Capacity of Burst Noise-Erasure Channels With and Without Feedback and Input Cost

Lin Song, Fady Alajaji, and Tamás Linder

Abstract

A class of burst noise-erasure channels which incorporate both errors and erasures during transmission is studied. The channel, whose output is explicitly expressed in terms of its input and a stationary ergodic noise-erasure process, is shown to have a so-called "quasi-symmetry" property under certain invertibility conditions. As a result, it is proved that a uniformly distributed input process maximizes the channel's block mutual information, resulting in a closed-form formula for its non-feedback capacity in terms of the noise-erasure entropy rate and the entropy rate of an auxiliary erasure process. The feedback channel capacity is also characterized, showing that feedback does not increase capacity and generalizing prior related results. The capacity-cost function of the channel with and without feedback is next investigated. A sequence of finite-letter upper bounds for the capacity-cost function without feedback is derived. Finite-letter lower bonds for the capacity-cost function with feedback are obtained using a specific encoding rule. Based on these bounds, it is demonstrated both numerically and analytically that feedback can increase the capacity-cost function for a class of channels with Markov noise-erasure processes.

Index Terms

Channels with burst errors and erasures, channels with memory, channel symmetry, non-feedback and feedback capacities, non-feedback and feedback capacity-cost functions, input cost constraints, stationary ergodic and Markov processes.

I. INTRODUCTION

The stationary memoryless binary erasure channel (BEC) and the binary symmetric channel (BSC) play fundamental roles in information theory, since they model two types of common channel distortions in digital communications. In a BEC, at each time instance, the transmitter sends a bit (0 or 1) and the receiver either gets the bit correctly or as an erasure denoted by the symbol "e." The BEC models communication systems where signals are either transmitted noiselessly or lost. The loss may be caused by packet collisions, buffer overflows, excessive delay, or corrupted data. In a BSC, the transmitter similarly sends a bit, but the receiver obtains it either correctly or flipped. The BSC is a standard model for binary communication systems with noise. For example, in a memoryless additive Gaussian noise channel used with antipodal signaling and hard-decision demodulation, when the noise level is high, a decision error may occur at the receiver which is characterized by flipping the transmitted bit in the system's BSC representation. As opposed to the BSC, the BEC is, in a sense, noiseless. However in realistic systems, erasures and errors usually co-exist and often occur in bursts due to their time-correlated statistical behavior. In this paper, we introduce the q-ary noise-erasure channel (NEC) with memory which incorporates both erasures and noise. This model, which subsumes both the BEC and the BSC, as well as their extensions with non-binary alphabets and memory, provides a useful model for real-world channels, where data packets can be corrupted or dropped in a bursty fashion. Such channels include wireless systems where the receiver can identify deep fades and designate them as erasure bursts (while less detrimental fades are treated as regular error-prone transmissions), hybrid internet-wireless communications and magnetic storage and recording devices [2]–[4].

A. The burst erasure and additive noise channels

Given integer $q \geq 2$, let $X_i \in \mathcal{X} = \{0, 1, 2, .., q-1\} \triangleq \mathcal{Q}$ denote the channel input at time i and $Y_i \in \mathcal{Y} = \mathcal{Q} \cup \{e\}$ denote the corresponding channel output (we assume throughout that $e \notin \mathcal{Q}$). For the general q-ary burst erasure channel (EC), the input-output relationship can be expressed by

$$Y_i = X_i \cdot \mathbf{1}\{\tilde{Z}_i \neq e\} + e \cdot \mathbf{1}\{\tilde{Z}_i = e\}, \text{ for } i = 1, 2, \ldots,$$

where $\{\tilde{Z}_i\}_{i=1}^{\infty}$ is a correlated erasure process (which is independent of the message conveyed by the input sequence) with alphabet $\{0, e\}$, $\mathbf{1}(\cdot)$ is the indicator function, and by definition

$a + 0 = a$, $a \cdot 0 = 0$, and $a \cdot 1 = a$ for all $a \in \mathcal{Q} \cup \{e\}$. When $\{\tilde{Z}_i\}_{i=1}^{\infty}$ is stationary memoryless (i.e., independent and identically distributed) and $q = 2$, the channel reduces to the BEC. The above burst EC also includes the Gilbert-Elliott erasure model (e.g., [5]–[7]) as a special instance. In this case, the erasure process $\{\tilde{Z}_i\}$ is a hidden Markov source driven by a two-state Markov chain according to the well-known Gilbert-Elliott model [8]–[10] (where each state is governed by a BEC). The performance of coding techniques for burst ECs has been extensively studied; see for example [2], [4]–[7], [11]–[13] and the references therein. Channel capacity studies include [14] and [15], where the feedback and non-feedback capacities of BECs with no-consecutive-ones at the input were respectively investigated. Furthermore, explicit computations of the feedback and non-feedback capacities of energy harvesting BECs were given in [16], where it was shown that feedback increases the capacity of such channels.

A discrete q-ary additive noise channel (ANC) with memory has identical input and output alphabets $\mathcal{X} = \mathcal{Y} = \mathcal{Q}$ and is described as $Y_i = X_i \oplus_q Z_i$ for $i = 1, 2, \cdots$, where $\{Z_i\}_{i=1}^{\infty}$ is a q-ary correlated noise process (that is independent of the input message) and $\oplus_q$ denotes modulo-q addition. The BSC is a special case of the ANC: when $\{Z_i\}_{i=1}^{\infty}$ is binary-valued and memoryless, the ANC reduces to the BSC. Furthermore, the Gilbert-Elliott burst noise channel [8]–[10] (whose states are each governed by a BSC) and the more recent infinite and finite-memory Polya contagion channel [17] and its queue-based variation [18] are interesting instances of the ANC, which have been used to model time-correlated fading channels (e.g., see [19], [20] and related work). In [21], it was shown that feedback does not increase the capacity of ANCs with arbitrary noise memory. In particular, denoting the capacity with and without feedback by C_{FB}^{ANC} and C^{ANC}, respectively, it is proved in [21] that $C^{\text{ANC}} = C_{FB}^{\text{ANC}} = \log q - \overline{H}_{sp}(\boldsymbol{Z})$, where $\overline{H}_{sp}(\boldsymbol{Z})$ denotes the spectral sup-entropy rate [22], [23] of the noise process $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$. The result of [21], which can also be proved for a larger class of channels [24], was recently extended in [25] to the family of compound channels with additive noise. Furthermore, it was shown in [26] that feedback can increase the capacity-cost function of an ANC with Markov noise.

B. NEC model: a burst channel for both errors and erasures

In this paper, we consider the NEC, a channel with both burst erasures and errors whose output $Y_i \in \mathcal{Y} = \mathcal{Q} \cup \{e\}$ at time i is given by

$$Y_i = h(X_i, Z_i) \cdot \mathbf{1}\{Z_i \neq e\} + e \cdot \mathbf{1}\{Z_i = e\} \triangleq \theta(X_i, Z_i) \tag{1}$$

where $X_i \in \mathcal{X} = \mathcal{Q}$ is the input, $\{Z_i\}_{i=1}^n \triangleq \mathbf{Z}$ is a noise-erasure process with alphabet $\mathcal{Z} = \{0, ..., q'-1\} \cup \{e\} \triangleq \mathcal{Q}' \cup \{e\}$, $1 \leq q' \leq q$, which is independent of the input message, and $h : \mathcal{Q} \times \mathcal{Q}' \to \mathcal{Q}$ is a deterministic function. Note that strictly speaking, $h(x_i, z_i)$ is undefined when $z_i = e$. However, since $\mathbf{1}\{z_i \neq e\} = 0$ when $z_i = e$, this is remedied by setting the product of an undefined quantity and zero as equal to zero. Indeed, (1) means that

$$Y_i = \begin{cases} h(X_i, Z_i), & \text{if } Z_i \neq e, \\ e, & \text{if } Z_i = e. \end{cases}$$

Setting $q' = 1$ and $h(x, z) = x$ for all $z \in \mathcal{Z}$, reduces the NEC to the EC. Setting $q' = q$, $h(x, z) = x \oplus_q z$, and $P_{Z_i}(e) = 0$, turns the NEC into the ANC. Also, a Gilbert-Elliott burst model combining (in general non-binary) errors with erasures is an example of an NEC (in such model, each state is governed by a memoryless channel whose inputs can be received in error or erased). Finally, we note that the NEC resembles the cascade channel of [4], where its non-feedback capacity is derived and its performance under low-density parity-check coding is analyzed. This cascade channel is motivated by magnetic data storage systems and it consists of a concatenation of an indecomposable finite-state channel with a burst erasure channel where symbols are erased consecutively in each output sequence. The NEC differs from the cascade model of [4] in a number of aspects, including the fact that the NEC is governed via an explicit input-output functional relationship in terms of a single noise-erasure process as given in (1) without requiring consecutive erasures in the output sequences while the cascade model has independent error and erasure processes emanating from its separate channel components.

We study the non-feedback and feedback capacities and capacity-cost functions of the NEC under certain invertibility conditions on the function h in (1). The class of NECs satisfying these conditions readily include the ANC, the EC, the discrete symmetric channel of [24] (with identical input, noise and output alphabets) and the following channel which differs significantly from the latter three channels.

Data storage channel: This channel is described by (1) where we set $q' = 2$, q to be a power of 2, and the function $h(\cdot, \cdot)$ to be

$$h(x, z) = x \cdot \mathbf{1}\{z = 0\} + (q - 1 - x) \cdot \mathbf{1}\{z = 1\}. \tag{2}$$

This channel can model storage devices with errors and erasures where the data is stored in binary form, using a natural binary code (NBC) of length $\log_2 q$ bits. When $z = e$, an erasure occurs. In the non-erasure mode, if $z = 0$, no error occurs and the storage device returns x

perfectly, and if $z = 1$, a hard failure occurs causing the storage device to flip all the NBC bits representing x, which is equivalent to the operation $q - 1 - x$.

In general, the capacity of well-behaving channels with memory (such as stationary information stable channels) is given as the limit of the n-fold mutual information sequence [22], [27]–[29], while the feedback capacity is expressed via the limit of the n-fold directed information [30]–[34]. For some special cases, single-letter expressions or exact values of such capacities can be obtained. Examples of channels where the feedback capacity is explicitly determined include the ANC [21], the finite-state channel with states known at both transmitter and receiver [35], the trapdoor channel [36], the Ising channel [37], the symmetric finite-state Markov channel [38], and the BEC [14] and the binary-input binary-output channel [39] with both channels subjected to a no consecutive ones input constraint.

C. Contributions

In this paper, we introduce an auxiliary erasure process $\{\tilde{Z}_i\}_{i=1}^{\infty} \triangleq \tilde{\boldsymbol{Z}}$, a binary process defined via the noise-erasure process $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$, and we prove that the non-feedback capacity of the NEC with a stationary ergodic noise-erasure process is given by $(1 - \varepsilon) \log q - [\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})]$ (Theorem 1), where $\bar{H}(\cdot)$ denotes entropy rate and $\varepsilon = P_{Z_i}(e)$ is the probability of an erasure which is defined in Section III. The proof consists of showing, via two intermediate lemmas (Lemmas 7 and 8) that make use of the structure of the channel function h in (1), that the n-fold NEC is quasi-symmetric (as per Definition 6) and hence its n-fold mutual information is maximized by a uniformly distributed input process. The derived NEC capacity formula recovers the capacity expressions of the ANC and the EC, when the NEC is specialized to the latter channels. We briefly explore the calculation of the capacity for Markov noise-erasure processes. We further show that, unlike the EC, for which memory in the erasure process does not increase capacity (e.g., see [5], [11]), the capacity of the NEC is strictly larger than the capacity of its memoryless counterpart (i.e., a channel with a memoryless noise-erasure process with identical marginal distribution as the NEC's stationary ergodic noise-erasure process) for non-trivial correlated noise-erasure processes such as non-degenerate stationary, irreducible and aperiodic Markov processes. We also investigate the NEC with ideal output feedback. We prove a converse for the feedback capacity and show that the feedback capacity coincides with the non-feedback capacity (Theorem 2). This shows that feedback does not increase the capacity of the NEC and generalizes the feedback capacity results of [21] and [24].

The capacity-cost functions of the NEC with and without feedback is next investigated. We establish a sequence of finite letter upper bounds on the capacity-cost function without feedback (Theorem 3) and a sequence of finite letter lower bounds on the capacity-cost function with feedback based on a constructive feedback encoding rule and an achievability result (Theorem 4). For a class of NECs with stationary irreducible and aperiodic Markov noise-erasure processes with transition probability matrices satisfying some uniformity conditions for one of their rows and for the column corresponding to the erasure state, we prove that feedback does increase the capacity-cost function in a certain cost range (Theorem 5). This result, which generalizes a similar result in [26] for the ANC, is further demonstrated to hold for more general NECs by numerically comparing the lower bound of the capacity-cost function with feedback and the upper bound of the capacity-cost function without feedback. Finally, we point out that the proof techniques of this paper are significant extensions of the methods used for the ANC in the derivation of the feedback capacity [21] and the study of the capacity-cost function with and without feedback [26]. We note that, in addition to allowing for erased output symbols, the NEC can significantly differ from the ANC as its noise alphabet can be much smaller than the input alphabet (e.g., see the data storage example in (2)). Furthermore, while the ANC is symmetric [45], the NEC satisfies a considerably weaker notion of symmetry, quasi-symmetry. These differences necessitate a technically more sophisticated approach for proving the capacity results.

The rest of this paper is organized as follows. We first provide preliminary results in Section II. In Section III, we present the invertibility properties imposed on the NEC and derive the NEC non-feedback capacity. We also examine the calculation of the capacity expression under Markov noise-erasure processes and the effect of memory on the NEC capacity. In Section IV, we study the feedback capacity of the NEC and show that feedback does not increase capacity. We investigate the NEC capacity-cost functions with and without feedback in Sections V and VI, respectively. We conclude the paper in Section VII.

II. PRELIMINARIES

A. Non-Feedback/Feedback Capacity and Capacity-Cost Function

We use capital letters such as X, Y, and Z to denote random variables and the corresponding script letters $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ to denote their alphabets. The distribution of X is denoted by P_X, where the subscript may be omitted if there is no ambiguity. In this paper, all random variables

have finite alphabets. A channel $\boldsymbol{W}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is statistically modeled as a sequence of conditional distributions $\boldsymbol{W} = \{W^n(\cdot|\cdot)\}_{n=1}^{\infty}$, where $W^n(\cdot|x^n)$ is a probability distribution on $\mathcal{Y}^n$ for every $x^n \in \mathcal{X}^n$, which we call the n-fold channel of $\boldsymbol{W}$. Finally, let X^n and Y^n denote the n-fold channel's input and output sequences, respectively, where $X^n = (X_1, X_2, ..., X_n)$ and $Y^n = (Y_1, Y_2, ..., Y_n)$.

Definition 1: A feedback channel code with block length n and rate $R \geq 0$ consists of a sequence of encoding functions $f_i^{(n)} : \mathcal{M} \times \mathcal{Y}^{i-1} \to \mathcal{X}$ for $i = 1, ..., n$ and a decoder $g^{(n)} : \mathcal{Y}^n \to \mathcal{M}$, where $\mathcal{M} = \{1, 2, ..., 2^{nR}\}$ is the message set.

When there is no feedback, the sequence of encoders simplifies to the sequence $f^{(n)} : \mathcal{M} \to \mathcal{X}^n$ of encoders whose domain is just the message set. The encoder conveys message M, which is uniformly distributed over $\mathcal{M}$, by sending the sequence X^n over the channel which in turn is received as Y^n at the receiver. For the non-feedback case, $X^n = f^{(n)}(M)$, while for the feedback case, the encoder takes into account the previously received channel outputs and sends $X_i = f_i^{(n)}(M, Y^{i-1})$ for $i = 1, \cdots, n$. Upon estimating the sent message via $g^{(n)}(Y^n)$, the resulting decoding error probability is $P_e^{(n)} = \Pr(g^{(n)}(Y^n) \neq M)$.

In general, the use of the channel is not free. For example, a binary *on-off keyed* physical channel emits a pulse signal when sending the bit 1 (which requires a certain expenditure of energy) and stays idle (using no energy) when sending the bit 0 (e.g., [40]); this results in different cost constraints on the input alphabet of the equivalent discrete channel. Let $b : \mathcal{X} \to \mathbb{R}$ be a cost function and define the cost of an input sequence x^n as $b(x^n) = \sum_{i=1}^n b(x_i)$ [41].

Definition 2: A channel code with block length n and rate R for the n-fold channel of $\boldsymbol{W}$ is β-admissible if $b(x^n) \leq n\beta$ for all codewords x^n in the codebook $\mathcal{C}$ which, when there is no feedback, is given by

$$\mathcal{C} = \left\{ x^n \in \mathcal{X}^n : x^n = f^{(n)}(m) \text{ for some } m \in \mathcal{M} \right\},$$

while, when there is feedback, is given by

$$\mathcal{C} = \left\{ x^n \in \mathcal{X}^n : x_i = f_i^{(n)}(m, y^{i-1}), i = 1, ..., n, m \in \mathcal{M}, y^n \in \mathcal{Y}^n, W^n(y^n|x^n) \neq 0 \right\}.$$

Definition 3: The feedback capacity-cost function of a channel, denoted by $C_{FB}(\beta)$, is the supremum of all rates R for which there exists a sequence of β-admissible feedback channel codes with block length n and rate R, such that $\lim_{n \to \infty} P_e^{(n)} = 0$.

The non-feedback capacity-cost function, feedback capacity, and non-feedback capacity are defined similarly and are denoted by $C(\beta)$, C_{FB}, and C, respectively. When there is no cost

constraint, or equivalently $\beta = \infty$, the capacity-cost function (with or without feedback) reduces to the capacity (with or without feedback).

Recall that the channel $\boldsymbol{W}$ is memoryless if $W^n(y^n|x^n) = \Pi_{i=1}^n W^1(y_i|x_i)$ for all $n \geq 1$, $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, when there is no feedback. Thus, a memoryless channel is defined by its input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $W^1(y|x)$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. For memoryless channels, the superscript "1" is usually omitted. Shannon's channel coding theorem [42] establishes that

$$C = \max_{P_X} I(X;Y) \tag{3}$$

for memoryless channels, where $I(X;Y)$ is the mutual information between X and Y. This coding theorem can be extended to show that (e.g., see [22], [27]–[29], [43])

$$C = \sup_n C_n = \lim_{n \to \infty} C_n \tag{4}$$

for stationary and information stable channels,[1] where

$$C_n = \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^n).$$

For memoryless channels, the feedback and non-feedback capacities are equal [44]. In general, $C_{FB} \geq C$, since the class of feedback codes includes non-feedback codes as a special case, and $C_{FB} > C$ for certain channels with memory.

Definition 4: For an input random vector X^n with distribution P_{X^n}, the average cost of sending X^n over the channel is defined by

$$E[b(X^n)] = \sum_{x^n} P_{X^n}(x^n) b(x^n) = \sum_{i=1}^n E[b(X_i)].$$

Definition 5: The distribution P_{X^n} of an n-dimensional input random vector X^n that satisfies

$$\frac{1}{n} E[b(X^n)] \leq \beta$$

is called a β-admissible input distribution. We denote the set of n-dimensional β-admissible input distributions by $\tau_n(\beta)$:

$$\tau_n(\beta) = \{P_{X^n} : \frac{1}{n} E[b(X^n)] \leq \beta\}.$$

---

[1]In this paper we focus on stationary and information stable channels. A channel is stationary if every stationary channel input process results in a stationary joint input-output process. Furthermore, loosely speaking, a channel is information stable if the input process that maximizes the channel's block mutual information yields a joint input-output process that behaves ergodically (see for example [27], [28], [43] for a precise definition). Note that $\sup_n C_n = \lim_{n \to \infty} C_n$ holds since the sequence $\{nC_n\}_{n=1}^\infty$ is superadditive in light of the channel stationarity (e.g., see [29, Lemma 2, pp. 112-113]).

The capacity-cost function of stationary information stable channels is given by (e.g., [26], [41])

$$C(\beta) = \sup_n C_n(\beta) = \lim_{n \to \infty} C_n(\beta), \tag{5}$$

where $C_n(\beta)$ is the nth capacity-cost function given by

$$C_n(\beta) \triangleq \max_{P_{X^n} \in \tau_n(\beta)} \frac{1}{n} I(X^n; Y^n). \tag{6}$$

Lemma 1: [41, p. 51] The nth capacity-cost function $C_n(\beta)$ is concave and strictly increasing in β for $\beta_{\min} \leq \beta \leq \beta_{\max}^{(n)}$ and is equal to C_n for $\beta \geq \beta_{\max}^{(n)}$, where

$$\beta_{\min} \triangleq \min_{x \in \mathcal{X}} b(x),$$

and

$$\beta_{\max}^{(n)} \triangleq \min_{P_{X^n}} \left\{ \frac{1}{n} E[b(X^n)] : \frac{1}{n} I(X^n; Y^n) = C_n \right\}.$$

Lemma 2: [26, Lemma 2] The capacity-cost function $C(\beta)$ given by (5) is concave and strictly increasing in β for $\beta_{\min} \leq \beta \leq \beta_{\max}$, and is equal to C for $\beta \geq \beta_{\max}$, where

$$\beta_{\max} \triangleq \min_{P_{\boldsymbol{X}}} \left\{ \lim_{n \to \infty} \frac{1}{n} \mathbf{E}[b(X^n)] : \lim_{n \to \infty} \frac{1}{n} I(X^n; Y^n) = C \right\}$$

and $\boldsymbol{X}$ denotes the random process $\{X_i\}_{i=1}^{\infty}$.

B. Quasi-symmetry

In general, the optimization problem in (3) is difficult to solve analytically. However, it is shown in [29], [41], [45] that when the channel satisfies certain "symmetry" properties, the optimal input distribution in (3) is uniform and the channel capacity can be expressed in closed-form. This result was further extended to so-called "quasi-symmetric" channels in [46].

The transition matrix of a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $\{W(y|x)\}$ is the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix $\mathbb{Q}$ with the entry $W(y|x)$ in the xth row and yth column. For simplicity, let $p_{x,y} \triangleq W(y|x)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

A DMC is *symmetric* if the rows of its transition matrix $\mathbb{Q}$ are permutations of each other and the columns of $\mathbb{Q}$ are permutations of each other. The DMC is *weakly-symmetric* if the rows of $\mathbb{Q}$ are permutations of each other and all the column sums of $\mathbb{Q}$ are identical [41], [45].

Lemma 3 ([41], [45]): The capacity of a weakly-symmetric DMC is attained by the uniform input distribution and is given by $C = \log |\mathcal{Y}| - H(p_1, p_2, ..., p_{|\mathcal{Y}|})$, where $(p_1, p_2, ..., p_{|\mathcal{Y}|})$ is an arbitrary row of $\mathbb{Q}$ and $H(p_1, p_2, ..., p_{|\mathcal{Y}|}) = -\sum_{i=1}^{|\mathcal{Y}|} p_i \log p_i$.

It readily follows that a symmetric DMC is weakly-symmetric. We also note that Gallager's notion for a symmetric channel [29, p. 94] is a generalization of the above symmetry definition in terms of partitioning $\mathbb{Q}$ into symmetric sub-matrices. In turn, Gallager-symmetry is subsumed by the notion of quasi-symmetry below.

Definition 6 ([46]): A DMC with transition matrix $\mathbb{Q}$ is *quasi-symmetric* if, for some $m \geq 1$, $\mathbb{Q}$ can be partitioned along its columns into m weakly-symmetric sub-matrices, $\tilde{\mathbb{Q}}_1, \tilde{\mathbb{Q}}_2, ..., \tilde{\mathbb{Q}}_m$, where $\tilde{\mathbb{Q}}_i$ is a sub-matrix of size $|\mathcal{X}| \times |\mathcal{Y}_i|$ for $i = 1, ..., m$, with $\mathcal{Y}_1 \cup ... \cup \mathcal{Y}_m = \mathcal{Y}$ and $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, for any $i \neq j$, $i, j = 1, 2, ..., m$.

Lemma 4 ([46]): The capacity of a quasi-symmetric DMC is attained by the uniform input distribution and is given by

$$C = \sum_{i=1}^{m} a_i C_i,$$

where, for $i = 1, \cdots, m$, $a_i \triangleq \sum_{y \in \mathcal{Y}_i} p_{x,y}$ is the sum of any row $(p_{x,y_1}, p_{x,y_2}, \cdots, p_{x,y_{|\mathcal{Y}_i|}})$ of $\mathbb{Q}_i$ (corresponding to an arbitrary input symbol $x \in \mathcal{X}$), and

$$C_i = \log |\mathcal{Y}_i| - H\left(\text{any row of } \frac{1}{a_i}\mathbb{Q}_i\right) = \log |\mathcal{Y}_i| - H\left(\frac{p_{x,y_1}}{a_i}, \frac{p_{x,y_2}}{a_i}, \cdots, \frac{p_{x,y_{|\mathcal{Y}_i|}}}{a_i}\right)$$

is the capacity of the ith weakly-symmetric sub-channel whose transition matrix is $\frac{1}{a_i}\mathbb{Q}_i$.

III. NEC NON-FEEDBACK CAPACITY

We study a class of NECs with memory as defined in (1) for which the function $h : \mathcal{Q} \times \mathcal{Q}' \to \mathcal{Q}$ satisfies the following invertibility conditions:[2]

- (S-I) Given any $x \in \mathcal{Q}$, the function $h(x, \cdot) : \mathcal{Q}' \to \mathcal{Y}_x$ is invertible with inverse $\tilde{h}(x, \cdot) : \mathcal{Y}_x \to \mathcal{Q}'$, where $\mathcal{Y}_x \triangleq \{y \in \mathcal{Q} : \exists\, z \in \mathcal{Q}' \text{ such that } h(x, z) = y\}$.
- (S-II) Given any $y \in \mathcal{Q}$, the function $\tilde{h}(\cdot, y) : \mathcal{X}_y \to \mathcal{Q}'$ invertible, where $\mathcal{X}_y \triangleq \{x \in \mathcal{Q} : \exists\, z \in \mathcal{Q}' \text{ such that } h(x, z) = y\}$.

The above properties enable us to obtain the channel's noise-erasure from the input and output values via the relationship

$$z_i = \tilde{h}(x_i, y_i) \cdot \mathbf{1}\{y_i \neq e\} + e \cdot \mathbf{1}\{y_i = e\} \tag{7}$$

where $\tilde{h} : \mathcal{A} \to \mathcal{Q}'$ with

$$\mathcal{A} \triangleq \bigcup_{x \in \mathcal{Q}} \{x\} \times \mathcal{Y}_x = \bigcup_{y \in \mathcal{Q}} \mathcal{X}_y \times \{y\} \subseteq \mathcal{Q} \times \mathcal{Q},$$

---

[2]These conditions are similar to but more general than the ones considered in [24].

or equivalently,

$$\mathcal{A} = \{(x, y) \in \mathcal{Q} \times \mathcal{Q} : \exists\, z \in \mathcal{Q}' \text{ such that } h(x, z) = y\}.$$

For any input-output pair $(x, y) \in \mathcal{Q} \times \mathcal{Q}$ with $(x, y) \notin \mathcal{A}$, there is no $z \in \mathcal{Q}'$ such that $y = h(x, z)$; thus for such a (x, y) pair, we have $\Pr(X_i = x, Y_i = y) = 0$.

It can be straightforwardly verified that examples of NECs that satisfy conditions S-I and S-II above include the EC, the ANC, the discrete symmetric channel of [24], and the data storage channel described in (2). In this paper, the noise-erasure process $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ is assumed to be stationary and ergodic and independent of the transmitted message. Throughout the paper, except in Section V, it is assumed that the NEC satisfies properties S-I and S-II above. We next present our first main result, which is proved at the end of this section.

Theorem 1: The capacity of an NEC without feedback is given by

$$C = (1 - \varepsilon) \log q - (\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})),$$

where $\varepsilon = P_{Z_i}(e)$ is the probability of an erasure, $\bar{H}(\cdot)$ denotes the entropy rate and $\tilde{\boldsymbol{Z}} = \{\tilde{Z}_i\}_{i=1}^{\infty}$ is an auxiliary erasure process derived from the noise-erasure process $\boldsymbol{Z}$ as follows

$$\tilde{Z}_i = \begin{cases} 0 & \text{if } Z_i \neq e \\ e & \text{if } Z_i = e. \end{cases} \tag{8}$$

Before proving the theorem, we state the following observations.

Observation 1 (Important Special Cases):

- If $\{Z_i\}_{i=1}^{\infty}$ is memoryless, then

$$\begin{aligned} C &= (1 - \varepsilon) \log q - (\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})) \\ &= (1 - \varepsilon) \log q - H(Z_1 | \tilde{Z}_1). \end{aligned} \tag{9}$$

- If we set $q' = 1$ and $h(x, z) = x$, then $Z_i = \tilde{Z}_i$ and $C = (1 - \varepsilon) \log q$, recovering the capacity of the burst EC [11].

- If $q' = q$, $h(x, z) = x \oplus_q z$ and $\varepsilon = 0$, then $C = \log q - \bar{H}(\boldsymbol{Z})$ and we recover the capacity of the discrete symmetric channel in [24] which subsumes the ANC [21].

Observation 2 (Capacity calculation): The calculation of the NEC capacity given in Theorem 1 hinges on the evaluation of the entropy rates $\bar{H}(\boldsymbol{Z})$ and $\bar{H}(\tilde{\boldsymbol{Z}})$ of the noise-erasure and auxiliary erasure processes, respectively. As both processes are stationary, we have $\bar{H}(\boldsymbol{Z}) \leq H(Z_l | Z^{l-1})$

and $\bar{H}(\tilde{\mathbf{Z}}) \leq H(\tilde{Z}_l|\tilde{Z}^{l-1})$ for any fixed integer $l \geq 1$, and estimates of the entropy rates (whose accuracy improve with l) can be obtained. We next examine how to determine these entropy rates when the noise-erasure process $\mathbf{Z}$ is a first-order Markov source.

- *Special Markov noise-erasure process:* If the noise-erasure process $\mathbf{Z}$ is first-order Markov and satisfies

$$P_{Z_i|Z_{i-1}}(e|z_{i-1}) = \varepsilon'$$

  for some $0 \leq \varepsilon' \leq 1$ and all $z_{i-1} \in \mathcal{Q}'$, then the corresponding auxiliary erasure process $\tilde{\mathbf{Z}}$ is also Markov[3] with transition distribution given by

$$P_{\tilde{Z}_i|\tilde{Z}_{i-1}}(e|0) = P_{Z_i|Z_{i-1}}(e|0) = \varepsilon'$$

  and

$$P_{\tilde{Z}_i|\tilde{Z}_{i-1}}(e|e) = P_{Z_i|Z_{i-1}}(e|e).$$

  In this case, we directly have $\bar{H}(\mathbf{Z}) = H(Z_2|Z_1)$ and $\bar{H}(\tilde{\mathbf{Z}}) = H(\tilde{Z}_2|\tilde{Z}_1)$, hence simplifying the NEC capacity to the following expression

$$C = (1 - \varepsilon)\log q - (H(Z_2|Z_1) - H(\tilde{Z}_2|\tilde{Z}_1)),$$

  which can be exactly determined.

- *General Markov noise-erasure process:* For a general (stationary ergodic first-order) Markov noise-erasure process $\mathbf{Z}$, the auxiliary erasure process $\tilde{\mathbf{Z}}$ is not Markovian; it is a hidden Markov process as $\tilde{Z}_i$ is a deterministic function of Z_i given by $\tilde{Z}_i = e \cdot \mathbf{1}(Z_i = e)$, yielding

$$\Pr(\tilde{Z}_j = \tilde{z}_j|Z_j = z_j) = \mathbf{1}(\tilde{z}_j = e)\mathbf{1}(z_j = e) + \mathbf{1}(\tilde{z}_j = 0)\mathbf{1}(z_j \neq e) \qquad (10)$$

  for any $j \geq 1$. As noted above, $\bar{H}(\tilde{\mathbf{Z}})$ is upper bounded by $H(\tilde{Z}_l|\tilde{Z}^{l-1})$ for any positive l. Furthermore by [45, Section 4.5], it is lower bounded by $H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1)$, $l \geq 1$, whose initial values are given by

$$H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1) = \begin{cases} H(\tilde{Z}_1|Z_1) = 0 & \text{for } l = 1, \\ H(\tilde{Z}_2|Z_1) & \text{for } l = 2. \end{cases}$$

---

[3]This can be shown, along the same lines as equations (49) and (51) in Section VI, by noting that if the conditional term $Y^{i-1} = y^{i-1}$ is removed, both equations still hold.

Thus, when $\boldsymbol{Z}$ is a general Markov process, the NEC capacity satisfies

$$(1-\varepsilon)\log q - H(Z_2|Z_1) + H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1)$$

$$\leq C \leq (1-\varepsilon)\log q - H(Z_2|Z_1) + H(\tilde{Z}_l|\tilde{Z}^{l-1}) \tag{11}$$

for any $l \geq 1$, where the lower and upper bounds in (11) are asymptotically tight as $l \to \infty$ [45, Section 4.5]. In light of the channel structure, the conditional entropies $H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1)$ and $H(\tilde{Z}_l|\tilde{Z}^{l-1})$ can be efficiently computed via the joint distributions $\Pr(Z_1 = z_1, \tilde{Z}^i = \tilde{z}^i)$ and $\Pr(\tilde{Z}^i = \tilde{z}^i)$, $i \geq 1$, which we can determine recursively as follows:

(i) Initial distributions: we have

$$\Pr(\tilde{Z}_1 = e) = 1 - \Pr(\tilde{Z}_1 = 0) = P_{Z_1}(e)$$

and

$$\Pr(Z_1 = z_1, \tilde{Z}_1 = \tilde{z}_1) = P_{Z_1}(z_1)\Pr(\tilde{Z}_1 = \tilde{z}_1|Z_1 = z_1)$$

for $z_1 \in \mathcal{Z} = \mathcal{Q}' \cup \{e\}$, $\tilde{z}_1 \in \{0, e\}$, where $\Pr(\tilde{Z}_j = \tilde{z}_j|Z_j = z_j)$ is given in (10).

(ii). For any $i \geq 2$, $z_1, z_i \in \mathcal{Z}$ and $\tilde{z}^i \in \{0, e\}^i$, we have the recursion

$$\Pr(Z_1 = z_1, Z_i = z_i, \tilde{Z}^i = \tilde{z}^i)$$
$$= \sum_{z_2,\dots,z_{i-1}} \Pr(Z^i = z^i, \tilde{Z}^i = \tilde{z}^i)$$
$$= \sum_{z_2,\dots,z_{i-1}} \Pr(Z^{i-1} = z^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1})P_{Z_i|Z_{i-1}}(z_i|z_{i-1})\Pr(\tilde{Z}_i = \tilde{z}_i|Z_i = z_i)$$
$$= \sum_{z_{i-1}} \Pr(Z_1 = z_1, Z_{i-1} = z_{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1})P_{Z_i|Z_{i-1}}(z_i|z_{i-1})\Pr(\tilde{Z}_i = \tilde{z}_i|Z_i = z_i) \tag{12}$$

where for $i = 2$, $\Pr(Z_1 = z_1, Z_{i-1} = z_{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1}) = \Pr(Z_1 = z_1, \tilde{Z}_1 = \tilde{z}_1)$, which is given above. Finally, in light of (12), we have

$$\Pr(Z_1 = z_1, \tilde{Z}^i = \tilde{z}^i) = \sum_{z_i} \Pr(Z_1 = z_1, Z_i = z_i, \tilde{Z}^i = \tilde{z}^i)$$

and

$$\Pr(\tilde{Z}^i = \tilde{z}^i) = \sum_{z_1} \Pr(Z_1 = z_1, \tilde{Z}^i = \tilde{z}^i)$$

for $i \geq 2$.

Example: We illustrate the estimation of C by computing the above lower and upper bounds on the entropy rate $\bar{H}(\tilde{\boldsymbol{Z}})$ of the auxiliary erasure process. We consider two cases of

the stationary (first-order) Markov noise-erasure process $\{Z_i\}$. In the first case, $\{Z_i\}$ has alphabet $\{0, 1, e\}$ (i.e., $q' = 2$) and a transition probability matrix given by

$$\mathbf{\Pi} = \begin{bmatrix} P_{Z_i|Z_{i-1}}(0|0) & P_{Z_i|Z_{i-1}}(1|0) & P_{Z_i|Z_{i-1}}(e|0) \\ P_{Z_i|Z_{i-1}}(0|1) & P_{Z_i|Z_{i-1}}(1|1) & P_{Z_i|Z_{i-1}}(e|1) \\ P_{Z_i|Z_{i-1}}(0|e) & P_{Z_i|Z_{i-1}}(1|e) & P_{Z_i|Z_{i-1}}(e|e) \end{bmatrix} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.4 & 0.4 & 0.2 \end{bmatrix}.$$

In the second case, $\{Z_i\}$ has alphabet $\{0, 1, 2, 3, e\}$ ($q' = 4$) and transition probability

$$\hat{\mathbf{\Pi}} = \begin{bmatrix} 0.8 & 0 & 0 & 0.2 & 0 \\ 0 & 0.7 & 0 & 0 & 0.3 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0.4 & 0 & 0.6 & 0 \\ 0.1 & 0 & 0.1 & 0 & 0.8 \end{bmatrix}.$$

We plot in Figs. 1 and 2, the upper and lower bounds on the entropy rate $\bar{H}(\tilde{\mathbf{Z}})$ when $\{\tilde{Z}_i\}$ is generated by $\{Z_i\}$ with transition matrices $\mathbf{\Pi}$ and $\hat{\mathbf{\Pi}}$, respectively. The figures indicate a relatively fast agreement between the upper and lower bounds. For source $\mathbf{\Pi}$, the bounds converge for $l = 8$ with $\bar{H}(\tilde{\mathbf{Z}}) \approx 0.5884$. For source $\hat{\mathbf{\Pi}}$, the bounds take longer to converge (which is expected, as this source is significantly more bursty than $\mathbf{\Pi}$); we have $\bar{H}(\tilde{\mathbf{Z}}) \approx 0.529$ around $l = 17$. With the accurate calculation of $\bar{H}(\tilde{\mathbf{Z}})$ in each case, we obtain a reliable estimate for channel capacity: $C = (1 - \varepsilon) \log q - (H(Z_2|Z_1) - \bar{H}(\tilde{\mathbf{Z}}))$. We have numerically observed a similar behavior of the tightness of these bounds for other examples of Markov noise-erasure processes with various degrees of burstiness.

We close this observation by noting that the upper and lower bounds on C in (11) can be directly extended to NECs with M'th-order Markov noise-erasure processes:

$$(1 - \varepsilon) \log q - H(Z_{M+1}|Z_M, \ldots, Z_1) + H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_M, \ldots, Z_1)$$

$$\leq C \leq (1 - \varepsilon) \log q - H(Z_{M+1}|Z_M, \ldots, Z_1) + H(\tilde{Z}_l|\tilde{Z}^{l-1}) \qquad (13)$$

for $l \geq M$, with the lower and upper bounds in (13) asymptotically coinciding as l grows without bound.

Observation 3 (Effect of memory on NEC capacity): Here we examine the effect of memory on the capacity of the NEC with stationary noise-erasure process $\mathbf{Z} = \{Z_i\}_{i=1}^{\infty}$. Let $\mathbf{Z}' = \{Z_i'\}_{i=1}^{\infty}$ be a memoryless noise-erasure process with the same marginal distribution as $\mathbf{Z}$ and let C^{DMC} denote the capacity of the NEC with noise-erasure process $\mathbf{Z}'$ (which is the memoryless
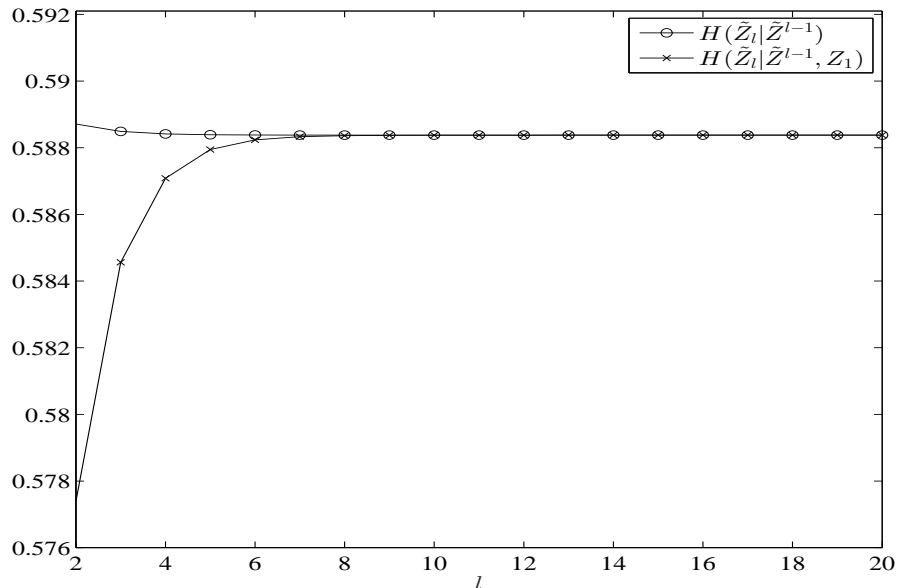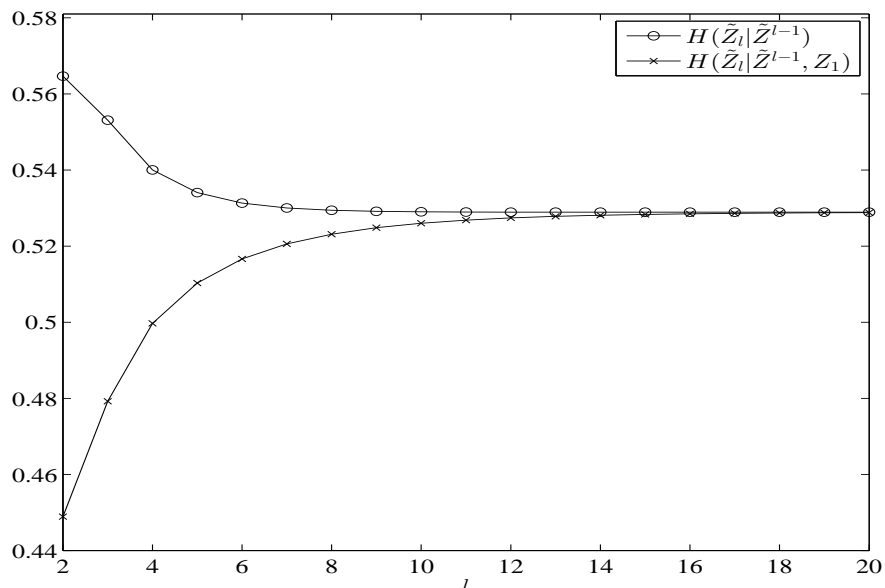
Fig. 1. Comparison as a function of l of the entropy rate upper and lower bounds, $H(\tilde{Z}_l|\tilde{Z}^{l-1})$ and $H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1)$, for the hidden Markov source $\{\tilde{Z}_i\}$ generated by the Markov noise-erasure process given by $\mathbf{\Pi}$.

Fig. 2. Comparison as a function of l of the entropy rate upper and lower bounds, $H(\tilde{Z}_l|\tilde{Z}^{l-1})$ and $H(\tilde{Z}_l|\tilde{Z}^{l-1}, Z_1)$, for the hidden Markov source $\{\tilde{Z}_i\}$ generated by the Markov noise-erasure process given by $\hat{\mathbf{\Pi}}$.

counterpart channel to the NEC). Similarly, let $\tilde{\boldsymbol{Z}}'$ be the memoryless erasure process obtained from $\boldsymbol{Z}'$. Since the channel is stationary and information stable, we readily obtain from (4) that $C \geq C_1 = C^{DMC}$; see also [47]. We have

$$C^{DMC} = (1 - \varepsilon) \log q - \left[\bar{H}(\boldsymbol{Z}') - \bar{H}(\tilde{\boldsymbol{Z}}') \right]$$

$$= (1 - \varepsilon) \log q - H(Z'_1 | \tilde{Z}'_1)$$

$$= (1 - \varepsilon) \log q - H(Z_1 | \tilde{Z}_1).$$

Therefore, $C > C^{DMC}$ if and only if $\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}}) < H(Z_1 | \tilde{Z}_1)$. If $q' = 1$ and $h(x, z) = z$, then $\boldsymbol{Z} = \tilde{\boldsymbol{Z}}$ and the NEC reduces to the EC; in this case, $C = C^{DMC} = (1 - \varepsilon) \log q$, which is the well-known result that memory does not increase capacity of the burst EC [5], [11]. If $\varepsilon = 0$ (i.e., no erasures occur) and $\boldsymbol{Z}$ has memory, then $C = \log q - H(\boldsymbol{Z}) > \log q - H(Z_1) = C^{DMC}$. For the NEC with general noise-erasure process (noisy, with $\varepsilon \neq 0$ and not memoryless), it is not obvious whether $C > C^{DMC}$ since we need to evaluate the difference of the entropy rates of two random process with memory. We provide an answer in the case of Markov noise-erasure processes in the following lemma.

Lemma 5: For NECs with (first-order) Markov noise-erasure process $\boldsymbol{Z}$, if there exist $z_1, z'_1, z_2 \in \mathcal{Q}'$ with $z_1 \neq z'_1$ such that $P_{Z_1}(z_1) > 0$, $P_{Z_1}(z'_1) > 0$ and $P_{Z_2|Z_1}(z_2|z_1) \neq P_{Z_2|Z_1}(z_2|z'_1)$, then $C > C^{DMC}$.

Note the conditions in Lemma 5 readily hold for non-degenerate (i.e., non-memoryless) stationary, irreducible and aperiodic Markov noise-erasure processes.

Proof: To prove this lemma, we first need the following lemma whose proof in given in the appendix.

Lemma 6: Let $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ and $\tilde{\boldsymbol{Z}} = \{\tilde{Z}_i\}_{i=1}^{\infty}$ be the processes as in (8) and let $H_n \triangleq \frac{1}{n}[H(Z^n) - H(\tilde{Z}^n)]$. Then the sequence $\{nH_n\}_{n=1}^{\infty}$ is subadditive.

From Lemma 6, we have

$$\inf_n H_n = \lim_{n \to \infty} H_n = \lim_{n \to \infty} \frac{1}{n}[H(Z^n) - H(\tilde{Z}^n)] = \bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}}).$$

Thus, $\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}}) \leq H_2$. Note that

$$
\begin{aligned}
H_2 &= \frac{1}{2}\big[H(Z^2) - H(\tilde{Z}^2)\big] \\
&= \frac{1}{2}\big[H(Z_1) + H(Z_2|Z_1) - H(\tilde{Z}_1) - H(\tilde{Z}_2|\tilde{Z}_1)\big] \\
&= H(Z_1) - H(\tilde{Z}_1) + \frac{1}{2}\big[-H(Z_1) + H(Z_2|Z_1) + H(\tilde{Z}_1) - H(\tilde{Z}_2|\tilde{Z}_1)\big] \\
&= H(Z_1|\tilde{Z}_1) + \frac{1}{2}\big[-H(Z_2) + H(Z_2|Z_1) + H(\tilde{Z}_2) - H(\tilde{Z}_2|\tilde{Z}_1)\big] \\
&= H(Z_1|\tilde{Z}_1) + \frac{1}{2}\big[-I(Z_1; Z_2) + I(\tilde{Z}_1; \tilde{Z}_2)\big] \\
&\leq H(Z_1|\tilde{Z}_1),
\end{aligned}
\tag{14}
$$

where (14) holds since $\tilde{Z}_1 - Z_1 - Z_2 - \tilde{Z}_2$ form a Markov chain and where equality holds if and only if $Z_1 - \tilde{Z}_1 - \tilde{Z}_2 - Z_2$ also form a Markov chain. Therefore, for a (first-order) Markov noise-erasure process $\boldsymbol{Z}$, if there exist $z_1, z_1', z_2 \in \mathcal{Q}'$ and $z_1 \neq z_1'$ such that $P_{Z_1}(z_1) > 0$, $P_{Z_1}(z_1') > 0$ and $P_{Z_2|Z_1}(z_2|z_1) \neq P_{Z_2|Z_1}(z_2|z_1')$, then $H(\boldsymbol{Z}) - H(\tilde{\boldsymbol{Z}}) \leq H_2 < H(Z_1|\tilde{Z}_1)$, which implies that $C > C^{DMC}$. $\blacksquare$

We conclude this section with the proof of Theorem 1.

Proof of Theorem 1: An NEC with stationary and ergodic noise-erasure process $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$ is stationary and information stable. Therefore, its non-feedback capacity is given by (4):

$$C = \lim_{n \to \infty} C_n = \lim_{n \to \infty} \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^n).$$

Focusing on C_n, note that it can be viewed as the capacity of a discrete memoryless channel with input alphabet $\mathcal{X}^n$, output alphabet $\mathcal{Y}^n$, and input-output relationship $Y_i = h(X_i, Z_i) \cdot \mathbf{1}\{Z_i \neq e\} + e \cdot \mathbf{1}\{Z_i = e\}$, for $i = 1, 2, ..., n$. Let $W^n(\cdot|\cdot)$ and $\mathbb{Q}^{(n)}$ denote the transition probability and transition matrix of this channel, respectively, and let $\bar{q}_{y^n|X^n}$ denote the column of $\mathbb{Q}^{(n)}$ associated with the output y^n, i.e.,

$$\bar{q}_{y^n|X^n} \triangleq [W^n(y^n|x^n)]_{x^n \in \mathcal{X}^n}^T,$$

where the superscript "T" denotes transposition and the entries of $\bar{q}_{y^n|X^n}$ are listed in the lexicographic order. For example, for binary input alphabet and $n = 2$,

$$\bar{q}_{y^2|X^2} = [W^2(y^2|00), W^2(y^2|01), W^2(y^2|10), W^2(y^2|11)]^T.$$

For any $\mathcal{S} \subseteq \mathcal{N} \triangleq \{1, 2, ..., n\}$, define

$$\mathcal{Y}_{\mathcal{S}} \triangleq \{y^n : y_i = e \text{ for } i \in \mathcal{S}, y_i \neq e \text{ for } i \notin \mathcal{S}\},$$

and

$$\mathbb{Q}_{\mathcal{Y}_{\mathcal{S}}|X^n} \triangleq [\bar{q}_{y^n|X^n}]_{y^n \in \mathcal{Y}_{\mathcal{S}}},$$

where the columns of $\mathbb{Q}_{\mathcal{Y}_{\mathcal{S}}|X^n}$ are collected in the lexicographic order in $y^n \in \mathcal{Y}_{\mathcal{S}}$. We first show that the n-fold channel $\mathbb{Q}^{(n)}$ of the NEC is quasi-symmetric (recall Definition 6).[4] Note that $\{\mathbb{Q}_{\mathcal{Y}_{\mathcal{S}}|X^n}\}_{\mathcal{S} \subseteq \mathcal{N}}$ is a partition of $\mathbb{Q}^{(n)}$. Also in light of properties S-I and S-II, we have the following two lemmas (Lemma 7 and 8) which imply the quasi-symmetry of the NEC.

---

[4]The NEC, being quasi-symmetric, satisfies a weaker (and hence more general) notion of "symmetry" than the ANC [21] and the channel in [24] which are both symmetric.

Lemma 7: For any $\mathcal{S} \subseteq \mathcal{N}$, each row of $\mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}$ is a permutation of

$$\bar{p}_{\mathcal{Z}_\mathcal{S}} \triangleq [P_{Z^n}(z^n), \underbrace{0, ..., 0}_{q^{n-|\mathcal{S}|}-(q')^{n-|\mathcal{S}|}}]_{z^n \in \mathcal{Z}_\mathcal{S}},$$

where

$$\mathcal{Z}_\mathcal{S} \triangleq \{z^n : z_i = e \text{ for } i \in \mathcal{S}, z_i \neq e \text{ for } i \notin \mathcal{S}\},$$

and the entries of $\bar{p}_{\mathcal{Z}_\mathcal{S}}$ are collected in the lexicographic order in $z^n \in \mathcal{Z}_\mathcal{S}$.

Proof: Fixing an input vector $x^n \in \mathcal{X}^n$ and considering the element in $\mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}$ associated with the input sequence x^n and output sequence $y^n \in \mathcal{Y}_\mathcal{S}$, we have

$$W(y^n|x^n)$$
$$= \begin{cases} \Pr(Z_{\mathcal{N}\backslash\mathcal{S}} = \tilde{h}(x_{\mathcal{N}\backslash\mathcal{S}}, y_{\mathcal{N}\backslash\mathcal{S}}), Z_\mathcal{S} = e^{|\mathcal{S}|}), & \text{if } y_i \in \mathcal{Y}_{x_i} \text{ for all } i \in \mathcal{N} \backslash \mathcal{S}, \\ 0, & \text{if } \exists \ y_i \notin \mathcal{Y}_{x_i} \text{ for some } i \in \mathcal{N} \backslash \mathcal{S}, \end{cases} \quad (15)$$

where $Z_{\mathcal{N}\backslash\mathcal{S}}$ denotes $\{Z_i : i \in \mathcal{N} \backslash \mathcal{S}\}$ and similarly for $x_{\mathcal{N}\backslash\mathcal{S}}$ and $y_{\mathcal{N}\backslash\mathcal{S}}$, $\tilde{h}(x_\mathcal{A}, y_\mathcal{A})$ is short for $(\tilde{h}(x_i, y_i))_{i \in \mathcal{A}}$, and (15) follows from (7) and (1). Note that

$$\left\{ \tilde{h}(x_{\mathcal{N}\backslash\mathcal{S}}, y_{\mathcal{N}\backslash\mathcal{S}}) \right\}_{y_{\mathcal{N}\backslash\mathcal{S}} \in \Pi_{i \in \mathcal{N}\backslash\mathcal{S}} \mathcal{Y}_{x_i}} = \left(\mathcal{Q}' \right)^{n-|\mathcal{S}|},$$

since $\tilde{h}(x, \cdot)$ is invertible for all $x \in \mathcal{Q}$ (property S-I). Therefore,

$$\left\{ z^n : z_{\mathcal{N}\backslash\mathcal{S}} = \tilde{h}(x_{\mathcal{N}\backslash\mathcal{S}}, y_{\mathcal{N}\backslash\mathcal{S}}), z_\mathcal{S} = e^{|\mathcal{S}|}, \text{ for some } y_{\mathcal{N}\backslash\mathcal{S}} \in \Pi_{i \in \mathcal{N}\backslash\mathcal{S}} \mathcal{Y}_{x_i} \right\}$$
$$= \left\{ z^n : z_{\mathcal{N}\backslash\mathcal{S}} \in \left(\mathcal{Q}' \right)^{n-|\mathcal{S}|}, z_\mathcal{S} = e^{|\mathcal{S}|} \right\}$$
$$= \mathcal{Z}_\mathcal{S}. \quad (16)$$

From (15), we note that $W(y^n|x^n) = 0$ for all y^n such that $y^n \in \mathcal{Y}_\mathcal{S}$ with $y_{\mathcal{N}\backslash\mathcal{S}} \notin \Pi_{i \in \mathcal{N}\backslash\mathcal{S}} \mathcal{Y}_{x_i}$. Therefore, $[W(y^n|x^n)]_{y^n \in \mathcal{Y}_\mathcal{S}}$ is a permutation of $[P_{Z^n}(z^n), 0, ..., 0]_{z^n \in \mathcal{Z}_\mathcal{S}}$ where the number of consecutive zeros is $|\mathcal{Y}_\mathcal{S}| - |\Pi_{i \in \mathcal{N}\backslash\mathcal{S}} \mathcal{Y}_{x_i}| = |\mathcal{Y}_\mathcal{S}| - |\mathcal{Z}_\mathcal{S}| = q^{n-|\mathcal{S}|} - (q')^{n-|\mathcal{S}|}$. Note that the above argument does not depend on the input sequence x^n, and thus all rows of $\mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}$ are permutations of $\bar{p}_{\mathcal{Z}_\mathcal{S}}$. ∎

Lemma 8: For any $\mathcal{S} \subseteq \mathcal{N}$, the column sums of $\mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}$ are identical and are equal to

$$q^{|\mathcal{S}|} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S})),$$

where $\tilde{Z}_i$, $i = 1, \cdots, n$, is defined in (8), and $\tilde{z}(n, \mathcal{S})$ denotes the n-tuple with components

$$\tilde{z}_i(n, \mathcal{S}) = \begin{cases} 0 & \text{for } i \in \mathcal{N} \backslash \mathcal{S}, \\ e & \text{for } i \in \mathcal{S}. \end{cases} \quad (17)$$

Proof: Fixing an output sequence $y^n \in \mathcal{Y}_\mathcal{S}$ and considering the column sum of $\bar{q}_{y^n|X^n}$, we have

$$\sum_{x^n \in \mathcal{Q}^n} W^n(y^n|x^n)$$

$$= \sum_{x^n \in \mathcal{Q}^n : x_i \in \mathcal{X}_{y_i}, i \in \mathcal{N} \setminus \mathcal{S}} \Pr(Z_{\mathcal{N} \setminus \mathcal{S}} = \tilde{h}(x_{\mathcal{N} \setminus \mathcal{S}}, y_{\mathcal{N} \setminus \mathcal{S}}), Z_\mathcal{S} = e^{|\mathcal{S}|}) \tag{18}$$

$$= \sum_{x_\mathcal{S} \in \mathcal{Q}^{|\mathcal{S}|}} \sum_{x_{\mathcal{N} \setminus \mathcal{S}} \in \mathcal{Q}^{n-|\mathcal{S}|} : x_i \in \mathcal{X}_{y_i}, i \in \mathcal{N} \setminus \mathcal{S}} \Pr(Z_{\mathcal{N} \setminus \mathcal{S}} = \tilde{h}(x_{\mathcal{N} \setminus \mathcal{S}}, y_{\mathcal{N} \setminus \mathcal{S}}), Z_\mathcal{S} = e^{|\mathcal{S}|})$$

$$= q^{|\mathcal{S}|} \sum_{x_{\mathcal{N} \setminus \mathcal{S}} \in \mathcal{Q}^{n-|\mathcal{S}|} : x_i \in \mathcal{X}_{y_i}, i \in \mathcal{N} \setminus \mathcal{S}} \Pr(Z_{\mathcal{N} \setminus \mathcal{S}} = \tilde{h}(x_{\mathcal{N} \setminus \mathcal{S}}, y_{\mathcal{N} \setminus \mathcal{S}}), Z_\mathcal{S} = e^{|\mathcal{S}|})$$

$$= q^{|\mathcal{S}|} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \tag{19}$$

$$= q^{|\mathcal{S}|} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S})), \tag{20}$$

where (18) follows (7), $y^n \in \mathcal{Y}_\mathcal{S}$ and $W^n(y^n|x^n) = 0$ if $x_i \notin \mathcal{X}_{y_i}$ for some $i \in \mathcal{N} \setminus \mathcal{S}$, (19) follows from property S-II, and (20) follows from (8) and (17). Note that (20) does not depend on the output sequence y^n, and thus the column sums are identical. ∎

We are now ready to explicitly determine C_n. By Lemma 4, we have

$$C_n = \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \cdot \left[\log q^{n-|\mathcal{S}|} - H\left(\text{any row of } \frac{1}{\sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n)} \mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n} \right) \right]$$

$$= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \left[\log q^{n-|\mathcal{S}|} - H\left(\left(\frac{P_{Z^n}(\hat{z}^n)}{\sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n)} \right)_{\hat{z}^n \in \mathcal{Z}_\mathcal{S}} \right) \right]$$

$$= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \left[\log q^{n-|\mathcal{S}|} - H(Z^n|Z^n \in \mathcal{Z}_\mathcal{S}) \right]$$

$$= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \Pr(\tilde{Z}_{\mathcal{N}/\mathcal{S}} = 0^{n-|\mathcal{S}|}, \tilde{Z}_\mathcal{S} = e^{|\mathcal{S}|}) \left[\log q^{n-|\mathcal{S}|} - H(Z^n|\tilde{Z}_{\mathcal{N}/\mathcal{S}} = 0^{n-|\mathcal{S}|}, \tilde{Z}_\mathcal{S} = e^{|\mathcal{S}|}) \right] \tag{21}$$

$$= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S})) \left[\log q^{n-|\mathcal{S}|} - H(Z^n|\tilde{Z}^n = \tilde{z}(n, \mathcal{S})) \right]$$

$$= \frac{1}{n} \Big[n \log q \sum_{\mathcal{S} \subseteq \mathcal{N}} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S})) - \log q \sum_{\mathcal{S} \subseteq \mathcal{N}} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S}))|\mathcal{S}|$$

$$- \sum_{\mathcal{S} \subseteq \mathcal{N}} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S})) H(Z^n|\tilde{Z}^n = \tilde{z}(n, \mathcal{S})) \Big]$$

$$= \frac{1}{n} \Big[n \log q - \log q \sum_{\tilde{z}^n \in \tilde{\mathcal{Z}}^n} P_{\tilde{Z}^n}(\tilde{z}^n) \sum_{i=1}^{n} \mathbf{1}(\tilde{z}_i = e) - \sum_{\tilde{z}^n \in \tilde{\mathcal{Z}}^n} P_{\tilde{Z}^n}(\tilde{z}^n) H(Z^n|\tilde{Z}^n = \tilde{z}^n) \Big]$$

$$= \frac{1}{n}\Big[n\log q - \log q \cdot E\Big[\sum_{i=1}^{n} \mathbf{1}(\tilde{Z}_i = e)\Big] - H(Z^n|\tilde{Z}^n)\Big]$$

$$= \log q - \frac{1}{n}\log q \sum_{i=1}^{n} E\Big[\mathbf{1}(\tilde{Z}_i = e)\Big] - \frac{1}{n}H(Z^n|\tilde{Z}^n)$$

$$= (1-\varepsilon)\log q - \frac{1}{n}(H(Z^n) - H(\tilde{Z}^n)),$$

where (21) follows from (8). Taking the limit of C_n above and using the definition of entropy rate (which exists for both $\mathbf{Z}$ and $\tilde{\mathbf{Z}}$ by stationarity) yields

$$C = \lim_{n\to\infty} C_n = (1-\varepsilon)\log q - (\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}})).$$

∎

IV. NEC FEEDBACK CAPACITY

We next show that feedback does not increase the capacity of the NEC.

Theorem 2: Feedback does not increase the capacity of the NEC:

$$C_{FB} = C = (1-\epsilon)\log q - [\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}})],$$

where $\tilde{\mathbf{Z}} = \{\tilde{Z}_i\}_{i=1}^{\infty}$ is defined in (8).

Proof: For any sequence of feedback channel codes with rate R and error probability satisfying $\lim_{n\to 0} P_e^{(n)} = 0$, we have

$$nR = H(M)$$

$$= I(M;Y^n) + H(M|Y^n)$$

$$\leq I(M;Y^n) + n\epsilon_n \tag{22}$$

$$= \sum_{i=1}^{n} I(M;Y_i|Y^{i-1}) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, M) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, M, X^i) + n\epsilon_n \tag{23}$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, M, X^i, Z^{i-1}) + n\epsilon_n \tag{24}$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Z_i|Y^{i-1}, M, X^i, Z^{i-1}) + n\epsilon_n \tag{25}$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \sum_{i=1}^{n} H(Z_i|Z^{i-1}) + n\epsilon_n \tag{26}$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Z^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} H(Y_i|Y^{i-1}, \tilde{Z}^{i-1}) - H(Z^n) + n\epsilon_n \tag{27}$$

$$\leq \sum_{i=1}^{n} H(Y_i|\tilde{Z}^{i-1}) - H(Z^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} \sum_{\tilde{z}^{i-1}} \Pr(\tilde{Z}^{i-1} = \tilde{z}^{i-1}) H(Y_i|\tilde{Z}^{i-1} = \tilde{z}^{i-1}) - H(Z^n) + n\epsilon_n$$

$$\leq \sum_{i=1}^{n} \sum_{\tilde{z}^{i-1}} \Pr(\tilde{Z}^{i-1} = \tilde{z}^{i-1}) \max_{P_{X_i|\tilde{Z}^{i-1}}(\cdot|\tilde{z}^{i-1})} H(Y_i|\tilde{Z}^{i-1} = \tilde{z}^{i-1}) - H(Z^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} \sum_{\tilde{z}^{i-1}} P_{\tilde{Z}^{i-1}}(\tilde{z}^{i-1}) \Big[\big(1 - P_{Z_i|\tilde{Z}^{i-1}}(e|\tilde{z}^{i-1})\big) \log q + h_b\big(P_{Z_i|\tilde{Z}^{i-1}}(e|\tilde{z}^{i-1})\big) \Big] - H(Z^n) + n\epsilon_n \tag{28}$$

$$= \sum_{i=1}^{n} \sum_{\tilde{z}^{i-1}} P_{\tilde{Z}^{i-1}}(\tilde{z}^{i-1}) \Big[\big(1 - P_{Z_i|\tilde{Z}^{i-1}}(e|\tilde{z}^{i-1})\big) \log q + H(\tilde{Z}_i|\tilde{Z}^{i-1} = \tilde{z}^{i-1}) \Big] - H(Z^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} \Big[(1-\varepsilon) \log q + H(\tilde{Z}_i|\tilde{Z}^{i-1}) \Big] - H(Z^n) + n\epsilon_n$$

$$= n(1-\varepsilon) \log q + H(\tilde{Z}^n) - H(Z^n) + n\epsilon_n,$$

where ϵ_n goes to zero as $n \to \infty$. Here (22) follows from Fano's inequality, (23) holds since $X_i = f_i(M, Y^{i-1})$, $i = 1, 2, ..., n$, (24) follows from (7), (25) follows from (1) and (7), and (26) holds because Z^n and M are independent, and for $i \geq 2$,

$$H(Z_i|Z^{i-1}) = H(Z_i|Z^{i-1}, M)$$

$$= H(Z_i|Z^{i-1}, M, X_1) \tag{29}$$

$$= H(Z_i|Z^{i-1}, M, X_1, Y_1) \tag{30}$$

$$= H(Z_i|Z^{i-1}, M, X^2, Y_1) \tag{31}$$

$$= H(Z_i|Z^{i-1}, M, X^i, Y^{i-1}), \tag{32}$$

where (29) follows from $X_1 = f_1(M)$, (30) follows from (1), (31) holds since $X_2 = f_2(M, Y_1)$, and (32) is obtained by including more conditional terms as in (30) and (31). Furthermore,

equation (27) follows from (8) and (7). Finally, (28) follows from Corollary 1 in the appendix, and $h_b(\varepsilon) \triangleq -\varepsilon \log \varepsilon - (1 - \varepsilon) \log(1 - \varepsilon)$ is the binary entropy function. We thus have

$$
\begin{aligned}
C_{FB} &\leq (1 - \varepsilon) \log q - \lim_{n \to \infty} \frac{1}{n} \big[H(Z^n) - H(\tilde{Z}^n) \big] \\
&= (1 - \varepsilon) \log q - \big[\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}}) \big] \\
&= C.
\end{aligned}
$$

This inequality and the fact that $C_{FB} \geq C$ complete the proof. ∎

V. NEC CAPACITY-COST FUNCTION

In this section, we consider the capacity-cost function of NECs without feedback. The capacity-cost function given in (5) is a multi-letter expression and is not computable for general channels. We herein derive a set of finite-letter upper bounds for it.

Theorem 3: The capacity-cost function of the NEC satisfies

$$
C(\beta) \leq C_l(\beta) - \bar{H}(\mathbf{Z}) + \frac{1}{l} H(Z^l) \triangleq C_l^{ub}(\beta)
$$

for any positive integer l.

Proof: Consider a sequence of β-admissible channel codes with rate R such that $\lim_{n \to \infty} P_e^{(n)} = 0$. As in (22), it follows from Fano's inequality that

$$
R \leq \lim_{n \to \infty} \frac{1}{n} I(M; Y^n).
$$

For any fixed integer $l \geq 1$, let $n \triangleq kl + l'$ for some non-negative integers k and l', where $l' \in [0, l-1]$. Then we have

$$
\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} I(M; Y^n) &= \lim_{k \to \infty} \frac{1}{kl + l'} \big[I(M; Y^{kl}) + I(M; Y_{kl+1}^{kl+l'} | Y^{kl}) \big] \\
&\leq \lim_{k \to \infty} \frac{1}{kl + l'} \big[I(M; Y^{kl}) + l' \log |\mathcal{Y}| \big] \\
&= \lim_{k \to \infty} \frac{1}{kl + l'} I(M; Y^{kl}) \\
&\leq \lim_{k \to \infty} \frac{1}{kl} I(M; Y^{kl}),
\end{aligned}
\tag{33}
$$

where Y_i^j is a constant random variable, if $j < i$. Note that

$$
\begin{aligned}
I(M; Y^{kl}) &= I(M, X^{kl}; Y^{kl}) \\
&= H(Y^{kl}) - H(Y^{kl} | M, X^{kl})
\end{aligned}
$$

$$= H(Y^{kl}) - H(Z^{kl}|M, X^{kl})$$

$$= H(Y^{kl}) - H(Z^{kl})$$

$$\leq \sum_{i=1}^{k} H(Y_{(i-1)l+1}^{il}) - H(Z^{kl})$$

$$= \sum_{i=1}^{k} H(Y_{(i-1)l+1}^{il}) - \sum_{i=1}^{k} H(Y_{(i-1)l+1}^{il}|X_{(i-1)l+1}^{il}) + \sum_{i=1}^{k} H(Z_{(i-1)l+1}^{il}) - H(Z^{kl}) \quad (34)$$

$$= \sum_{i=1}^{k} I(X_{(i-1)l+1}^{il}; Y_{(i-1)l+1}^{il}) - H(Z^{kl}) + kH(Z^{l})$$

$$\leq \sum_{i=1}^{k} lC_l(\beta_i) - H(Z^{kl}) + kH(Z^{l})$$

$$\leq klC_l\Big(\frac{\sum_{i=1}^{k} \beta_i}{k}\Big) - H(Z^{kl}) + kH(Z^{l}) \quad (35)$$

$$\leq klC_l(\beta) - H(Z^{kl}) + kH(Z^{l}) \quad (36)$$

where $\beta_i \triangleq \frac{1}{l}E[b(X_{(i-1)l+1}^{il})]$, (34) follows from $H(Y_{(i-1)l+1}^{il}|X_{(i-1)l+1}^{il}) = H(Z_{(i-1)l+1}^{il}|X_{(i-1)l+1}^{il})$ and the independence of $Z_{(i-1)l+1}^{il}$ and $X_{(i-1)l+1}^{il}$, (35) follows from the concavity of $C_l(\beta)$ which is stated in Lemma 1, and (36) holds since $C_l(\beta)$ is monotone increasing by Lemma 1. Substituting (36) into (33), we have

$$\lim_{n\to\infty} \frac{1}{n} I(M; Y^n) \leq C_l(\beta) - \lim_{k\to\infty} \frac{1}{kl} H(Z^{kl}) + \frac{1}{l} H(Z^{l}).$$

Since

$$\lim_{n\to\infty} \frac{1}{n} H(Z^n) = \lim_{k\to\infty} \frac{1}{kl} H(Z^{kl}),$$

we obtain

$$C(\beta) \leq C_l(\beta) - \bar{H}(\boldsymbol{Z}) + \frac{1}{l} H(Z^{l}).$$

∎

The upper bounds for $C(\beta)$ given in Theorem 3, which hold for an arbitrary NEC (not necessarily satisfying conditions S-I and S-II), generalize the upper bounds for the capacity-cost function of the ANC shown in [26]. Note that these upper bounds are counterparts to the Wyner-Ziv lower bounds on the rate-distortion function of stationary sources [48], [49] and illustrate the functional duality between the capacity-cost and rate-distortion functions originally pointed out by Shannon [50]. For any positive integer l, $C_l(\beta)$ is a finite-letter lower bound to $C(\beta) : C(\beta) =$

$\sup_{n\geq 1} C_n(\beta) \geq C_l(\beta)$. The l-letter upper and lower bounds are asymptotically tight as the gap $\Delta_l \triangleq C_l^{ub}(\beta) - C_l(\beta) = \frac{1}{l}H(Z^l) - \bar{H}(\mathbf{Z})$ goes to zero as $l \to \infty$. Finally, note that for finite l, both $C_l(\beta)$ and $C_l^{ub}(\beta)$ can be numerically evaluated via Blahut's algorithm for the capacity-cost function [40], [51].

VI. NEC CAPACITY-COST FUNCTION WITH FEEDBACK

We next investigate the feedback capacity-cost function $C_{FB}(\beta)$ of the NEC. At time i, the transmitter obtains Y^{i-1} from the feedback link, and thus knows Z^{i-1} according to (7). Therefore, the input at time i can be generated according to $X_i = f_i(M, Z^{i-1})$. In general, the feedback encoding rule $f_i(M, Z^{i-1})$ is time-varying. In this section, we will choose an input cost function, a family of Markov noise-erasure processes and an appropriate time invariant feedback encoding rule to demonstrate that feedback can increase the capacity-cost function.

We first derive a lower bound to $C_{FB}(\beta)$ under time invariant feedback strategies. For the NEC with feedback and a fixed encoding rule $f^* : \mathcal{Q} \times (\mathcal{Q}' \cup \{e\}) \to \mathcal{Q}$, we define $C_n^{lb,FB}(\beta)$ as

$$C^{lb,FB}(\beta) = \sup_n C_n^{lb,FB}(\beta) = \lim_{n\to\infty} C_n^{lb,FB}(\beta),$$

where

$$C_n^{lb,FB}(\beta) = \max_{P_{V^n} \in \tilde{\tau}_n(\beta)} \frac{1}{n} I(V^n; Y^n),$$

$X_i = f^*(V_i, Z_{i-1})$, for $i = 1, 2, ..., n$, V^n is a q-ary n-tuple independent of Z^n, and

$$\tilde{\tau}_n(\beta) \triangleq \{P_{V^n} : \frac{1}{n} E[b(X^n)] \leq \beta\}.$$

Note that the cost constraint is imposed on the input vector X^n rather than V^n. We next state without proving the following theorem; the proof can be obtained by using a standard random coding argument as in the proof of [26, Theorem 2].

Theorem 4 (Achievability of $C^{lb,FB}(\beta) : C_{FB}(\beta) \geq C^{lb,FB}(\beta)$): Consider the NEC and a fixed time-invariant feedback encoding rule f^* as above. For any $R < C^{lb,FB}(\beta)$, there exists a sequence of β-admissible feedback codes of block length n and rate R such that $P_e^{(n)} \to 0$ as $n \to \infty$.

In the rest of this section, we consider the linear cost function $b(x) = x$ for $x \in \mathcal{Q}$ and the following specific encoding function f^* [26]. Let $V^n(M)$ be a q-ary n-tuple representing the message $M \in \{1, 2, ..., 2^{nR}\}$. Then, to transmit M, the encoder sends $X^n(M)$, where

$$X_1(M) = V_1(M); \quad X_i(M) = f^*(V_i(M), Z_{i-1}) \triangleq \begin{cases} V_i(M), & Z_{i-1} \neq \tilde{s} \\ 0, & Z_{i-1} = \tilde{s} \end{cases} \text{ if } i > 1, \quad (37)$$

and $\tilde{s}$ is some fixed preselected state. Note that $V^n(M)$ can be viewed as the input vector when there is no feedback; that is, if the channel is without feedback, then $X^n(M) = V^n(M)$. The encoder of the NEC with feedback can obtain the state Z_{i-1} at time i. If the encoder observes the "bad" state $\tilde{s}$, then it sends the least expensive symbol. In many cases (such as the examples considered in the figures below), the least expensive symbol has cost $b(0) = 0$.

In light of Theorems 3 and 4, a numerical comparison of $C_n^{lb,FB}(\beta)$ and $C_n^{ub}(\beta)$ for a given block length n can indicate whether it is possible for feedback to increase the capacity-cost function. Since $C(\beta) \leq C_n^{ub}(\beta)$ and $C^{lb,FB}(\beta) = \sup_n C_n^{lb,FB}(\beta)$, it suffices to show that $C_n^{lb,FB}(\beta) > C_n^{ub}(\beta)$ for some n and β to conclude that $C_{FB}(\beta) > C(\beta)$. To this end, consider an NEC with $q' = q = 2$, $h(x, z) = x \oplus_2 z$, a linear cost function and a first-order Markov noise-erasure process described by the transition matrix

$$\mathbf{\Pi}_1 = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.7 & 0.1 & 0.2 \\ 0.2 & 0.7 & 0.1 \end{bmatrix},$$

where the entries are ordered according to the order $(0, 1, e)$. In Fig. 3, we plot using Blahut's algorithm [40], [51] $C_n^{ub}(\beta)$ versus $C_n^{lb,FB}(\beta)$ (with f^* given by (37)) for $n = 6$. Fig. 3 clearly indicates that feedback increases the capacity-cost function of this NEC for a range of costs β.

We next formalize this result analytically for NECs with irreducible and aperiodic stationary Markov noise-erasure processes whose transition probability matrix has the property that the row corresponding to a given noise state $\tilde{s} \in \mathcal{Q}'$ and the column corresponding to the erasure state are nearly uniform. More specifically, we prove that for such channels using feedback encoding rule (37) (which is properly matched to the linear cost function), we can achieve the channel capacity with a cost that is lower than the cost incurred in the non-feedback case, hence extending a previous result in [26] from the family of ANCs to the wider class of NECs.
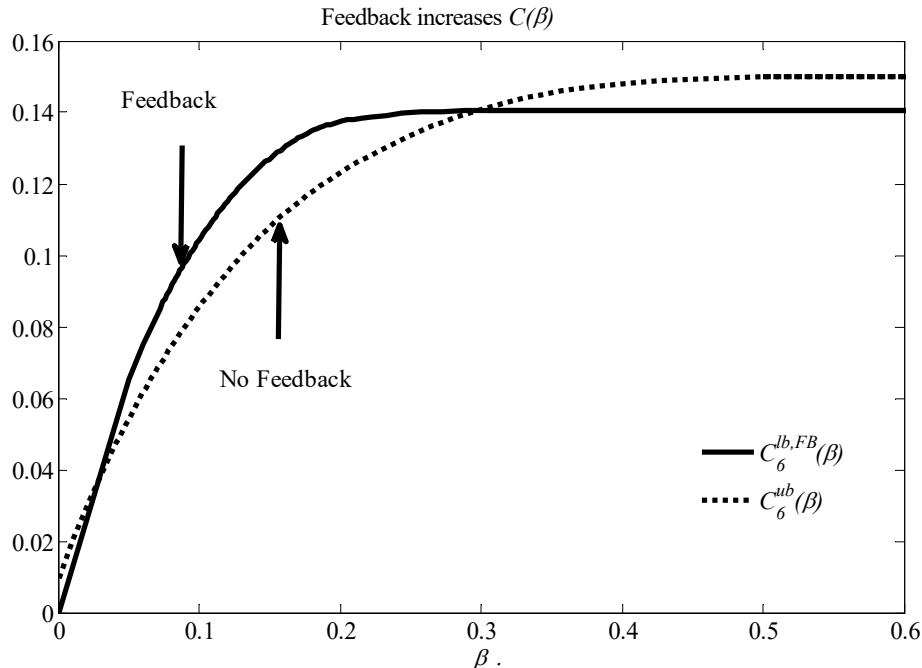
Fig. 3. Comparison of $C_6^{ub}(\beta)$ with $C_6^{lb,FB}(\beta)$ (in bits) for a binary input NEC with a Markov noise-erasure process given by Π_1 (recall that $C(\beta) \leq C_n^{ub}(\beta)$ and $C_n^{lb,FB}(\beta) \leq C_{FB}(\beta)$ for any n).

Theorem 5: Consider an NEC with stationary irreducible and aperiodic Markov noise-erasure process and feedback encoding rule given in (37). If the transition probabilities of the noise-erasure process satisfy that for a particular noise state $\tilde{s} \in \mathcal{Q}'$

$$P_{Z_i|Z_{i-1}}(z_i|\tilde{s}) = \begin{cases} \varepsilon', & \text{if } z_i = e \\ \frac{1-\varepsilon'}{q}, & \text{otherwise} \end{cases}$$

and

$$P_{Z_i|Z_{i-1}}(e|z_{i-1}) = \varepsilon'$$

for some $0 \leq \varepsilon' \leq 1$ and all $z_{i-1} \in \mathcal{Q}'$, then

$$C_{FB}(\beta) > C(\beta) \text{ for } \beta^{lb} \leq \beta < \frac{q-1}{2},$$

where

$$\beta^{lb} = [1 - P_Z(\tilde{s})]\frac{q-1}{2}.$$

Proof: Let $P_{V^n}^*(v^n) = \frac{1}{q^n}$ for any $v^n \in \mathcal{Q}^n$. For the non-feedback channel with input distribution $P_{V^n}^*$, since $P_{V^n}^*$ achieves C_n, we have

$$\beta_{\max}^{(n)} = \frac{1}{n}\sum_{v^n} P_{V^n}^*(v^n)b(v^n) = \sum_v P_{V_1}^*(v)b(v) = \frac{q-1}{2} = \beta_{\max},$$

$$C^{(n)}\Big(\frac{q-1}{2}\Big) = (1-\varepsilon)\log q - \frac{1}{n}[H(Z^n) - H(\tilde{Z}^n)],$$

and

$$C\Big(\frac{q-1}{2}\Big) = (1-\varepsilon)\log q - [\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})].$$

Thus, from Lemma 2, we have

$$C(\beta) < (1-\varepsilon)\log q - [\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})] \text{ for } \beta < \frac{q-1}{2}. \tag{38}$$

For the feedback channel with input distribution $P^*_{V^n}$ and the encoding rule f^*, we have

$$\beta^{lb}_n \triangleq \frac{1}{n}E[b(X^n)] = \frac{1}{n}\sum_{i=1}^{n} E[b(X_i)]$$

$$= \frac{1}{n}\sum_{v_1} P^*_{V_1}(v_1)b(v_1) + \frac{1}{n}\sum_{i=2}^{n} E[b(f^*(V_i, Z_{i-1}))]$$

$$= \frac{1}{n}\frac{q-1}{2} + \frac{n-1}{n}\sum_{v}\sum_{z} P_Z(z)P^*_V(v)b(f^*(v,z))$$

$$= \frac{1}{n}\frac{q-1}{2} + \frac{n-1}{n}\Big[\sum_{v} P_Z(\tilde{s})P^*_V(v)b(0) + \sum_{z\neq\tilde{s}}\sum_{v} P_Z(z)P^*_V(v)b(v)\Big]$$

$$= \frac{1}{n}\frac{q-1}{2} + \frac{n-1}{n}\sum_{z\neq\tilde{s}} P_Z(z)\frac{q-1}{2}$$

$$= \Big[1 - \frac{n-1}{n}P_Z(\tilde{s})\Big]\frac{q-1}{2}.$$

Note that since $\boldsymbol{Z}$ is an irreducible and aperiodic stationary Markov process, $P_Z(\tilde{s}) > 0$, and thus $\beta^{lb}_n < \frac{q-1}{2}$. The uniform input distribution $P^*_{V^n}$ may not be the optimal input distribution achieving $C^{lb,FB}_n(\beta^{lb}_n)$, implying that for V^n with distribution $P^*_{V^n}$, we have

$$C^{lb,FB}_n(\beta^{lb}_n) \geq \frac{1}{n}I(V^n; Y^n) = \frac{1}{n}[H(Y^n) - H(Y^n|V^n)]$$

$$= \frac{1}{n}\sum_{i=1}^{n} H(Y_i|Y^{i-1}) - \frac{1}{n}H(Y^n|V^n). \tag{39}$$

For the second term in (39), we have

$$H(Y^n|V^n) = H(Y^n|V^n, X_1) \tag{40}$$

$$= H(Z_1, Y^n|V^n, X_1) \tag{41}$$

$$= H(Z_1|V^n, X_1) + H(Y^n|V^n, X_1, Z_1)$$

$$= H(Z_1) + H(Y_2^n|V^n, X_1, Z_1) \tag{42}$$

$$= H(Z_1) + H(Y_2^n|V^n, X_1, Z_1, X_2) \tag{43}$$

$$= H(Z_1) + H(Z_2, Y_2^n|V^n, X_1, Z_1, X_2)$$

$$= H(Z_1) + H(Z_2|V^n, X_1, Z_1, X_2) + H(Y_2^n|V^n, X_1, Z_1, X_2, Z_2)$$

$$= H(Z_1) + H(Z_2|Z_1) + H(Y_3^n|V^n, X_1, Z_1, X_2, Z_2)$$

$$= H(Z^n), \tag{44}$$

where (40) holds since $X_1 = V_1$, (41) follows from (7), (42) follows form (1) and the fact that the noise process is independent of the message, (43) holds since $X_2 = f^*(V_2, Z_1)$ and (44) is obtained by repeating the steps (40)-(43). To analyze the first term in (39), we consider $\Pr(Y_i = y_i|Y^{i-1} = y^{i-1})$ for two cases:

- For $Y_i = e$, we have

$$\Pr(Y_i = e|Y^{i-1} = y^{i-1})$$

$$= \sum_{z_i} \Pr(Z_i = z_i, Y_i = e|Y^{i-1} = y^{i-1})$$

$$= \Pr(Z_i = e|Y^{i-1} = y^{i-1}) \tag{45}$$

$$= \Pr(\tilde{Z}_i = e|Y^{i-1} = y^{i-1}),$$

where (45) follows from the fact that $Y_i = e$ if and only if $Z_i = e$ by (1).

- For $Y_i = y_i \neq e$, we have

$$\Pr(Y_i = y_i|Y^{i-1} = y^{i-1})$$

$$= \sum_{z_i, z_{i-1}, v_i, x_i} \Pr(Z_{i-1} = z_{i-1}, Z_i = z_i, V_i = v_i, X_i = x_i, Y_i = y_i|Y^{i-1} = y^{i-1})$$

$$= \sum_{z_i, z_{i-1}, v_i, x_i} \Pr(Z_{i-1} = z_{i-1}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|z_{i-1}) P_{V_i}^*(v_i) \times$$

$$\mathbf{1}(x_i = f^*(v_i, z_{i-1})) \cdot \mathbf{1}(y_i = \theta(x_i, z_i)) \tag{46}$$

$$= \sum_{z_i, z_{i-1} \neq \tilde{s}, v_i, x_i} \Pr(Z_{i-1} = z_{i-1}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|z_{i-1}) P_{V_i}^*(v_i) \times$$

$$\mathbf{1}(x_i = f^*(v_i, z_{i-1})) \cdot \mathbf{1}(y_i = \theta(x_i, z_i))$$

$$+ \sum_{z_i, v_i, x_i} \Pr(Z_{i-1} = \tilde{s}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|\tilde{s}) P_{V_i}^*(v_i) \mathbf{1}(x_i = f^*(v_i, \tilde{s})) \cdot \mathbf{1}(y_i = \theta(x_i, z_i))$$

$$= \sum_{z_i, z_{i-1} \neq \tilde{s}, x_i} \Pr(Z_{i-1} = z_{i-1}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|z_{i-1}) \frac{1}{q} \cdot \mathbf{1}(y_i = \theta(x_i, z_i))$$

$$+ q \sum_{z_i} \Pr(Z_{i-1} = \tilde{s}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|\tilde{s}) \frac{1}{q} \cdot \mathbf{1}(y_i = \theta(0, z_i))$$

$$= \sum_{z_{i-1} \neq \tilde{s}, x_i \in \mathcal{X}_{y_i}} \Pr(Z_{i-1} = z_{i-1}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(\tilde{h}(x_i, y_i)|z_{i-1}) \frac{1}{q}$$

$$+ \Pr(Z_{i-1} = \tilde{s}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(\tilde{h}(0, y_i)|\tilde{s})$$

$$= \sum_{z_{i-1} \neq \tilde{s}, z_i \neq e} \Pr(Z_{i-1} = z_{i-1}|Y^{i-1} = y^{i-1}) P_{Z_i|Z_{i-1}}(z_i|z_{i-1}) \frac{1}{q}$$

$$+ \Pr(Z_{i-1} = \tilde{s}|Y^{i-1} = y^{i-1}) \frac{1 - \varepsilon'}{q}$$

$$= \frac{1}{q} \sum_{z_{i-1} \neq \tilde{s}, z_i \neq e} \Pr(Z_{i-1} = z_{i-1}, Z_i = z_i|Y^{i-1} = y^{i-1})$$

$$+ \frac{1}{q} \Pr(Z_{i-1} = \tilde{s}|Y^{i-1} = y^{i-1}) P(Z_i \neq e|Z_{i-1} = \tilde{s})$$

$$= \frac{1}{q} \Pr(Z_{i-1} \neq \tilde{s}, Z_i \neq e|Y^{i-1} = y^{i-1}) + \frac{1}{q} P(Z_{i-1} = \tilde{s}, Z_i \neq e|Y^{i-1} = y^{i-1})$$

$$= \frac{1}{q} \Pr(Z_i \neq e|Y^{i-1} = y^{i-1})$$

$$= \frac{1}{q} \Pr(\tilde{Z}_i = 0|Y^{i-1} = y^{i-1}),$$

where (46) follows from the chain rule, the encoding rule, (1) and the fact that the noise process is Markov and independent of the message.

From the preceding analysis, we have

$$H(Y_i|Y^{i-1})$$

$$= \sum_{y^{i-1}} \Pr(Y^{i-1} = y^{i-1}) H(Y_i|Y^{i-1} = y^{i-1})$$

$$= \sum_{y^{i-1}} \Pr(Y^{i-1} = y^{i-1}) \Big[-\Pr(\tilde{Z}_i = e|Y^{i-1} = y^{i-1}) \log \Pr(\tilde{Z}_i = e|Y^{i-1} = y^{i-1})$$

$$- q \frac{\Pr(\tilde{Z}_i = 0|Y^{i-1} = y^{i-1})}{q} \log \frac{\Pr(\tilde{Z}_i = 0|Y^{i-1} = y^{i-1})}{q} \Big]$$

$$= \sum_{y^{i-1}} \Pr(Y^{i-1} = y^{i-1}) \Big[H(\tilde{Z}_i|Y^{i-1} = y^{i-1}) + \Pr(\tilde{Z}_i = 0|Y^{i-1} = y^{i-1}) \log q \Big]$$

$$= H(\tilde{Z}_i|Y^{i-1}) + (1 - \varepsilon) \log q$$

$$= H(\tilde{Z}_i|Y^{i-1}, \tilde{Z}^{i-1}) + (1 - \varepsilon) \log q. \tag{47}$$

Next, we consider $\Pr(\tilde{Z}_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1})$ for $(y^{i-1}, \tilde{z}^{i-1})$ with $\Pr(Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1}) > 0$. We have

$$\Pr(\tilde{Z}_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1})$$

$$= \sum_{z_{i-1}} \Pr(Z_{i-1} = z_{i-1}, \tilde{Z}_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1})$$

$$= \sum_{z_{i-1}} \Pr(Z_{i-1} = z_{i-1} | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1}) \Pr(\tilde{Z}_i = e | Z_{i-1} = z_{i-1})$$

$$= \sum_{z_{i-1} \neq e} \Pr(Z_{i-1} = z_{i-1} | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1}) \varepsilon'$$

$$\quad + \Pr(Z_{i-1} = e | Y^{i-1} = y^{i-1}, \tilde{Z}^{i-1} = \tilde{z}^{i-1}) \Pr(\tilde{Z}_i = e | Z_{i-1} = e). \tag{48}$$

If $\tilde{Z}_{i-1} = e$, then since $\boldsymbol{Z}$ is Markovian,

$$\Pr(\tilde{Z}_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = e, \tilde{Z}^{i-2} = \tilde{z}^{i-2})$$

$$= \Pr(Z_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = e, \tilde{Z}^{i-2} = \tilde{z}^{i-2})$$

$$= \Pr(Z_i = e | Z_{i-1} = e)$$

$$= \Pr(\tilde{Z}_i = e | \tilde{Z}_{i-1} = e). \tag{49}$$

If $\tilde{Z}_{i-1} = 0$, then using (48), we have

$$\Pr(\tilde{Z}_i = e | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = 0, \tilde{Z}^{i-2} = \tilde{z}^{i-2})$$

$$= \sum_{z_{i-1} \neq e} \Pr(Z_{i-1} = z_{i-1} | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = 0, \tilde{Z}^{i-2} = \tilde{z}^{i-2}) \varepsilon'$$

$$\quad + \Pr(Z_{i-1} = e | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = 0, \tilde{Z}^{i-2} = \tilde{z}^{i-2}) \Pr(\tilde{Z}_i = e | Z_{i-1} = e)$$

$$= \sum_{z_{i-1}} \Pr(Z_{i-1} = z_{i-1} | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = 0, \tilde{Z}^{i-2} = \tilde{z}^{i-2}) \varepsilon'$$

$$= \varepsilon' \tag{50}$$

$$= \Pr(\tilde{Z}_i = e | \tilde{Z}_{i-1} = 0), \tag{51}$$

where (50) holds since $\Pr(Z_{i-1} = e | Y^{i-1} = y^{i-1}, \tilde{Z}_{i-1} = 0, \tilde{Z}^{i-2} = \tilde{z}^{i-2}) = 0$. Since $\tilde{Z}_i$ is binary, (49) and (51) show that $\tilde{Z}_i - \tilde{Z}_{i-1} - (Y^{i-1}, \tilde{Z}^{i-2})$ form a Markov chain and thus $H(\tilde{Z}_i | Y^{i-1}, \tilde{Z}^{i-1}) = H(\tilde{Z}_i | \tilde{Z}_{i-1})$. From this, (39), (44) and (47), we have

$$C_n^{lb,FB}(\beta_n^{lb}) \geq (1 - \varepsilon) \log q - \frac{1}{n}[H(Z^n) - H(\tilde{Z}^n)]. \tag{52}$$

Recalling that

$$
\begin{aligned}
\lim_{n\to\infty} \beta_n^{lb} &= \lim_{n\to\infty} \left(1 - \frac{n-1}{n} P_Z(\tilde{s})\right) \frac{q-1}{2} \\
&= \left(1 - P_Z(\tilde{s})\right) \frac{q-1}{2} = \beta^{lb},
\end{aligned}
$$

the following lemma is proved in the appendix.

Lemma 9:

$$
\lim_{n\to\infty} C_n^{lb,FB}(\beta_n^{lb}) = C^{lb,FB}(\beta^{lb}).
$$

In view of the preceding lemma, taking the limit of both sides of (52) yields

$$
C^{lb,FB}(\beta^{lb}) \geq (1-\varepsilon)\log q - [\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}})].
$$

Since $C^{lb,FB}(\beta^{lb})$ is a lower bound to $C_{FB}(\beta^{lb})$ and $C_{FB}(\beta^{lb}) \leq C_{FB} = (1-\varepsilon)\log q - [\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}})]$, we have

$$
C_{FB}(\beta) = (1-\varepsilon)\log q - [\bar{H}(\mathbf{Z}) - \bar{H}(\tilde{\mathbf{Z}})] \text{ for } \beta^{lb} \leq \beta \leq \frac{q-1}{2},
$$

and by (38) we conclude that $C_{FB}(\beta) > C(\beta)$ for $\beta^{lb} \leq \beta < \frac{q-1}{2}$. ∎

Note that the NEC of Fig. 3 has a Markov transition matrix that satisfies exactly the conditions of Theorem 5. We next provide numerical results for Markov NECs which do not precisely meet these conditions. In Figs. 4 and 5, we plot $C_6^{ub}(\beta)$ and $C_6^{lb,FB}(\beta)$ (under the linear cost function and f^* given by (37)) for NECs with $q' = q = 2$, $h(x,z) = x \oplus_2 z$ and Markov transition matrices

$$
\mathbf{\Pi}_2 = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \end{bmatrix} \quad \text{and} \quad \mathbf{\Pi}_3 = \begin{bmatrix} 0.45 & 0.35 & 0.2 \\ 0.7 & 0.2 & 0.1 \\ 0.2 & 0.7 & 0.1 \end{bmatrix},
$$

respectively. These figures show that in fact Theorem 5 holds for a more general class of NECs and a wider range of costs β. Similar numerical results can be obtained for NECs with non-binary input alphabets.

Without input cost constraints, the channel, being quasi-symmetric, has the uniform distribution as the optimal capacity achieving input distribution under both feedback and non-feedback regimes. Thus the entropy rate of the channel output remains unchanged in the presence of feedback and feedback does not increase capacity. However, upon imposing an input cost constraint, the uniform input distribution is no longer optimal. In this case, feedback provides the
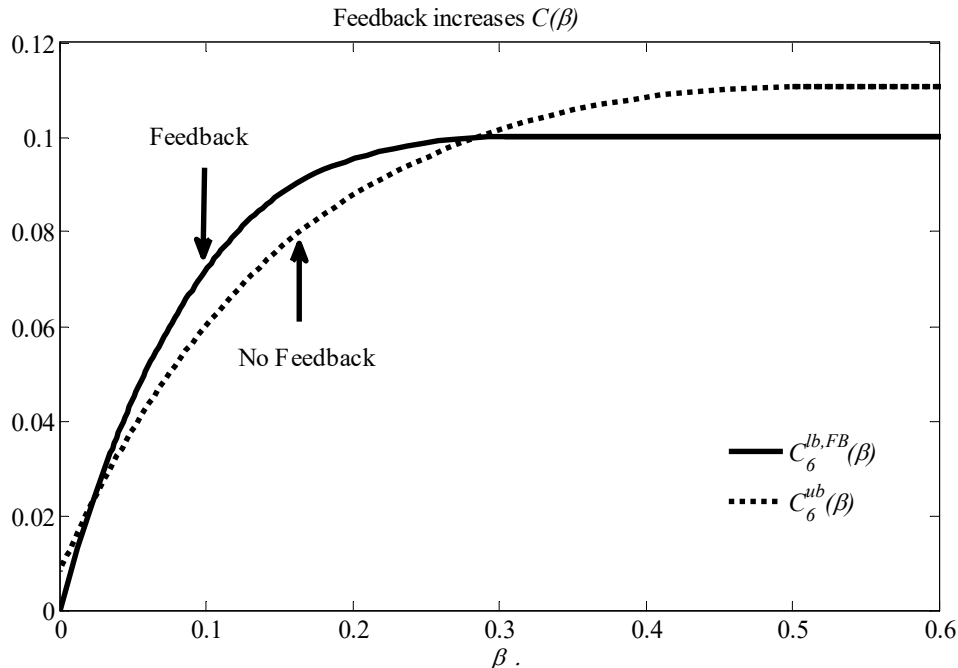
Fig. 4. Comparison of $C_6^{ub}(\beta)$ with $C_6^{lb,FB}(\beta)$ (in bits) for a binary input NEC with Markov noise-erasure given by Π_2 (recall that $C(\beta) \le C_n^{ub}(\beta)$ and $C_n^{lb,FB}(\beta) \le C_{FB}(\beta)$ for any n).

encoder useful information that, under feedback encoding rules judiciously selected in accordance with the cost function, can drive the input distribution to improve the entropy rate of the output, resulting in an increase in the channel capacity-cost function.

VII. CONCLUSION

We investigated a class of NECs satisfying invertibility conditions which can be viewed as a generalization of the EC and ANC with memory. The non-feedback capacity was derived in closed-form with the aide of an auxiliary erasure process with memory and proving that the n-fold channel is quasi-symmetric for all n. We then showed that the feedback capacity is identical to the non-feedback capacity, demonstrating that feedback does not increase capacity. We note that these results can be generalized to NECs with an arbitrary noise-erasure process (not necessarily stationary or information stable) using generalized spectral information measures [21]–[23]. The capacity-cost function of the NEC with and without feedback were also studied. We demonstrated, both analytically and numerically, that for a class of NECs with linear input costs and Markov noise-erasure processes, feedback does increase the capacity-cost function.
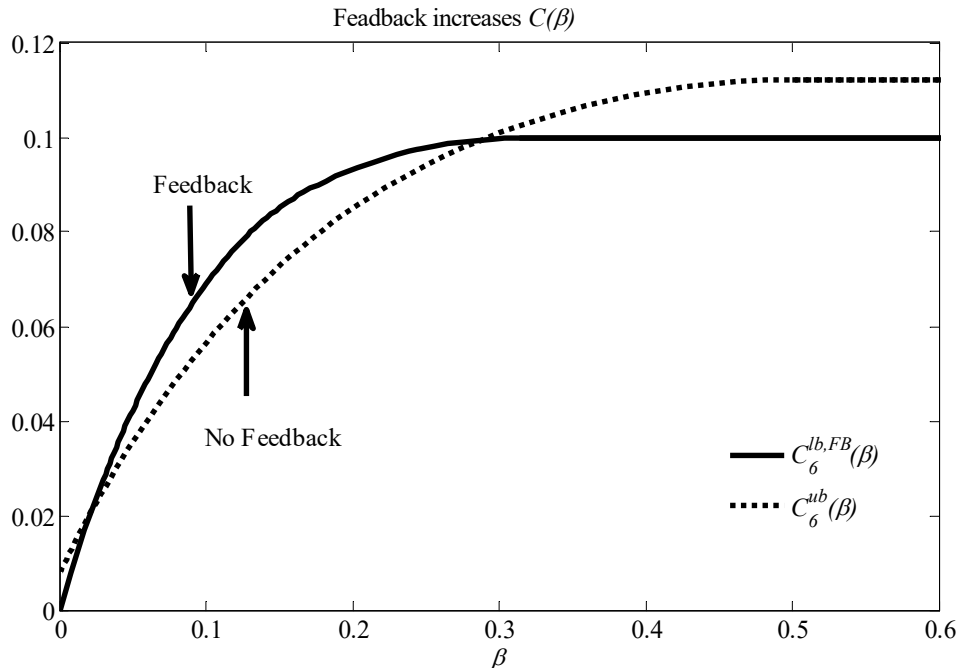
Fig. 5. Comparison of $C_6^{ub}(\beta)$ with $C_6^{lb,FB}(\beta)$ (in bits) for a binary input NEC with Markov noise-erasure given by $\mathbf{\Pi}_3$ (recall that $C(\beta) \leq C_n^{ub}(\beta)$ and $C_n^{lb,FB}(\beta) \leq C_{FB}(\beta)$ for any n).

Future work include deriving the non-feedback and feedback capacities of non quasi-symmetric NECs and of compound channels with NEC components.

APPENDIX

A. Proof of Lemma 6

Proof: If N and n are two integers such that $N > n \geq 1$, then we have

$$
\begin{aligned}
NH_N &= H(Z^N) - H(\tilde{Z}^N) \\
&= H(Z^N|\tilde{Z}^N) \\
&= H(Z^n, Z_{n+1}^N|\tilde{Z}^N) \\
&= H(Z^n|\tilde{Z}^N) + H(Z_{n+1}^N|\tilde{Z}^N, Z^n) \\
&\leq H(Z^n|\tilde{Z}^n) + H(Z_{n+1}^N|\tilde{Z}_{n+1}^N) \\
&= H(Z^n|\tilde{Z}^N) + H(Z^{N-n}|\tilde{Z}^{N-n}) \\
&= nH_n + (N-n)H_{N-n},
\end{aligned}
\tag{53}
$$

where (53) follows by writing $I(Z^N; \tilde{Z}^N)$ in two different ways and noting that $H(\tilde{Z}^N|Z^N) = 0$. Dividing both sides by N, we have that $H_N \leq \frac{n}{N} H_n + \frac{N-n}{N} H_{N-n}$, and hence the sequence $\{nH_n\}_{n=1}^{\infty}$ is subadditive. ∎

B. Corollary 1

We first prove the following lemma.

Lemma 10: If Y denotes the output of the NEC with invertibility conditions S-I and S-II, the input X and the noise Z are independent, and $\varepsilon = P_Z(e)$, then

$$\max_{P_X} H(Y) = (1 - \varepsilon) \log q - h_b(\varepsilon).$$

Proof: Noting that $\tilde{Z} = 0$ if $Z \neq e$ and that $\tilde{Z} = e$ if $Z = e$, we have

$$\max_{P_X} H(Y) = \max_{P_X}[I(X;Y) + H(Y|X)]$$

$$= \max_{P_X} I(X;Y) + H(Z)$$

$$= (1 - \varepsilon) \log q - (1 - \varepsilon)H(Z|\tilde{Z} \neq e) + H(Z) \tag{54}$$

$$= (1 - \varepsilon) \log q - (1 - \varepsilon)H(Z|\tilde{Z} \neq e) + H(Z, \tilde{Z})$$

$$= (1 - \varepsilon) \log q - (1 - \varepsilon)H(Z|\tilde{Z} \neq e) + H(\tilde{Z}) + H(Z|\tilde{Z})$$

$$= (1 - \varepsilon) \log q - (1 - \varepsilon)H(Z|\tilde{Z} \neq e) + h_b(\varepsilon) + (1 - \varepsilon)H(Z|\tilde{Z} \neq e)$$

$$= (1 - \varepsilon) \log q - h_b(\varepsilon),$$

where (54) follows from (9). ∎

Corollary 1: If in the setup of Lemma 10, random variable A is jointly distributed with Z and is conditionally independent of X and Y given Z, then

$$\max_{P_X} H(Y|A = a) = (1 - \varepsilon_a) \log q - h_b(\varepsilon_a),$$

for all $a \in \mathcal{A}$, where $\varepsilon_a = P(Z = e|A = a)$.

C. Proof of Lemma 9

We will show that $\{C_n^{lb,FB}\}$ is an equicontinuous family of functions on $[\beta^{lb}, \beta_{\max}]$. Since $C_n^{lb,FB}$ converges to $C^{lb,FB}$ pointwise, this will imply by Problem 16 of [52, Chap. 7] that the sequence of functions $\{C_n^{lb,FB}\}$ converges to $C^{lb,FB}$ uniformly over the interval $[\beta^{lb}, \beta_{\max}]$. Since

$\beta_n^{lb} \in [\beta^{lb}, \beta_{\max}]$ for all n and $\lim_{n\to\infty} \beta_n^{lb} = \beta^{lb}$, Problem 9 in [52, Chap. 7] will in turn imply that $C_n^{lb,FB}(\beta_n^{lb}) \to C^{lb,FB}(\beta^{lb})$ as $n \to \infty$ as claimed.

Since each $C_n^{lb,FB}$ is an increasing concave function with $C_n^{lb,FB}(0) = 0$, for all n and $\beta, \beta' \in [\beta^{lb}, \beta_{\max}]$ with $\beta' < \beta$, we have

$$0 \leq \frac{C_n^{lb,FB}(\beta) - C_n^{lb,FB}(\beta')}{\beta - \beta'} \leq \frac{C_n^{lb,FB}(\beta^{lb}) - C_n^{lb,FB}(0)}{\beta^{lb} - 0}$$
$$= \frac{C_n^{lb,FB}(\beta^{lb})}{\beta^{lb}} \leq \frac{C^{lb,FB}(\beta^{lb})}{\beta^{lb}},$$

where the last inequality holds since $C^{lb,FB}(\beta) = \sup_{n \geq 1} C_n^{lb,FB}(\beta)$. Thus for all n,

$$\left| C_n^{lb,FB}(\beta) - C_n^{lb,FB}(\beta') \right| \leq \frac{C^{lb,FB}(\beta^{lb})}{\beta^{lb}} \left| \beta - \beta' \right|$$

showing that $\{C_n^{lb,FB}\}$ is equicontinuous on $[\beta^{lb}, \beta_{\max}]$ and the lemma follows. $\qquad\square$

ACKNOWLEDGMENTS

REFERENCES

[1] L. Song, F. Alajaji and T. Linder, "On the capacity of burst noise-erasure channels with and without feedback," in *Proc. IEEE Int. Symp. Inf. Theory,* Aachen, Germany, June 2017, pp. 206-210.

[2] S. Johnson, "Burst erasure correcting LDPC codes," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 641-652, Mar. 2009.

[3] A. R. Krishnan, R. Radhakrishnan, B. Vasic, A. Kavcic, W. Ryan and F. Erden, "2-D Magnetic recording: Read channel modeling and detection," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3830-3836, Oct. 2009.

[4] K. Li, A. Kavcic, R. Venkataramani, M. Erden, "Channels with both random errors and burst erasures: Capacities, LDPC code thresholds, and code performances," in *Proc. IEEE Int. Symp. Inf. Theory,* Austin, Texas, June 2010, pp. 699-703.

[5] A. R. Iyengar, M. Papaleo, P. H.Siegel, J. K. Wolf, A. Vanelli-Coralli and G. E. Corazza, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2303-2320, Apr. 2012.

[6] P. Parag, J.-F. Chamberland, H. D. Pfister and K. Narayanan, "Code-rate selection, queueing behavior, and the correlated erasure channel," *IEEE Trans. Inf. Theory*, vol. 59, pp. 397-407, Jan. 2013.

[7] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, "Streaming codes with partial recovery over channels with burst and isolated erasures," *IEEE J. Sel. Topics Signal Processing*, vol. 9, pp. 501-516, Apr. 2015.

[8] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253-1266, Sep. 1960.

[9] E. O. Elliott, "Estimates of error rates for codes on burst-noise channel," *Bell Syst. Tech. J.*, vol. 42, pp. 1977-1997, Sep. 1963.

[10] M. Mushkin, I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inf. Theory*, vol. 35, pp. 1277-1290, Nov. 1989.

[11] J. Hamkins. "Optimal codes for the burst erasure channel," *IPN Progress Report* , pp. 42-174, 2008.

[12] F. Etezadi, A. Khisti and M. Trott, "Zero-delay sequential transmission of Markov sources over burst erasure channels," *IEEE Trans. Inf. Theory*, vol. 60, pp. 4584-4613, Jan. 2014.

[13] F. Etezadi, A. Khisti and J. Chen, "Price of perfection: Limited prediction for streaming over erasure channels," in *Proc. IEEE Int. Symp. Inf. Theory,* Hong Kong, June 2015, pp. 1392-1396.

[14] O. Sabag, H. Permuter and N. Kashyap, "The feedback capacity of the binary erasure channel with a no-consecutive-ones input constraint," *IEEE Trans. Inf. Theory*, vol. 62, pp. 8-22 Jan. 2016.

[15] Y. Li, and G. Han, "Asymptotics of input-constrained erasure channel capacity," *arXiv preprint arXiv:1605.02175*, 2016.

[16] D. Shaviv, A. Ozgur and H. Permuter, "Can feedback increase the capacity of the energy harvesting channel?" in *Proc. IEEE Inf. Theory Workshop*, pp. 1-5, 2015.

[17] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inf. Theory*, vol. 40, pp. 2035-2041, Nov. 1994.

[18] L. Zhong, F. Alajaji and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inf. Theory*, vol. 53, pp. 2815-2840, Aug. 2007.

[19] C. Pimentel, T. H. Falk and L. Lisbôa, "Finite-state Markov modeling of correlated Rician fading channels," *IEEE Trans. Veh. Technol.,* vol. 53, pp. 1491-1501, Sep. 2004.

[20] L. Zhong, F. Alajaji and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Vehicular Technology*, vol. 57, pp. 79-89, Jan. 2008.

[21] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inf. Theory*, vol. 41, pp. 546-549, Feb. 1995.

[22] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1147-1157, July 1994.

[23] T. S. Han, *Information-Spectrum Methods in Information Theory*, Springer, 2003.

[24] F. Alajaji and T. Fuja, "Effect of feedback on the capacity of discrete additive channels with memory," in *Proc. IEEE Int. Symp. Inf. Theory*, Trondheim, June 1994, p. 464.

[25] S. Loyka and C. D. Charalambous, "Feedback does not increase the capacity of compound channels with additive noise," in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, July 2016, pp. 1267-1271.

[26] F. Alajaji, and N. Whalen "The capacity-cost function of discrete additive noise channels with and without feedback," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1131-1140, Mar. 2000.

[27] R. Dobrushin, "General formulation of Shannon's main theorem in information theory," *Amer. Math. Soc. Trans.*, vol. 33, pp. 323-438, 1963.

[28] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes,* Holden-Day, 1964.

[29] R. G. Gallager, *Information Theory and Reliable Communication,* Wiley, 1968.

[30] J. Massey, "Causality, feedback, and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Nov. 1990, pp. 303-305.

[31] G. Kramer, *Directed Information for Channels with Feedback*, Ph.D. Dissertation, ser. ETH Series in Information Processing. Konstanz, Switzerland: Hartung-Gorre Verlag, vol. 11, 1998.

[32] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inform. Theory*, vol. 54, pp. 1488-1499, Apr. 2008.

[33] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, pp. 323-349, Jan. 2009.

[34] H. Permuter, T. Weissman and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inform. Theory*, vol. 55, pp. 644-662, Feb. 2009.

[35] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, pp. 780-798, Mar. 2005.

[36] H. Permuter, P. Cuff, B. Van Roy and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3150-3165, July 2008.

[37] O. Elischo and H. Permuter, "Capacity and coding for the Ising channel with feedback," *IEEE Trans. Inf. Theory*, vol. 60, pp. 5138-5149, Sep. 2014.

[38] N. Sen, F. Alajaji and S. Yüksel, "Feedback capacity of a class of symmetric finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 57, pp. 4110-4122, July 2011.

[39] O. Sabag, H. Permuter and N. Kashyap, "Feedback capacity and coding for the BIBO channel with a no-repeated-ones input constraint," *arXiv:1701.06174*, 2017.

[40] R. E. Blahut, *Principles and Practice of Information Theory,* Addison Wesley, Massachusetts, 1988.

[41] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication,* Addison Wesley, Massachusetts, 1977.

[42] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379-423, 623-656, July, Oct. 1948.

[43] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, pp. 44-54, Jan. 1995.

[44] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, pp. 8-19, 1956.

[45] T. M. Cover and J. Thomas, *Elements of Information Theory,* 2nd ed., Wiley, 2006.

[46] F. Alajaji and P.-N. Chen, *An Introduction to Single-User Information Theory,* Springer, 2018.

[47] R. L. Dobrushin and M. S. Pinsker, "Memory increases transmission capacity", *Probl. Pered. Inform.*, vol. 5, pp. 94-95, 1969.

[48] A. Wyner and J. Ziv, "Bounds on the rate-distortion function for stationary sources with memory", *IEEE Trans. Inf. Theory*, vol. 17, pp. 508-513, Sep. 1971.

[49] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, 1971.

[50] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pt. 4, pp. 142-163, Mar. 1959.

[51] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, pp. 460-473, July 1972.

[52] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.