The proof of (169) is completed by noting that the choice

$$Q(x) V(y|x) = \frac{P(x) W(y|x)^{\frac{1}{1+\rho}} \alpha(y)}{\sum_{x' \in \mathcal{X}, y' \in \mathcal{Y}} P(x') W(y'|x')^{\frac{1}{1+\rho}} \alpha(y')}$$

achieves equality in (171).

Combining (168) with (169) shows that

$$\begin{aligned} & (P \circ W)^n (T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq \gamma e^{n\rho \left(R - \frac{E_0(\rho, P)}{\rho} \right)}, \quad (Q, V) \notin \mathcal{G}(R), \quad \rho > 0. \quad (173) \end{aligned}$$

We now turn to the case where $(Q, V) \in \mathcal{G}(R)$. We treat separately the subcases $\rho \geq 1$ and $0 < \rho < 1$, beginning with the former. From (167) and the fact that relative entropies are nonnegative, it follows that

$$(P \circ W)^n (T_{Q \circ V}) \leq e^{-n\rho^{-1}(D(Q||P) + D(V||W|Q))}, \quad \rho \geq 1. \quad (174)$$

Combining (174) with (165) and (169) gives

$$\begin{aligned} & (P \circ W)^n (T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq (P \circ W)^n (T_{Q \circ V}) + \gamma e^{n \left(R - \frac{E_0(\rho, P)}{\rho} \right)}, \\ & \quad (Q, V) \in \mathcal{G}(R), \quad \rho \geq 1. \quad (175) \end{aligned}$$

It remains to treat the case where $(Q, V) \in \mathcal{G}(R)$ and $0 < \rho < 1$. In this case,

$$\begin{aligned} & R - K(QV) + \frac{D(V||W|Q) + H(V|Q)}{1 + \rho} \\ & \leq \rho \left(R - K(QV) + \frac{D(V||W|Q) + H(V|Q)}{1 + \rho} \right), \\ & \quad (Q, V) \in \mathcal{G}(R), \quad 0 < \rho < 1. \quad (176) \end{aligned}$$

Using (176) to upper-bound the right-hand side of (165), we obtain

$$\begin{aligned} & \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq 1 + \gamma e^{n\rho \left(R - K(QV) + \frac{D(V||W|Q) + H(V|Q)}{1 + \rho} \right)}, \\ & \quad (Q, V) \in \mathcal{G}(R), \quad 0 < \rho < 1. \quad (177) \end{aligned}$$

Combining (177) with (167) and (169) yields

$$\begin{aligned} & (P \circ W)^n (T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq (P \circ W)^n (T_{Q \circ V}) + \gamma e^{n\rho \left(R - \frac{E_0(\rho, P)}{\rho} \right)}, \\ & \quad (Q, V) \in \mathcal{G}(R), \quad 0 < \rho < 1. \quad (178) \end{aligned}$$

Combining (175) with (173) and using the fact that the number of types and conditional types is polynomial in n , we obtain

$$\begin{aligned} & \sum_{Q, V} (P \circ W)^n (T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq 1 + e^{n \left(R - \frac{E_0(\rho, P)}{\rho} + \delta_n \right)} + e^{n\rho \left(R - \frac{E_0(\rho, P)}{\rho} + \delta_n \right)}, \\ & \quad \rho \geq 1. \quad (179) \end{aligned}$$

Similarly, combining (178) with (173), we obtain

$$\begin{aligned} & \sum_{Q, V} (P \circ W)^n (T_{Q \circ V}) \mathbb{E} \left[\left(1 + \sum_{m=2}^{e^{nR}} \tilde{B}_m(Q, V) \right)^\rho \right] \\ & \leq 1 + e^{n\rho \left(R - \frac{E_0(\rho, P)}{\rho} + \delta_n \right)}, \quad 0 < \rho < 1. \quad (180) \end{aligned}$$

This completes the proof of (151) because the right-hand sides of (179) and (180) tend to one as n tends to infinity provided that $R < E_0(\rho, P)/\rho$, and we may choose a P that maximizes the right-hand side. \square

APPENDIX C

A PROOF THAT (27) IS AT LEAST AS TIGHT AS (25)

As pointed out in [3], we may add the constraint $V' \ll W$ in the minimization in (27) without increasing the value of the minimum. For any input PMF P and any two auxiliary channels $V, V' \ll W$ satisfying $PV = PV'$,

$$\begin{aligned} & -\rho^{-1} \log \sum_{y \in \mathcal{Y}} (PW)(y) P(\mathcal{X}(y))^\rho \\ & \leq -\rho^{-1} \log \sum_{y \in \text{supp}(PV')} (PW)(y) P(\mathcal{X}(y))^\rho \quad (181) \end{aligned}$$

$$= -\rho^{-1} \log \sum_{y \in \text{supp}(PV')} (PV')(y) \frac{(PW)(y) P(\mathcal{X}(y))^\rho}{(PV')(y)} \quad (182)$$

$$\leq -\rho^{-1} \sum_{y \in \text{supp}(PV')} (PV')(y) \log \frac{(PW)(y) P(\mathcal{X}(y))^\rho}{(PV')(y)} \quad (183)$$

$$= \rho^{-1} D(PV' || PW) + \sum_{y \in \text{supp}(PV')} (PV')(y) \log \frac{1}{P(\mathcal{X}(y))}, \quad (184)$$

where (181) follows because the support of PV' is a subset of the support of PW (because $V' \ll W$); where (182) follows by multiplying and dividing the summand by $(PV')(y)$; and where (183) follows from Jensen's Inequality. By the Log-Sum Inequality [9, Lemma 3.1]

$$D(PV' \| PW) \leq D(V' \| W | P). \quad (185)$$

The second term on the right-hand side of (184) can be upper-bounded as follows.

$$\begin{aligned} & \sum_{y \in \text{supp}(PV')} (PV')(y) \log \frac{1}{P(\mathcal{X}(y))} \\ &= \sum_{y \in \text{supp}(PV)} (PV)(y) \log \frac{1}{P(\mathcal{X}(y))} \end{aligned} \quad (186)$$

$$\begin{aligned} &= \sum_{y \in \text{supp}(PV)} (PV)(y) \log \frac{(PV)(y)}{(PV)(y)P(\mathcal{X}(y))} \\ &\leq \sum_{y \in \text{supp}(PV)} (PV)(y) \log \frac{(PV)(y)}{(PV)(y) \sum_{x: V(y|x) > 0} P(x)} \end{aligned} \quad (187)$$

$$\leq \sum_{y \in \text{supp}(PV)} \sum_{x: V(y|x) > 0} P(x) V(y|x) \log \frac{V(y|x)}{(PV)(y)} \quad (188)$$

$$= I(P, V), \quad (189)$$

where (186) follows because $PV = PV'$; where (187) follows because $V \ll W$; and where (188) follows from the Log-Sum Inequality. Combining (184) with (185) and (189) shows that the right-hand side of (25) never exceeds the right-hand side of (27). \square

APPENDIX D

A PROPERTY OF GALLAGER'S E_0 FUNCTION

Gallager [8] defined the function

$$E_0(\rho, P) = -\log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \quad (190)$$

for all $\rho \geq 0$ and all PMFs P on \mathcal{X} . Here we show that

$$\lim_{\rho \rightarrow \infty} \max_P \frac{E_0(\rho, P)}{\rho} = -\log \pi_0, \quad (191)$$

where π_0 is defined in (32). This identity is noted without proof in [8]. To establish (191), we first show that for any P

$$\lim_{\rho \rightarrow \infty} \frac{E_0(\rho, P)}{\rho} = -\log \max_{y \in \mathcal{Y}} P(\mathcal{X}(y)). \quad (192)$$

We then use Lemma 5.1 (Appendix E) to justify the interchange of limit and maximization. The lemma applies because $E_0(\rho, P)/\rho$ is nonincreasing and continuous in $\rho > 0$ and continuous on the set of all PMFs on \mathcal{X} (a compact subset of $\mathbb{R}^{|\mathcal{X}|}$).

To prove (192) for a given P , we distinguish two cases: Assume first that there exists $y_0 \in \mathcal{Y}$ such that $W(y_0|x) > 0$ for all $x \in \mathcal{X}$ with $P(x) > 0$. In this case, the right-hand side of (192) is equal to zero because $P(\mathcal{X}(y_0)) = 1$. As to the left-hand side of (192), note that replacing the sum over all $y \in \mathcal{Y}$

on the right-hand side of (190) with the term corresponding to y_0 shows that

$$E_0(\rho, P) \leq -(1 + \rho) \log \sum_{x \in \mathcal{X}} P(x) W(y_0|x)^{\frac{1}{1+\rho}}. \quad (193)$$

Using L'Hospital's Rule,

$$\begin{aligned} & \lim_{\rho \rightarrow \infty} (1 + \rho) \log \sum_{x \in \mathcal{X}} P(x) W(y_0|x)^{\frac{1}{1+\rho}} \\ &= \lim_{\xi \searrow 0} \frac{\log \sum_{x \in \mathcal{X}} P(x) W(y_0|x)^\xi}{\xi} \\ &= \sum_{x \in \mathcal{X}} P(x) \log W(y_0|x). \end{aligned} \quad (194)$$

Combining (194) and (193),

$$\lim_{\rho \rightarrow \infty} E_0(\rho, P) \leq - \sum_{x \in \mathcal{X}} P(x) \log W(y_0|x). \quad (195)$$

Since the right-hand side of (195) is a finite number, and $E_0(\rho, P) \geq 0$, it follows that

$$\lim_{\rho \rightarrow \infty} \frac{E_0(\rho, P)}{\rho} = 0. \quad (196)$$

This establishes (192) for the first case. It remains to check the case where for every $y \in \mathcal{Y}$ there is some $x_y \in \mathcal{X}$ for which $P(x_y) > 0$ and $W(y|x_y) = 0$. In this case, for every $y \in \mathcal{Y}$,

$$\begin{aligned} \left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} &\leq (1 - P(x_y))^{1+\rho} \\ &\rightarrow 0, \quad (\rho \rightarrow \infty). \end{aligned} \quad (197)$$

Consequently, $E_0(\rho, P) \rightarrow \infty$ as $\rho \rightarrow \infty$, so by L'Hospital's Rule

$$\lim_{\rho \rightarrow \infty} \frac{E_0(\rho, P)}{\rho} = \lim_{\rho \rightarrow \infty} \frac{\partial E_0(\rho, P)}{\partial \rho}. \quad (198)$$

Straightforward computations show that

$$\begin{aligned} \frac{\partial E_0(\rho, P)}{\partial \rho} &= - \sum_{y \in \mathcal{Y}} \frac{\left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}}{\sum_{y' \in \mathcal{Y}} \left(\sum_{x' \in \mathcal{X}} P(x') W(y'|x')^{\frac{1}{1+\rho}} \right)^{1+\rho}} \\ &\quad \times \left(\varepsilon(\rho) + \log \sum_{x'' \in \mathcal{X}} P(x'') W(y|x'')^{\frac{1}{1+\rho}} \right), \end{aligned} \quad (199)$$

where $\varepsilon(\rho) \rightarrow 0$ as $\rho \rightarrow \infty$. For each $y \in \mathcal{Y}$, the expression

$$\left(\sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} = e^{(1+\rho) \log \sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}}} \quad (200)$$

is either zero for all $\rho > 0$ or decays exponentially with ρ . Noting that

$$\lim_{\rho \rightarrow \infty} \sum_{x \in \mathcal{X}} P(x) W(y|x)^{\frac{1}{1+\rho}} = P(\mathcal{X}(y)), \quad (201)$$

we see that the slowest decay in (200) occurs for those $y \in \mathcal{Y}$ that maximize $P(\mathcal{X}(y))$. This implies that the right-hand side of (199) approaches the right-hand side of (192) as ρ tends to infinity. \square

APPENDIX E
A MINIMAX LEMMA

Lemma 5.1: Let \mathcal{C} be a compact subset of \mathbb{R}^n , let $\mathcal{I} = [a, \infty)$ for some $a \in \mathbb{R}$, and let $f: \mathcal{I} \times \mathcal{C} \rightarrow \mathbb{R}$ be such that $f(\cdot, \pi)$ is nonincreasing and continuous for every $\pi \in \mathcal{C}$ and $f(\rho, \cdot)$ is continuous for every $\rho \in \mathcal{I}$. Then

$$\lim_{\rho \rightarrow \infty} \max_{\pi \in \mathcal{C}} f(\rho, \pi) = \max_{\pi \in \mathcal{C}} \lim_{\rho \rightarrow \infty} f(\rho, \pi). \quad (202)$$

Proof: We first show that the maximum on the right-hand side of (202) is attained. Select a sequence π_1, π_2, \dots in \mathcal{C} such that

$$\lim_{n \rightarrow \infty} \lim_{\rho \rightarrow \infty} f(\rho, \pi_n) = \sup_{\pi \in \mathcal{C}} \lim_{\rho \rightarrow \infty} f(\rho, \pi). \quad (203)$$

By compactness of \mathcal{C} , we can find a convergent subsequence $\pi_{n_k} \rightarrow \pi_\infty \in \mathcal{C}$ as $k \rightarrow \infty$. By continuity and monotonicity we have for every $\rho_0 \in \mathcal{I}$ that

$$\begin{aligned} f(\rho_0, \pi_\infty) &= \lim_{k \rightarrow \infty} f(\rho_0, \pi_{n_k}) \\ &\geq \lim_{k \rightarrow \infty} \lim_{\rho \rightarrow \infty} f(\rho, \pi_{n_k}) \\ &= \sup_{\pi \in \mathcal{C}} \lim_{\rho \rightarrow \infty} f(\rho, \pi). \end{aligned} \quad (204)$$

Taking $\rho_0 \rightarrow \infty$ thus shows that π_∞ attains the maximum on the right-hand side of (202).

To prove that equality holds in (202), first note that the left-hand side is clearly never smaller than the right-hand side, so it remains to prove the reverse inequality. If the left-hand side equals $-\infty$, then there is nothing left to prove. Otherwise select real numbers a and b such that

$$a < b < \lim_{\rho \rightarrow \infty} \max_{\pi \in \mathcal{C}} f(\rho, \pi) \quad (205)$$

and define the sets

$$\mathcal{A}(\pi) = \{\rho \in \mathcal{I} : f(\rho, \pi) \leq a\}, \quad (206a)$$

$$\mathcal{B}(\pi) = \{\rho \in \mathcal{I} : f(\rho, \pi) \leq b\}. \quad (206b)$$

Our choice of a and b implies that $\mathcal{A}(\pi) \subseteq \mathcal{B}(\pi)$ and $\bigcap_{\pi \in \mathcal{C}} \mathcal{B}(\pi) = \emptyset$. For a fixed $\pi \in \mathcal{C}$, the set $\mathcal{B}(\pi)$ is either empty or, by monotonicity and continuity, an interval of the form $[\lambda, \infty)$. If $\mathcal{B}(\pi_0) = \emptyset$ for some $\pi_0 \in \mathcal{C}$, then $f(\rho, \pi_0) > b$ for every $\rho \in \mathcal{I}$, so $\lim_{\rho \rightarrow \infty} f(\rho, \pi_0) \geq b$, and hence $\max_{\pi \in \mathcal{C}} \lim_{\rho \rightarrow \infty} f(\rho, \pi) \geq b > a$. If $\mathcal{B}(\pi) \neq \emptyset$ for every $\pi \in \mathcal{C}$, then, since $\bigcap_{\pi \in \mathcal{C}} \mathcal{B}(\pi) = \emptyset$, we can find a sequence π_1, π_2, \dots in \mathcal{C} such that $\mathcal{B}(\pi_n) = [\lambda_n, \infty)$ where $\lambda_n \rightarrow \infty$ as $n \rightarrow \infty$. By compactness of \mathcal{C} , we can then find a convergent subsequence $\pi_{n_k} \rightarrow \pi_\infty \in \mathcal{C}$ as $k \rightarrow \infty$. We claim that $\mathcal{A}(\pi_\infty) = \emptyset$. Indeed, for if $\rho_0 \in \mathcal{A}(\pi_\infty)$, i.e., if $f(\rho_0, \pi_\infty) \leq a$, then by continuity $f(\rho_0, \pi_{n_k}) \leq b$ for all sufficiently large k , i.e., $\rho_0 \in \mathcal{B}(\pi_{n_k})$ for all sufficiently large k . This leads to a contradiction because $\mathcal{B}(\pi_{n_k}) = [\lambda_{n_k}, \infty)$ and $\lambda_{n_k} \rightarrow \infty$ as $k \rightarrow \infty$ so $\lambda_{n_k} > \rho_0$ for sufficiently large k . Thus, $\mathcal{A}(\pi_\infty) = \emptyset$ and hence $\lim_{\rho \rightarrow \infty} f(\rho, \pi_\infty) \geq a$, so $\max_{\pi \in \mathcal{C}} \lim_{\rho \rightarrow \infty} f(\rho, \pi) \geq a$. Letting $a \nearrow \lim_{\rho \rightarrow \infty} \max_{\pi \in \mathcal{C}} f(\rho, \pi)$ completes the proof. \square

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their helpful comments.

REFERENCES

- [1] J. Körner and A. Orłitsky, "Zero-error information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2207–2229, Oct. 1998.
- [2] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, Jan. 1996.
- [3] I. E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1977–1982, Nov. 1997.
- [4] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.
- [5] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.
- [6] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 24–49, Jan. 2012.
- [7] C. Bunte and A. Lapidoth, "The zero-undetected-error capacity of discrete memoryless channels with feedback," in *Proc. 50th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2012, pp. 1838–1842.
- [8] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2011.
- [10] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, Mar. 1968.
- [11] I. E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, May 1992.
- [12] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity approaches the Sperner capacity," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3825–3833, Jul. 2014.
- [13] M. S. Pinsker and A. Y. Sheverdyaev, "Transmission capacity with zero error and erasure," *Problemy Peredachi Inf.*, vol. 6, no. 1, pp. 20–24, 1970.
- [14] R. Ahlswede, "Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 25, no. 3, pp. 239–252, 1973.
- [15] C. Bunte and A. Lapidoth, "Encoding tasks and Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5065–5076, Sep. 2014.
- [16] S. Arimoto, "Information measures and capacity of order a for discrete memoryless channels," in *Topics in Information Theory*, vol. 17, I. Csiszár and P. Elias, Eds. Amsterdam, The Netherlands: North Holland, 1977, pp. 41–52.
- [17] E. Arkan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, Jan. 1996.

Christoph Bunte received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering in 2010, 2011, and 2014, all from ETH Zurich, Switzerland.

He is currently a research assistant in the Signal and Information Processing Laboratory at ETH Zurich.

His research interests are in the area of Shannon theory.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in mathematics (summa cum laude, 1986), the B.Sc. degree in electrical engineering (summa cum laude, 1986), and the M.Sc. degree in electrical engineering (1990) all from the Technion—Israel Institute of Technology. He received the Ph.D. degree in electrical engineering from Stanford University in 1995.

In the years 1995–1999 he was an Assistant and Associate Professor at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, and was the KDD Career Development Associate Professor in Communications and Technology. He is now Professor of Information Theory at ETH Zurich in Switzerland. He is the author of the book *A Foundation in Digital Communication* (Cambridge University Press, 2009). His research interests are in digital communications and information theory.

Dr. Lapidoth served in the years 2003–2004 and 2009 as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.