On the Average-Listsize Capacity and the Cutoff Rate of Discrete Memoryless Channels with Feedback

Christoph Bunte and Amos Lapidoth
ETH Zurich
Switzerland
Email: {bunte,lapidoth}@isi.ee.ethz.ch

Abstract—We study the cutoff rate and the average-listsize capacity of discrete memoryless channels (DMCs) with feedback. We show that feedback can increase the average-listsize capacity but not the cutoff rate. For DMCs with positive zero-error capacity, we show that the average-listsize capacity with feedback is equal to the cutoff rate. For all other DMCs, we derive a lower bound on the average-listsize capacity with feedback. The bound is asymptotically tight for low-noise channels. We also show that a multi-letter version of Forney's lower bound on the average-listsize capacity of DMCs without feedback is asymptotically tight.

I. INTRODUCTION

Consider a discrete memoryless channel (DMC) with transition law W and finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. The *average-listsize capacity* $C_{\text{a-l}}$ [1] (a.k.a. *zero-error list capacity* [2]) is defined as the supremum of all rates R that are achievable in the following sense: there exists a sequence of rate-R blocklength-n encoders

$$\mathbf{f}_n \colon \{1, \ldots, 2^{nR}\} \to \mathcal{X}^n, \quad n = 1, 2, \ldots \quad (1)$$

such that

$$\lim_{n \to \infty} \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{\mathbf{y}} L(\mathbf{y}) W^n(\mathbf{y}|\mathbf{f}_n(m)) = 1, \quad (2)$$

where

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{1 \le j \le n} W(y_j|x_j), \quad \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n, \quad (3)$$

and where $L(\mathbf{y})$ denotes the number of messages that cannot be ruled out by the receiver after observing $\mathbf{y}$

$$L(\mathbf{y}) = \big| \{m : W^n(\mathbf{y}|\mathbf{f}_n(m)) > 0\} \big|. \quad (4)$$

The *cutoff rate* R_{cutoff} is defined similarly with $L(\mathbf{y})$ in (2) replaced with

$$L(\mathbf{y}, m) = \big| \{m' : W^n(\mathbf{y}|\mathbf{f}_n(m')) \ge W^n(\mathbf{y}|\mathbf{f}_n(m))\} \big|, \quad (5)$$

i.e., with the number of messages that are at least as likely as the transmitted message given that $\mathbf{y}$ is received. It can be expressed in closed-form as [3]

$$R_{\text{cutoff}} = \max_Q E_0(1, Q), \quad (6)$$

where E_0 is Gallager's exponent function [4]

$$E_0(\rho, Q) = -\log \sum_y \left[\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (7)$$

The definitions of the *zero-undetected-error capacity* $C_{\text{0-u}}$ [2], [5] (a.k.a. *erasures-only capacity* [6] and *zero-error erasure capacity* [1]) and the (ordinary) *capacity* C are obtained by replacing $L(\mathbf{y})$ on the LHS of (2) with the indicator function of the events $\{L(\mathbf{y}) \ge 2\}$ and $\{L(\mathbf{y}, m) \ge 2\}$, respectively, and by replacing 1 on the RHS of (2) with 0.

For DMCs with feedback we denote the quantities above $C_{\text{a-l,fb}}$, $R_{\text{cutoff,fb}}$, $C_{\text{0-u,fb}}$, C_{fb}, and we replace $\mathbf{f}_n$ with

$$f_n^{(j)} \colon \{1, \ldots, 2^{nR}\} \times \mathcal{Y}^{j-1} \to \mathcal{X}, \quad j = 1, \ldots, n \quad (8)$$

and $W^n(\mathbf{y}|\mathbf{f}_n(m))$ with

$$\prod_{1 \le j \le n} W\big(y_j | f_n^{(j)}(m, y^{j-1})\big). \quad (9)$$

It is clear from the definition that

$$C_{\text{a-l}} \le R_{\text{cutoff}} \quad (10)$$

and

$$C_{\text{a-l,fb}} \le R_{\text{cutoff,fb}}. \quad (11)$$

Observe that equality holds in (10) whenever there exist positive functions A and B such that

$$W(y|x) = A(x)B(y), \quad \text{whenever } W(y|x) > 0. \quad (12)$$

Indeed, in this case $L(\mathbf{y})$ and $L(\mathbf{y}, m)$ coincide whenever $W^n(\mathbf{y}|\mathbf{f}_n(m)) > 0$ and constant composition codes (see [7]) are used. (It was shown in [6] that $C_{\text{0-u}} = C$ when (12) holds). This observation includes as a special case the result that equality holds in (10) whenever the bipartite channel graph is acyclic [2]. In general, however, R_{cutoff} and $C_{\text{a-l}}$ behave very differently. For example, compare the conditions for positivity: while $R_{\text{cutoff}} > 0$ if, and only if, (iff) $C > 0$ [4], we have $C_{\text{a-l}} > 0$ iff there exists a triple (x, x', y) such that $W(y|x) > 0$ and $W(y|x') = 0$ [2]. In fact, determining $C_{\text{a-l}}$ for arbitrary DMCs, in particular those not satisfying (12), is an open problem.

It is known since Shannon that $C_{\text{fb}} = C$. In Section II we show that $R_{\text{cutoff,fb}} = R_{\text{cutoff}}$. In view of (11) this implies

$$C_{\text{a-l,fb}} \le R_{\text{cutoff}}. \quad (13)$$

In Section III we show that equality holds in (13) whenever the zero-error capacity (see, e.g., [7]) is positive. By providing an example where $C_{\text{a-l}} < C_{\text{a-l,fb}} = R_{\text{cutoff}}$, we show that

feedback enlarges the set of channels for which the average-listsize capacity is equal to the cutoff rate. For DMCs whose zero-error capacity is zero, we derive a lower bound on $C_{\text{a-l,fb}}$ (Section IV). The bound is asymptotically tight for a class of low-noise channels (Section V). Section VI contains a proof that a multi-letter version of Forney's [8] lower bound on $C_{\text{a-l}}$ is asymptotically tight. We concluded in Section VII with some remarks about the relationship between the different notions of capacity appearing in this paper.

II. FEEDBACK DOES NOT INCREASE THE CUTOFF RATE

Theorem 2.1: For every DMC $R_{\text{cutoff,fb}} = R_{\text{cutoff}}$.

Proof: The proof uses Arikan's lower bound [3, Theorem 1] and the Kuhn-Tucker conditions for Gallager's exponent function [4, Theorem 5.6.5]. Fix a sequence of rate-R blocklength-n encoders as in (8) and recall the definition of $L(\mathbf{y}, m)$ for channels with feedback (see (5) and subsequent paragraph). For each $\mathbf{y} \in \mathcal{Y}^n$, list the messages in decreasing order of their likelihood

$$\prod_{1 \leq j \leq n} W\big(y_j \big| f_n^{(j)}(m, y^{j-1})\big), \tag{14}$$

(resolving ties arbitrarily) and 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 and

$$G(m|\mathbf{y}) \leq L(\mathbf{y}, m). \tag{15}$$

(Equality holds if no message other than m has the same likelihood as m). Consequently, $G(\cdot|\cdot)$ is a guessing function in the sense of [3, Theorem 1], and its expectation can thus be lower bounded as

$$\frac{1}{2^{nR}} \sum_m \sum_{\mathbf{y}} G(m|\mathbf{y}) \prod_{1 \leq j \leq n} W\big(y_j \big| f_n^{(j)}(m, y^{j-1})\big)$$

$$\geq \frac{1}{1+nR} \sum_{\mathbf{y}} \left[\sum_m \sqrt{\frac{1}{2^{nR}} \prod_{1 \leq j \leq n} W\big(y_j \big| f_n^{(j)}(m, y^{j-1})\big)} \right]^2$$

$$= \frac{2^{nR}}{1+nR} \sum_{\mathbf{y}} \left[\sum_m \frac{1}{2^{nR}} \sqrt{\prod_{1 \leq j \leq n} W\big(y_j \big| f_n^{(j)}(m, y^{j-1})\big)} \right]^2$$

$$= \frac{2^{nR}}{1+nR} \sum_{\mathbf{y}} \left[\sum_{\mathbf{f}} \widetilde{Q}(\mathbf{f}) \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f})} \right]^2, \tag{16}$$

where $\widetilde{W}_n$ is the channel whose input alphabet is the set of all n-tuples $\mathbf{f} = (f^{(1)}, \ldots, f^{(n)})$ of functions of the form $f^{(j)} \colon \mathcal{Y}^{j-1} \to \mathcal{X}$, whose output alphabet is $\mathcal{Y}^n$, and whose transition law is

$$\widetilde{W}_n(\mathbf{y}|\mathbf{f}) = \prod_{1 \leq j \leq n} W\big(y_j \big| f^{(j)}(y^{j-1})\big), \tag{17}$$

and where $\widetilde{Q}$ is the PMF on the input alphabet of $\widetilde{W}_n$ induced by uniform messages and the encoder

$$\widetilde{Q}(\mathbf{f}) = \frac{\big| \{m : (f_n^{(1)}(m), \ldots, f_n^{(n)}(m, \cdot)) = \mathbf{f}\} \big|}{2^{nR}}. \tag{18}$$

The proof is completed once we establish that

$$\sum_{\mathbf{y}} \left[\sum_{\mathbf{f}} \widetilde{Q}(\mathbf{f}) \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f})} \right]^2 \geq 2^{-nR_{\text{cutoff}}}, \tag{19}$$

because it will then follow using (15) and (16) that the expectation of $L(\mathbf{y}, m)$ cannot tend to one unless $R \leq R_{\text{cutoff}}$. To establish (19), let $Q^\star$ be a PMF on $\mathcal{X}$ that minimizes

$$\sum_y \left[\sum_x Q(x) \sqrt{W(y|x)} \right]^2 \tag{20}$$

and thus achieves the cutoff rate of the channel W. From the Kuhn-Tucker conditions it follows that $Q^\star$ must satisfy

$$\sum_y \sqrt{W(y|x)} \alpha(y, Q^\star) \geq \sum_y \big[\alpha(y, Q^\star) \big]^2, \quad x \in \mathcal{X}, \tag{21}$$

where

$$\alpha(y, Q) = \sum_x Q(x) \sqrt{W(y|x)}, \quad y \in \mathcal{Y}, \tag{22}$$

and where equality must hold in (21) whenever $Q^\star(x) > 0$. We now use the Kuhn-Tucker conditions to show that

$$\widetilde{Q}^\star(\mathbf{f}) = \begin{cases} \prod_{j=1}^n Q^\star(x_j) & f^{(1)} \equiv x_1, \ldots, f^{(n)} \equiv x_n, \\ 0 & \text{otherwise}, \end{cases} \tag{23}$$

minimizes the LHS of (19) over all PMFs on the input alphabet of $\widetilde{W}_n$. (The notation $f^{(j)} \equiv x_j$ means that $f^{(j)}(y^{j-1}) = x_j$ for all $y^{j-1} \in \mathcal{Y}^{j-1}$.) To this end, observe that

$$\sum_{\mathbf{y}} \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f})} \sum_{\mathbf{f}'} \widetilde{Q}^\star(\mathbf{f}') \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f}')}$$

$$= \sum_{\mathbf{y}} \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f})} \sum_{\mathbf{x}} (Q^\star)^n(\mathbf{x}) \sqrt{W^n(\mathbf{y}|\mathbf{x})}$$

$$= \sum_{y_1} \sqrt{W(y_1|f^{(1)})} \alpha(y_1, Q^\star)$$

$$\times \sum_{y_2} \sqrt{W\big(y_2 \big| f^{(2)}(y_1)\big)} \alpha(y_2, Q^\star)$$

$$\times \cdots$$

$$\times \sum_{y_n} \sqrt{W\big(y_n \big| f^{(n)}(y^{n-1})\big)} \alpha(y_n, Q^\star). \tag{24}$$

Applying (21) to the innermost of the nested sums on the RHS of (24) (the sum over y_n), then to the second innermost (the sum over y_{n-1}), and so on, we obtain

$$\sum_{\mathbf{y}} \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f})} \sum_{\mathbf{f}'} \widetilde{Q}^\star(\mathbf{f}') \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f}')}$$

$$\geq \left[\sum_y \left[\sum_x Q^\star(x) \sqrt{W(y|x)} \right]^2 \right]^n$$

$$= \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} (Q^\star)^n(\mathbf{x}) \sqrt{W^n(\mathbf{y}|\mathbf{x})} \right]^2$$

$$= \sum_{\mathbf{y}} \left[\sum_{\mathbf{f}'} \widetilde{Q}^\star(\mathbf{f}') \sqrt{\widetilde{W}_n(\mathbf{y}|\mathbf{f}')} \right]^2, \tag{25}$$

with equality if $f^{(1)} \equiv x_1, \ldots, f^{(n)} \equiv x_n$ and $Q^\star(x_j) > 0$ for all $j \in \{1, \ldots, n\}$, i.e., with equality if $\widetilde{Q}^\star(\mathbf{f}) > 0$.

(a) A DMC with $C_{\text{a-l,fb}} > C_{\text{a-l}}$ (b) A DMC with $0 < C_{\text{a-l,fb}} < R_{\text{cutoff}}$
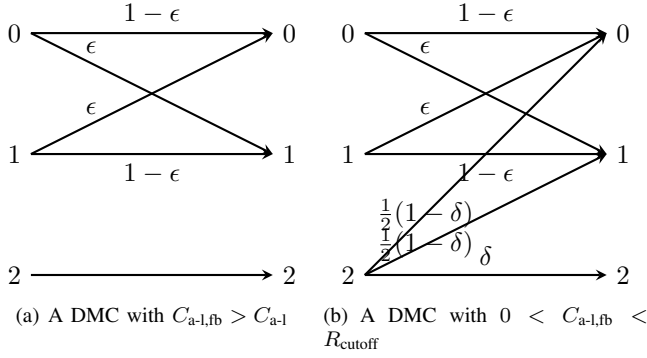
Fig. 1. Two instructive examples. We always assume $0 < \epsilon, \delta < 1$.

The PMF $\widetilde{Q}^\star$ thus satisfies the Kuhn-Tucker conditions for minimizing the LHS of (19); the value of this minimum is

$$\left[\sum_y \left[\sum_x Q^\star(x)\sqrt{W(y|x)}\right]^2\right]^n = 2^{-nR_{\text{cutoff}}}, \qquad (26)$$

where the equality follows from the assumption that $Q^\star$ achieves the cutoff rate of W. ∎

III. THE FEEDBACK AVERAGE-LISTSIZE CAPACITY OF DMCS WITH POSITIVE ZERO-ERROR CAPACITY

Let C_0 denote the zero-error capacity.

Theorem 3.1: If C_0 is positive, then $C_{\text{a-l,fb}} = R_{\text{cutoff}}$.

We can use Theorem 3.1 to prove:

Proposition 3.2: Feedback can increase the average-listsize capacity.

Proof: The channel in Figure 1(a) has positive zero-error capacity and its cutoff rate approaches $\log 3$ as ϵ tends to zero. On the other hand, $C_{\text{a-l}} = 1$ for every $\epsilon \in (0,1)$. Indeed, for the purpose of producing a list of all messages with positive likelihood given a particular output sequence, we may combine Outputs 0 and 1 into a single output, and clearly $C_{\text{a-l}} \leq 1$ for channels with binary outputs. ∎

Proof of Theorem 3.1: It follows from (11) and Theorem 2.1 that $C_{\text{a-l,fb}}$ cannot exceed R_{cutoff}. It thus remains to show that $C_{\text{a-l,fb}} \geq R_{\text{cutoff}}$. We propose a 3-phase coding scheme. (We need some basic results about types, which can be found in [7, Chapter 2].) In the first phase, we send one of 2^{nR} messages using the length-n codewords $(\mathbf{x}_m)_{1 \leq m \leq 2^{nR}}$. In the second phase, after observing $\mathbf{y} \in \mathcal{Y}^n$ through the feedback link, we use a zero-error code of rate 1 to describe the type P of the codeword and the conditional type V of $\mathbf{y}$ given the codeword.[1] Since the total number of types and conditional types is polynomial in n, this requires at most $o(n)$ additional channel uses. Let $\mathcal{M}(\mathbf{y}, P, V) \subseteq \{1, \ldots, 2^{nR}\}$ denote the set of all messages whose codeword is of type P and given which $\mathbf{y}$ has conditional type V. We fix some $\alpha > 0$ and partition $\mathcal{M}(\mathbf{y}, P, V)$ into $2^{n\alpha}$ lists of lengths at most $\lceil 2^{-n\alpha}|\mathcal{M}(\mathbf{y}, P, V)|\rceil$. In the third phase, we send the index of the list containing the correct message using a zero-error code (of rate 1). This requires at most $\lceil n\alpha \rceil$ additional

---

[1] If the zero-error capacity of a DMC is positive, then it is at least 1.

channel uses. Using the inequality $\lceil \xi \rceil < 1 + \xi$ and the fact that

$$W^n(\mathbf{y}|\mathbf{x}) = 2^{-n(D(V\|W|P) + H(V|P))}, \quad \mathbf{x} \in \mathcal{M}(\mathbf{y}, P, V),$$

we can bound the average listsize of the coding scheme by

$$1 + \frac{\sum_{\mathbf{y}} \sum_{P,V} 2^{-n(D(V\|W|P)+H(V|P))}|\mathcal{M}(\mathbf{y}, P, V)|^2}{2^{n(R+\alpha)}}$$

$$\leq 1 + \frac{\sum_{\mathbf{y}} \left[\sum_{P,V} 2^{-\frac{n}{2}(D(V\|W|P)+H(V|P))}|\mathcal{M}(\mathbf{y}, P, V)|\right]^2}{2^{n(R+\alpha)}}$$

$$= 1 + 2^{-n(R+\alpha)} \sum_{\mathbf{y}} \left[\sum_m \sqrt{W^n(\mathbf{y}|\mathbf{x}_m)}\right]^2, \qquad (27)$$

where we used the inequality

$$\sum_i a_i^2 \leq \left[\sum_i a_i\right]^2, \quad a_i \geq 0. \qquad (28)$$

Averaging the sum on the RHS of (27) over codebooks of size 2^{nR} drawn IID from a PMF Q on $\mathcal{X}$, we obtain

$$\sum_{\mathbf{x}_1, \ldots, \mathbf{x}_{2^{nR}}} \left[\prod_m Q^n(\mathbf{x}_m)\right] \sum_{\mathbf{y}} \left[\sum_{m'} \sqrt{W^n(\mathbf{y}|\mathbf{x}_{m'})}\right]^2$$

$$= 2^{nR} + 2^{nR}(2^{nR}-1) \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} Q^n(\mathbf{x})\sqrt{W^n(\mathbf{y}|\mathbf{x})}\right]^2$$

$$= 2^{nR} + 2^{nR}(2^{nR}-1)2^{-nE_0(1,Q)}, \qquad (29)$$

where we arrived at the second line by expanding the square

$$\left[\sum_m \sqrt{W^n(\mathbf{y}|\mathbf{x}_m)}\right]^2$$

$$= \sum_m W^n(\mathbf{y}|\mathbf{x}_m) + \sum_{m \neq m'} \sqrt{W^n(\mathbf{y}|\mathbf{x}_m)}\sqrt{W^n(\mathbf{y}|\mathbf{x}_{m'})}.$$

Combining (27) and (29), we see that the average listsize, averaged over all realizations of the codebook, is upper bounded by

$$1 + 2^{-n\alpha} + 2^{-n(E_0(1,Q)+\alpha-R)}, \qquad (30)$$

and this tends to 1 provided that $R < E_0(1,Q) + \alpha$. Observe that the rate of the coding scheme approaches $R/(1+\alpha)$ as n tends to infinity. By letting $\alpha \to 0$, we thus see that all rates below $E_0(1,Q)$ are achievable. The proof is completed by choosing a Q that achieves the cutoff rate. ∎

IV. A LOWER BOUND ON $C_{\text{a-l,fb}}$

If $C_0 = 0$, then we cannot use Theorem 3.1. In such cases we propose the following lower bound.

Theorem 4.1: If $C_0 = 0$ and $C_{\text{a-l}} > 0$, then

$$C_{\text{a-l,fb}} \geq \frac{R^\star}{1 + \frac{R^\star}{\log \frac{1}{1-q^\star}}}, \qquad (31)$$

where

$$R^\star = \sup_{\rho \geq 0} \max_Q \frac{E_0(\rho, Q)}{1+\rho}, \qquad (32)$$

and where $q^\star$ is the maximum of $W(\mathcal{E}|x_1)$ taken over all $x_1 \in \mathcal{X}$ and over the subsets $\mathcal{E} \subset \mathcal{Y}$ for which there exists some $x_0 \in \mathcal{X}$ with $W(\mathcal{E}|x_0) = 0$.

Before we give a proof, some remarks are in order.

(i) If $C_{\text{a-l}} = 0$, then also $C_{\text{a-l,fb}} = 0$. Indeed, if $C_{\text{a-l}} = 0$, then every output that can be reached from some input can be reached from every input (see Section I). Consequently, no message can be ruled out based on the output sequence even when feedback is used.

(ii) Using the fact that $E_0(\rho, Q)$ is nondecreasing in ρ and $E_0(\rho, Q)/\rho$ is nonincreasing in ρ [4], we obtain

$$\frac{E_0(\rho, Q)}{1 + \rho} \leq \frac{E_0(1 + \rho, Q)}{1 + \rho}$$
$$\leq E_0(1, Q), \quad \rho \geq 0. \tag{33}$$

Consequently, $R^\star \leq R_{\text{cutoff}}$. Moreover, $C_0 = 0$ implies $q^\star < 1$, whereas $C_{\text{a-l}} > 0$ implies $q^\star > 0$. Thus, the lower bound in (31) is always strictly smaller than the cutoff rate and is interesting only when (12) does not hold.

(iii) We could have included the case $C_0 > 0$ with the resulting lower bound being $R^\star$, but this is not interesting in view of Theorem 3.1.

(iv) To see that Theorem 4.1 is useful, consider the channel in Figure 1(b). Using an argument similar to that given in Section III for the channel in Figure 1(a), we see that $C_{\text{a-l}} \leq 1$. Since the zero-error capacity of the channel is zero, Theorem 3.1 does not apply. But if $\epsilon \to 0$ and $\delta \to 1$, then $q^\star \to 1$ and $\max_Q E_0(\rho, Q) \to \rho \log 3$, so from Theorem 4.1 we obtain that $C_{\text{a-l,fb}} \to \log 3$. We will generalize this observation in Section V.

Proof of Theorem 4.1: Fix a positive integer L_0 and let $x_0, x_1, \mathcal{E}$ achieve $q^\star$. We propose a 3-phase coding scheme. In the first phase, we use a rate-R blocklength-n encoder (as in (1)) paired with a decoder that produces a list of the L_0 most likely messages given the received sequence $y_1, \ldots, y_n$ (resolving ties arbitrarily). As shown in [4, Exercise 5.20], for every PMF Q on $\mathcal{X}$ we can find a sequence of rate-R blocklength-n encoders for which the probability of the correct message not being on the list is at most $2^{-n(E_0(\rho, Q) - \rho R)}$ for every $0 \leq \rho \leq L_0$. Thanks to the feedback, the transmitter knows which messages are on the decoder's list, and if the correct message is among them, then it tries to convey this fact to the receiver in the second phase by sending n' times the symbol x_1; otherwise it sends n' times the symbol x_0. Accordingly, if at least one symbol in $\mathcal{E}$ is observed at the output during the second phase, then the receiver knows with certainty that the correct message is on the list (because $W(\mathcal{E}|x_0) = 0$); otherwise it ignores the third phase and produces a list of all 2^{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_0 auxiliary codewords $\mathbf{x}_1, \ldots, \mathbf{x}_{L_0}$ of length kL_0, where k is a positive integer, as follows. The components $(j-1)k+1, \ldots, jk$ of $\mathbf{x}_j$ 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 $\mathcal{E}$ is observed at the output during the third phase (because $W(\mathcal{E}|x_0) = 0$ and the x_1 patterns are disjoint); otherwise it produces the list of size L_0. 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 kL_0 times the symbol x_0.

To analyze the average listsize of this coding scheme, we define the events

$E_1 = \{$correct message not on the list after 1st phase$\}$,
$E_2 = \{$no symbol in $\mathcal{E}$ is observed in 2nd phase$\}$,
$E_3 = \{$no symbol in $\mathcal{E}$ is observed in 3rd phase$\}$.

The average listsize can then be bounded as

$$\mathrm{E}[L] \leq 1 + \mathrm{E}[L|E_1]\Pr(E_1) + \mathrm{E}[L|E_1^c \cap E_2]\Pr(E_1^c \cap E_2)$$
$$+ \mathrm{E}[L|E_1^c \cap E_2^c \cap E_3]\Pr(E_1^c \cap E_2^c \cap E_3). \tag{34}$$

We bound the RHS of (34) term by term, beginning with

$$\mathrm{E}[L|E_1]\Pr(E_1) \leq 2^{nR}2^{-n(E_0(\rho, Q) - \rho R)}$$
$$= 2^{-n(1+\rho)\left(\frac{E_0(\rho, Q)}{1+\rho} - R\right)}, \quad 0 \leq \rho \leq L_0.$$

The RHS approaches zero as $n \to \infty$ provided that $R < R^\star$ and L_0 is large enough so that we can pick a ρ in the interval $[0, L_0]$ and a PMF Q that achieve a value of $E_0(\rho, Q)/(1 + \rho)$ close enough to the supremum in the definition of $R^\star$. The next term on the RHS of (34) can be bounded as

$$\mathrm{E}[L|E_1^c \cap E_2]\Pr(E_1^c \cap E_2) \leq 2^{nR}(1 - q^\star)^{n'}$$
$$= 2^{nR}2^{-n' \log \frac{1}{1-q^\star}}. \tag{35}$$

The RHS of (35) approaches zero as $n \to \infty$ if we choose

$$n' = n(1 + \delta)\frac{R}{\log \frac{1}{1-q^\star}} \tag{36}$$

for an arbitrarily small $\delta > 0$. Finally,

$$\mathrm{E}[L|E_1^c \cap E_2^c \cap E_3]\Pr(E_1^c \cap E_2^c \cap E_3) \leq L_0(1 - q^\star)^k, \tag{37}$$

and the RHS can be made arbitrarily small by choosing k sufficiently large. Observe that the rate of the scheme is

$$\frac{R}{1 + \frac{n'}{n} + \frac{kL_0}{n}}. \tag{38}$$

Choosing first L_0 sufficiently large, then R close to $R^\star$, then n' as in (36) with δ sufficiently small, then k sufficiently large, and finally n sufficiently large shows that that all rates strictly less than the RHS of (31) are achievable. ∎

V. Low-Noise Channels

For $0 < \epsilon < 1$ let $\mathcal{W}_\epsilon(\mathcal{X})$ denote the class of DMCs with input and output alphabet equal to $\mathcal{X}$ and transition law satisfying the low-noise condition

$$W(x|x) \geq 1 - \epsilon, \quad x \in \mathcal{X}. \tag{39}$$

For a similar class of channels, Ahlswede et al. showed that $C_{\text{a-l}}$ approaches the Sperner capacity of the channel graph as ϵ tends to zero [1]. With feedback we can do better:

Theorem 5.1: If $W_\epsilon \in \mathcal{W}_\epsilon(\mathcal{X})$ and $C_{\text{a-l}}(W_\epsilon) > 0$, then

$$\lim_{\epsilon \to 0} C_{\text{a-l,fb}}(W_\epsilon) = \log|\mathcal{X}|. \tag{40}$$

Proof: If $W_\epsilon \in \mathcal{W}_\epsilon(\mathcal{X})$ and $C_{\text{a-l}}(W_\epsilon) > 0$, then, in the notation of Theorem 4.1, $q^\star \geq 1 - \epsilon$. The result now follows from Theorem 4.1 by noting that $\lim_{\epsilon \to 0} \max_Q E_0(\rho, Q) = \rho \log|\mathcal{X}|$ for all channels in $\mathcal{W}_\epsilon(\mathcal{X})$. ∎

VI. A Multi-Letter Characterization of $C_{\text{a-l}}$ Based on Forney's Lower Bound

Forney derived the lower bound [8]

$$C_{\text{a-l}} \geq \max_Q \log \frac{1}{\sum_y (QW)(y)Q(\mathcal{X}(y))}, \qquad (41)$$

where we use the notation

$$(QW)(y) = \sum_x Q(x)W(y|x) \qquad (42)$$

and

$$Q(\mathcal{X}(y)) = \sum_{x:W(y|x)>0} Q(x). \qquad (43)$$

The bound can be strengthened by applying it to W^n and normalizing the result by $1/n$. The next result shows that the strengthened bound is asymptotically tight:

Theorem 6.1: For any DMC,

$$C_{\text{a-l}} = \lim_{n\to\infty} \frac{1}{n} \max_Q \log \frac{1}{\sum_{\mathbf{y}} (QW^n)(\mathbf{y})Q(\mathcal{X}^n(\mathbf{y}))}, \qquad (44)$$

where the maximum can be restricted to PMFs that are uniform over subsets of $\mathcal{X}^n$.

A different multi-letter characterization of $C_{\text{a-l}}$ was given in [1], [2].

Proof of Theorem 6.1: That the upper limit of the sequence on the RHS of (44) is a lower bound to $C_{\text{a-l}}$ follows from (41). To prove a converse, fix a sequence $(\mathbf{f}_n)_{n\geq 1}$ of rate-R blocklength-n encoders with average listsize tending to one. For simplicity assume that each $\mathbf{f}_n$ is one-to-one.[2] Let Q_n be the uniform PMF on the range of $\mathbf{f}_n$. Then

$$Q_n(\mathcal{X}^n(\mathbf{y})) = 2^{-nR}L(\mathbf{y}), \qquad (45)$$

where $L(\mathbf{y})$ is defined as in (4). For the logarithm of the average listsize we then have

$$\log \frac{1}{2^{nR}} \sum_m \sum_{\mathbf{y}} L(\mathbf{y})W^n\big(\mathbf{y}\big|\mathbf{f}_n(m)\big)$$
$$= nR + \log \sum_{\mathbf{y}} (Q_n W^n)(\mathbf{y})Q_n\big(\mathcal{X}^n(\mathbf{y})\big). \qquad (46)$$

Since the average listsize tends to one, it follows from (46) upon dividing by n and letting $n \to \infty$ that

$$R = \lim_{n\to\infty} \frac{1}{n} \log \frac{1}{\sum_{\mathbf{y}} (Q_n W^n)(\mathbf{y})Q_n(\mathcal{X}^n(\mathbf{y}))}, \qquad (47)$$

and the RHS of (47) is upper bounded by the lower limit of the sequence on the RHS of (44). ∎

VII. Final Remarks

It is interesting to note that the relationship between $C_{\text{a-l}}$ and R_{cutoff} is analogous to the relationship between $C_{\text{0-u}}$ and C, and the relationship between $C_{\text{a-l}}$ and $C_{\text{0-u}}$ is analogous to the relationship between R_{cutoff} and C:

(i) $C_{\text{a-l}} > 0$ iff $C_{\text{0-u}} > 0$, and $R_{\text{cutoff}} > 0$ iff $C > 0$.

---

[2]The result remains true without this assumption.

(ii) $C_{\text{a-l}} \leq R_{\text{cutoff}}$ and $C_{\text{0-u}} \leq C$. Both hold with equality whenever (12) holds.
(iii) $C_{\text{a-l}} \leq C_{\text{0-u}}$ and $R_{\text{cutoff}} \leq C$.
(iv) $C_{\text{0-u,fb}} > 0$ only if $C_{\text{0-u}} > 0$, and $C_{\text{a-l,fb}} > 0$ only if $C_{\text{a-l}} > 0$.
(v) Feedback increases neither C nor R_{cutoff}, but it can increase $C_{\text{0-u}}$ [5] and $C_{\text{a-l}}$.
(vi) There are single-letter expressions for C and R_{cutoff}, but none are currently known for $C_{\text{0-u}}$ and $C_{\text{a-l}}$.

It was shown in [5], [9] that $C_{\text{0-u,fb}}$ equals C whenever $C_{\text{0-u}} > 0$. One might thus be tempted to conjecture that $C_{\text{a-l,fb}}$ equals R_{cutoff} whenever $C_{\text{a-l}} > 0$. This, however, is not true. The assumption that $C_0 > 0$ in Theorem 3.1 cannot be replaced with $C_{\text{a-l}} > 0$:

Proposition 7.1: A positive value of $C_{\text{a-l}}$ does not guarantee that $C_{\text{a-l,fb}} = R_{\text{cutoff}}$.

Proof: A counterexample is the channel in Figure 1(b). It has $C_{\text{a-l}} > 0$, and for small ϵ the cutoff rate is at least close to 1. If the received sequence contains only the symbols 0 and 1, then the decoder cannot rule out any of the messages and produces a list of length 2^{nR}. But regardless of the choice of the encoder, the probability of observing only the symbols 0 and 1 at the output is at least $(1-\delta)^n$. Thus, the expected listsize is at least $2^{n(R+\log(1-\delta))}$, and $\log(1-\delta)$ is close to zero for small $\delta > 0$. ∎

As a final remark, many of the results in this paper can be generalized to the *ρ-th moment cutoff rate and the ρ-th moment listsize capacity*, whose definitions are analogous to those of the cutoff rate and the average-listsize capacity except that $L(\mathbf{y})$ and $L(\mathbf{y}, m)$ are replaced with $L(\mathbf{y})^\rho$ and $L(\mathbf{y}, m)^\rho$ for some positive number ρ [10].

References

[1] 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, 1996.

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

[3] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[4] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[5] C. Bunte and A. Lapidoth, "The zero-undetected-error capacity of discrete memoryless channels with feedback," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1838–1842.

[6] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, 1995.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York: Cambridge University Press, 2011.

[8] G. Forney Jr, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[9] 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, 2012.

[10] C. Bunte and A. Lapidoth, manuscript in preparation.