On the Capacity of Burst Noise-Erasure Channels With and Without Feedback

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 satisfy a so-called "quasi-symmetry" condition 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.**

***Index Terms*—Channels with errors and erasures, channels with memory, symmetry, non-feedback and feedback capacities.**

I. Introduction

The 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 communication systems. 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 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 above the signal's amplitude, a decision error occurs 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 communication 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 noise-erasure channel (NEC) which incorporates both erasures and noise. This channel model, which subsumes both the BEC and the BSC, as well as their extensions with non-binary alphabets and memory, provides a potentially useful model for wireless networks, where data packets can be corrupted or dropped in a bursty fashion.

The authors are with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada (Emails: lin.song@queensu.ca, {fady, linder}@mast.queensu.ca). This work was supported in part by NSERC of Canada.

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. For the general q-ary erasure channel (EC), the input-output relationship can be expressed by

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

where $\{\tilde{Z}_i\}_{i=1}^{\infty}$ is an erasure process (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 $a \in \mathcal{Q} \cup e$. Coding schemes for burst ECs were studied in [1] and it was proved that maximum distance separable codes offer optimal burst erasure protection. The sequential transmission of Markov sources over burst ECs was considered in [2]. The feedback and non-feedback capacities of BECs with no-consecutive-ones at the input were investigated in [3] and [4], respectively. Explicit computations of the feedback and non-feedback capacities of energy harvesting BECs were also given in [5], where it was shown that feedback increases the capacity of such channels.

A discrete q-ary additive noise channel (ANC) with identical input and output alphabets $\mathcal{X} = \mathcal{Y} = \mathcal{Q}$ 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 noise process (that is independent of the input message) and $\oplus_q$ denotes modulo-q addition. Note that the BSC is a special case of an ANC: when $\{Z_i\}_{i=1}^{\infty}$ is binary-valued and memoryless (i.e., the Z_is are independent and identically distributed), the ANC reduces to the BSC. In [6] it was shown that feedback does not increase the capacity of ANCs with arbitrary memory. In particular, denoting the capacity with and without feedback by C_{FB}^{ANC} and C^{ANC}, respectively, it is proved in [6] 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 [7] of the noise process $\boldsymbol{Z} = \{Z_i\}_{i=1}^{\infty}$. The result of [6], which can also be proved for a larger class of channels [8], was recently extended in [9] for the family of compound channels with additive noise.

In this paper, we consider the NEC, a channel with both erasures and errors whose output Y_i at time i is given by

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

where input $X_i \in \mathcal{X} = \mathcal{Q}$, $Y_i \in \mathcal{Y} = \mathcal{Q} \cup e$, $\{Z_i\}_{i=1}^{n}$ is a noise-erasure process with alphabet $\mathcal{Z} = \mathcal{Q} \cup e$ which is independent of the input message, and $h : \mathcal{Q} \times \mathcal{Q} \to \mathcal{Q}$ is a deterministic function. ECs and ANCs are special cases of NECs. If $h(x, z) = x$ for any $x \in \mathcal{Q}$ and $z \in \mathcal{Q}$, then the NEC reduces to an EC. If $h(x, z) = x \oplus_q z$ and $P_{Z_i}(e) = 0$, then the NEC reduces to the ANC. We study the non-feedback and feedback capacities of the NEC under certain invertibility conditions on the function h in (1). 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 [7], [10]–[12], while the feedback capacity is expressed via the limit of the n-fold directed information [13], [14]. For some special cases, single-letter expressions or exact values of the capacities can be obtained. Examples of channels where feedback capacity is explicitly determined include the finite-state channel with states known at both transmitter and receiver [15], the trapdoor channel [16], the Ising channel [17], and the symmetric finite-state Markov channel [18].

In this work, 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 $\boldsymbol{Z} \triangleq \{Z_i\}_{i=1}^{\infty}$, and we prove that the non-feedback capacity of an NEC with stationary and 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. The proof is based on showing that the n-fold NEC is quasi-symmetric (as per Definition 5) and hence its n-fold mutual information is maximized by a uniformly distributed input process. Next, we 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 implies that feedback does not increase the capacity of the NEC and generalizes the feedback capacity results of [6] and [8]. The rest of this paper is organized as follows. We first provide some preliminary results in Section II. In Sections III and IV, we study the non-feedback capacity and feedback capacity, respectively, of a class of NECs with certain invertibility conditions. We conclude the paper in Section V.

II. Preliminaries

A. Feedback and Non-Feedback Capacities

We use capital letters such as X, Y, and Z to denote discrete 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 non-feedback channel code with blocklength n and rate $R \geq 0$ for the n-fold channel of $\boldsymbol{W}$ consists of an encoder $f^{(n)} : \mathcal{M} \to \mathcal{X}^n$ and a decoder $g^{(n)} : \mathcal{Y}^n \to \mathcal{M}$, where $\mathcal{M} = \{1, 2, ..., 2^{nR}\}$.

The encoder conveys message M, which is uniformly distributed over $\mathcal{M}$, by sending the sequence $X^n = f^{(n)}(M)$ over the channel which in turn is received as Y^n at the receiver. Upon estimating the sent message via $g^{(n)}(Y^n)$, the resulting decoding error probability is $P_e^{(n)} = \mathbf{Pr}\{g^{(n)}(Y^n) \neq M\}$.

Definition 2: The non-feedback channel capacity, denoted by C, is defined as the supremum of all rates R for which there exists a sequence of non-feedback channel codes with blocklength n and rate R, such that $\lim_{n\to\infty} P_e^{(n)} = 0$.

Recall that 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 [19] establishes that

$$C = \max_{P_X} I(X;Y) \tag{2}$$

for memoryless channels. The above coding theorem can be extended to show that (e.g., see [7], [10]–[12], [20])

$$C = \sup_n C_n = \lim_{n\to\infty} C_n \tag{3}$$

for stationary and information stable channels[1] where

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

We next consider the situation where the channel is equipped with an ideal feedback channel such that at any time instant $i > 1$, the encoder has access to all previously received channel outputs (from time 1 to $i-1$).

Definition 3: A feedback channel code with blocklength n and rate R for the n-fold channel of $\boldsymbol{W}$ 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 decoding function $g^{(n)} : \mathcal{Y}^n \to \mathcal{M}$, $\mathcal{M} = \{1, 2, ..., 2^{nR}\}$, where $\mathcal{Y}^0$ is the empty set.

Under feedback, the encoder maps the message M (which is uniformly distributed over $\mathcal{M}$) by taking into account the previously received channel outputs; in other words, the encoder sends the input sequence X^n, where $X_i = f_i^{(n)}(M, Y^{i-1})$ for $i = 1, \cdots, n$, over the channel which is received as Y^n and decoded as $g^{(n)}(Y^n)$. The decoding error probability is again given by $P_e^{(n)} = \mathbf{Pr}\{g^{(n)}(Y^n) \neq M\}$.

Definition 4: The feedback channel capacity, denoted by C_{FB}, is defined as the supremum of all rates R for which there exists a sequence of feedback channel codes with blocklength n and rate R, such that $\lim_{n\to\infty} P_e^{(n)} = 0$.

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

B. Quasi-symmetry

In general, the optimization problem in (2) is difficult to solve analytically. However, it is shown in [12], [21], [22] that when the channel satisfies certain "symmetry" properties

[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 [10], [11], [20] for a precise definition).

the optimal input distribution in (2) is uniform and the channel capacity can be expressed analytically. This result was further extended to so-called "quasi-symmetric" channels in [23].

The transition matrix of a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $\{W(y|x)\}_{x\in\mathcal{X},y\in\mathcal{Y}}$ 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 $x\in\mathcal{X}$, $y\in\mathcal{Y}$.

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 in $\mathbb{Q}$ are identical [21], [22].

Lemma 1 ([21], [22]): The capacity of a weakly-symmetric DMC is attained by the uniform input distribution and satisfies

$$C=\log|\mathcal{Y}|-H(q_1,q_2,...,q_{|\mathcal{Y}|})$$

where $(q_1,q_2,...,q_{|\mathcal{Y}|})$ is an arbitrary row of $\mathbb{Q}$ and

$$H(q_1,q_2,...,q_{|\mathcal{Y}|})=-\sum_{i=1}^{|\mathcal{Y}|}q_i\log q_i.$$

It readily follows that a symmetric DMC is weakly-symmetric. We also note that Gallager's notion for a symmetric channel [12, 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 5 ([23]): A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and 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 $\tilde{\mathbb{Q}}_i$ 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 2 ([23]): The capacity of a quasi-symmetric DMC is attained by the uniform input distribution and is given by

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

where, for $i=1,\cdots,m$, $a_i\triangleq\sum_{y\in\mathcal{Y}_i}p_{x,y}$ is the sum of any row of $\mathbb{Q}_i$, and

$$C_i=\log|\mathcal{Y}_i|-H\left(\text{any row of }\frac{1}{a_i}\mathbb{Q}_i\right)$$

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

III. Non-feedback capacity of a class of NECs

In this paper, we study a class of NECs with memory as defined in (1) and 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 $h(x,\cdot)$ is one-to-one, i.e., if $h(x,z)=h(x,\tilde{z})$ then $z=\tilde{z}$ for any $x\in\mathcal{Q}$. This condition implies the existence of a function $\tilde{h}:\mathcal{Q}\times\mathcal{Q}\to\mathcal{Q}$ such that for any x, $\tilde{h}(x,\cdot)$ is one-to-one and $h(x,\tilde{h}(x,y))=y$.
- (S-II) Given any $y\in\mathcal{Q}$, the function $\tilde{h}(\cdot,y)$ is one-to-one.

The above properties and (1) enable us to explicitly express the channel's noise-erasure variable Z_i at time i in terms of the input X_i and the output Y_i as follows

$$Z_i=\tilde{h}(X_i,Y_i)\cdot\mathbf{1}\{Y_i\neq e\}+e\cdot\mathbf{1}\{Y_i=e\}. \quad (4)$$

We further assume that the noise-erasure process $\boldsymbol{Z}=\{Z_i\}_{i=1}^{\infty}$ is stationary and ergodic and independent of the transmitted message. We next present our first main result.

Theorem 1: The capacity of an NEC without feedback is

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

where $\varepsilon=P_{Z_i}(e)$ is the probability of an erasure, $\bar{H}(\cdot)$ denotes 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

$$\tilde{Z}_i=\begin{cases}0 & \text{if } Z_i\neq e\\ e & \text{if } Z_i=e.\end{cases} \quad (5)$$

Proof: 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 (3):

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

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

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

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.[3] 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 which imply the quasi-symmetry of the NEC and whose proofs we omit due to space limitations.

[2]These conditions are similar to the ones considered in [8].

[3]The NEC, being quasi-symmetric, satisfies a weaker (and hence more general) notion of "symmetry" than the ANC [6] and the channel in [8] which are both symmetric.

Lemma 3: For any $\mathcal{S} \subseteq \mathcal{N}$, each row of $\mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}$ is a permutation of

$$\bar{p}_{\mathcal{Z}_\mathcal{S}} \triangleq [P_{Z^n}(z^n)]_{z^n \in \mathcal{Z}_\mathcal{S}},$$

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 $z^n \in \mathcal{Z}_\mathcal{S}$.

Lemma 4: 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 (5) and $\tilde{z}(n, \mathcal{S})$ denotes the n-tuple whose components satisfy

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

Now we are ready to calculate C_n. By Lemma 2, we have

$$\begin{aligned} C_n &= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \cdot \Big[\log q^{n-|\mathcal{S}|} \\ &\quad - H\Big(\text{any row of } \frac{1}{\sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n)} \mathbb{Q}_{\mathcal{Y}_\mathcal{S}|X^n}\Big)\Big] \\ &= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n) \Big[\log q^{n-|\mathcal{S}|} \\ &\quad -H\left(\left(\frac{P_{Z^n}(\hat{z}^n)}{\sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n)}\right)_{\hat{z}^n \in \mathcal{Z}_\mathcal{S}}\right)\Big] \\ &= \frac{1}{n}\sum_{\mathcal{S} \subseteq \mathcal{N}} \sum_{z^n \in \mathcal{Z}_\mathcal{S}} P_{Z^n}(z^n)\Big[\log q^{n-|\mathcal{S}|} - H(Z^n | Z^n \in \mathcal{Z}_\mathcal{S})\Big] \\ &= \frac{1}{n} \sum_{\mathcal{S} \subseteq \mathcal{N}} P_{\tilde{Z}^n}(\tilde{z}(n, \mathcal{S}))\Big[\log q^{n-|\mathcal{S}|} \qquad (6) \\ &\quad - H(Z^n | \tilde{Z}^n = \tilde{z}(n, \mathcal{S}))\Big] \\ &= \frac{1}{n}\Big[n \log q - \log q \sum_{\tilde{z}^n \in \tilde{\mathcal{Z}}^n} P_{\tilde{Z}^n}(\tilde{z}^n) \sum_{i=1}^n \mathbf{1}(\tilde{z}_i = e) \\ &\quad - \sum_{\tilde{z}^n \in \tilde{\mathcal{Z}}^n} P_{\tilde{Z}^n}(\tilde{z}^n) H(Z^n | \tilde{Z}^n = \tilde{z}^n)\Big] \\ &= \frac{1}{n}\Big[n \log q - \log q \cdot E\Big[\sum_{i=1}^n \mathbf{1}(\tilde{Z}_i = e)\Big] - H(Z^n | \tilde{Z}^n)\Big] \\ &= \log q - \frac{1}{n} \log q \sum_{i=1}^n E\Big[\mathbf{1}(\tilde{Z}_i = e)\Big] - \frac{1}{n} H(Z^n | \tilde{Z}^n) \\ &= (1 - \varepsilon) \log q - \frac{1}{n}(H(Z^n) - H(\tilde{Z}^n)), \end{aligned}$$

where (6) follows from (5). Therefore,

$$\begin{aligned} C &= \lim_{n \to \infty} C_n \\ &= (1 - \varepsilon) \log q - (\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})). \end{aligned}$$

■

Observation: We note the following special cases:

- If $\{Z_i\}_{i=1}^\infty$ is memoryless, then
$$\begin{aligned} C &= (1 - \varepsilon) \log q - (\bar{H}(\boldsymbol{Z}) - \bar{H}(\tilde{\boldsymbol{Z}})) \\ &= (1 - \varepsilon) \log q - H(Z_1 | \tilde{Z}_1). \end{aligned} \qquad (7)$$
- If we set $\mathcal{Z} = \{0, e\}$ and $h(x, z) = x$, then $Z_i = \tilde{Z}_i$ and $C = (1 - \varepsilon) \log q$, recovering the capacity of the burst EC [1]. We further remark that unlike the latter channel (for which memory in its erasure process does not increase capacity), the capacity of the NEC with memory can be strictly larger than the capacity of its memoryless counterpart (i.e., the channel with a memoryless noise-erasure process with identical marginal distribution as the NEC's stationary ergodic noise-erasure process). This can be for example realized for NECs with (strictly dependent) stationary mixing Markov noise-erasure processes.
- If $\varepsilon = 0$, then $C = \log q - \bar{H}(\boldsymbol{Z})$ and we recover the capacity of the discrete symmetric channel in [8] which subsumes the ANC [6].

IV. Feedback capacity of a class of NECs

We next show that feedback does not increase the capacity of NECs which satisfy our invertibility conditions on $h(\cdot, \cdot)$.

Theorem 2: Feedback does not increase the capacity of an NEC satisfying conditions S-I and S-II, i.e.,

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

where $\{\tilde{Z}_i\}_{i=1}^\infty$ is defined in (5).

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

$$\begin{aligned} nR &\le I(W; Y^n) + n\epsilon_n \qquad (8) \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Y_i | Y^{i-1}, W) + n\epsilon_n \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Z_i | Y^{i-1}, W, X^i, Z^{i-1}) + n\epsilon_n \qquad (9) \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - \sum_{i=1}^n H(Z_i | Z^{i-1}) + n\epsilon_n \qquad (10) \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}, \tilde{Z}^{i-1}) - H(Z^n) + n\epsilon_n \qquad (11) \\ &\le \sum_{i=1}^n H(Y_i | \tilde{Z}^{i-1}) - H(Z^n) + n\epsilon_n \\ &= \sum_{i=1}^n \sum_{\tilde{z}^{i-1}} P(\tilde{Z}^{i-1} = \tilde{z}^{i-1}) H(Y_i | \tilde{Z}^{i-1} = \tilde{z}^{i-1}) \\ &\quad - H(Z^n) + n\epsilon_n \\ &\le \sum_{i=1}^n \sum_{\tilde{z}^{i-1}} P(\tilde{Z}^{i-1} = \tilde{z}^{i-1}) \times \\ &\quad \max_{P_{X_i|\tilde{Z}^{i-1}}(\cdot|\tilde{z}^{i-1})} H(Y_i | \tilde{Z}^{i-1} = \tilde{z}^{i-1}) - H(Z^n) + n\epsilon_n \end{aligned}$$

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

$$+ h_b\big(P_{Z_i|\tilde{Z}^{i-1}}(e|\tilde{z}^{i-1})\big)\Big] - H(Z^n) + n\epsilon_n \quad (12)$$

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

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

where $\epsilon_n \to 0$ as $n \to \infty$, (8) follows from Fano's inequality, (9) holds since $X_i = f_i(W, Y^{i-1})$, $i = 1, 2, ..., n$, and Z_i is given by (4). Moreover, (10) holds because Z^n and W are independent of each other, (11) follows from (5) and (4), (12) 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

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

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

V. 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 based on introducing an auxiliary erasure process with memory and proving that the n-fold channel is quasi-symmetric. We then showed that the feedback capacity is identical to the non-feedback capacity, demonstrating that feedback does not increase capacity. We point out that our results in Theorems 1 and 2 can be generalized to NECs with arbitrary noise-erasure process (not necessarily stationary or information stable) using generalized spectral information measures [6], [7]. Future work include deriving the non-feedback and feedback capacities of compound channels with NEC components and examining the extension of the results of [9]. Another interesting direction is the study of the capacity-cost function of the NEC with and without feedback. Preliminary results indicate that for a class of NECs with linear input costs and a Markovian noise-erasure process, feedback does increase the capacity-cost function under a judiciously chosen feedback encoding strategy.

Appendix

Lemma 5: If Y denotes the output of the NEC with invertibility conditions S-I and S-II and $\varepsilon = P_Z(e)$, then

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

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

$$\max_{P_X} H(Y) = \max_{P_X}[I(X;Y) + H(Y|X)]$$
$$= \max_{P_X} I(X;Y) + H(Z)$$
$$= (1-\varepsilon)\log q - (1-\varepsilon)H(Z|\tilde{Z} \neq e) + H(Z) \quad (13)$$
$$= (1-\varepsilon)\log q - (1-\varepsilon)H(Z|\tilde{Z} \neq e) + H(\tilde{Z}) + H(Z|\tilde{Z})$$
$$= (1-\varepsilon)\log q - h_b(\varepsilon),$$

where (13) follows from (7). ■

Corollary 1: If 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 $a \in \mathcal{A}$, where $\varepsilon_a = P(Z = e|A = a)$.

References

[1] J. Hamkins. "Optimal codes for the burst erasure channel," *IPN Progress Report* , pp. 42-174, 2008.
[2] 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.
[3] Y. Li, and G. Han, "Asymptotics of input-constrained erasure channel capacity," *arXiv preprint arXiv:1605.02175*, 2016.
[4] 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.
[5] 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.
[6] 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.
[7] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1147-1157, July 1994.
[8] F. Alajaji and T. Fuja, "Effect of feedback on the capacity of discrete additive channels with memory," in *Proc. IEEE ISIT*, Trondheim, June 1994.
[9] S. Loyka and C. D. Charalambous, "Feedback does not increase the capacity of compound channels with additive noise," in *Proc. ISIT*, Barcelona, July 2016.
[10] R. Dobrushin, "General formulation of Shannon's main theorem in information theory," *Amer. Math. Soc. Trans.*, vol. 33, pp. 323-438, 1963.
[11] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes,* Holden-Day, 1964.
[12] R. G. Gallager, *Information Theory and Reliable Communication,* Wiley, 1968.
[13] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, pp. 323-349, Jan. 2009.
[14] 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.
[15] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. on Inform. Theory*, vol. 51, pp. 780-798, Mar. 2005.
[16] 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.
[17] 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.
[18] 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.
[19] C. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Journal*, vol. 27, pp. 379-423, 623-656, July, Oct. 1948.
[20] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, pp. 44-54, Jan. 1995.
[21] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication,* Cambridge Univ. Press, 1984.
[22] T. M. Cover and J. Thomas, *Elements of Information Theory,* 2nd ed., Wiley, 2006.
[23] F. Alajaji and P.-N. Chen, *Information Theory for Single-User Systems*, http://www.mast.queensu.ca/∼math474/it-lecture-notes.pdf.