Feedback Can Increase the Capacity-Cost Function of Discrete Channels with Memory†

Fady Alajaji and Tom Fuja

Electrical Engineering Department and Institute for Systems Research
University of Maryland, College Park, MD 20742

ABSTRACT

We consider a binary additive noise channel, where the noise process is a stationary mixing Markov process. Additive noise channels are *symmetric* in the sense that equiprobable input vectors maximize their block mutual information. We incorporate average power constraints on the possible input sequences of the additive channel, thus "destroying" its symmetry. We show that output feedback *increases* the capacity-cost function of this channel. This is achieved by establishing a tight upper bound to the capacity-cost function of the channel with no feedback ($C_{NFB}(\beta)$) and a lower bound to the capacity-cost function with feedback ($C_{FB}(\beta)$). The lower bound to $C_{FB}(\beta)$ is then shown numerically to exceed the upper bound to $C_{NFB}(\beta)$. An upper bound to $C_{FB}(\beta)$ is also obtained, by proving the converse to the coding theorem.

I. Introduction

We consider a binary communication channel with additive noise. The noise process is a stationary mixing Markov process. We assume that the channel is accompanied by a noiseless, delayless feedback channel with large capacity.

Additive noise channels are *symmetric* channels; by this we mean that the block mutual information between input and output vectors of an additive channel is maximized for equiprobable input blocks. In a previous related work [1], we showed that the capacity of discrete additive channels with feedback does not exceed their respective capacity without feedback. Indeed, this result holds for the class of discrete symmetric channels [4].

In achieving our result in [1], we did not introduce any restriction on the possible input sequences of the channels. However, in many communication channels, not every combination of the elements of the input alphabet can be used for transmission; this is due to physical restrictions that may arise in the channels. We therefore incorporate average cost constraints on the input sequences of the Markov additive channel and analyze its capacity-cost function with and without feedback.

Since this additive channel with memory becomes *non-symmetric* under input constraints, closed form expressions for its nonfeedback and feedback capacity-cost functions do not exist. This lead us to establish bounds on the non-feedback and feedback capacity-cost functions – denoted by $C_{NFB}(\beta)$ and $C_{FB}(\beta)$ respectively – in order to determine whether the introduction of output feedback brings about any potential increase in the capacity-cost function of the channel.

We first derive an upper bound to $C_{NFB}(\beta)$ which holds for all discrete channels with stationary mixing additive noise. The bound is the counterpart of the Wyner-Ziv lower bound to the rate-distortion function of stationary ergodic sources; this illustrates the striking duality that exists between the rate-distortion function and the capacity-cost function.

We then study the capacity-cost function of the channel with feedback. An encoding feedback strategy is developed for the case where the Markov noise process is assumed to be generated by the finite-memory contagion urn scheme of Polya [2]. It consists of adding at each instant of time, the maximum a posteriori (MAP) estimate of the current noise letter to the unencoded input letters representing the message to be sent. This results in a lower bound to $C_{FB}(\beta)$. An upper bound to $C_{FB}(\beta)$ is also derived, by proving the converse to the coding theorem.

The lower bound to $C_{FB}(\beta)$ and the upper bound to $C_{NFB}(\beta)$ are compared numerically, using Blahut's algorithm for the computation of the capacity-cost function. The results, computed assuming power cost constraints on the input, demonstrate that the lower bound to $C_{FB}(\beta)$ *exceeds* the upper bound to $C_{NFB}(\beta)$. For certain parameters of the channel, the gain in the capacity-cost function is $\geq 35.62\%$ if the average input cost is set at $\beta = 0.05$.

II. Preliminaries: The Capacity-Cost Function

Consider a discrete channel with finite input alphabet $\mathcal{X}$, finite output alphabet $\mathcal{Y}$ and n-fold transition probability $W^{(n)}(y_1, y_2, \dots, y_n | x_1, x_2, \dots, x_n)$, $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $i = 1, 2, \dots, n$. In general, the use of the channel is not free; we associate with each input letter x a nonnegative number $b(x)$, that we call the "cost" of x. The function $b(\cdot)$ is called the cost function. If we use the channel n consecutive times, i.e., we send an input vector $x^n = (x_1, x_2, \dots, x_n)$,

†Supported in part by NSF grant NCR-8957623; also by the NSF Engineering Research Centers Program, CDR-8803012.

the cost associated with this input vector is defined as

$$b(x^n) = \sum_{i=1}^{n} b(x_i).$$

For an input process $\{X_i\}_{i=1}^{\infty}$ with block input distribution $P^{(n)}(X^n = x^n)$ the *average cost* for sending X^n is defined by

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

Definition 1 An n-dimensional input random vector $X^n = (X_1, X_2, \ldots, X_n)$ that satisfies

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

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

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

We recall that a discrete channel is stationary if for every *stationary* input process $\{X_i\}_{i=1}^{\infty}$, the joint input-output process $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is stationary. Furthermore, a discrete channel is ergodic if for every *ergodic* input process $\{X_i\}_{i=1}^{\infty}$, the joint input-output process $\{(X_i, Y_i)\}_{i=1}^{\infty}$ is ergodic [10].

A discrete channel with stationary mixing (or weakly mixing) additive noise is stationary and ergodic [9, 12]. Furthermore, it has no anticipation and no input memory. Recall that a channel is said to have no anticipation (i.e., causal) if for a given input and a given input-output history, its current output is independent of future inputs. Furthermore, a channel is said to have no input memory (i.e., historyless) if its current output is independent of previous inputs. Refer to [9, 10] for more rigorous definitions of causal and historyless channels.

Definition 2 A channel block code of length n over $\mathcal{X}$ is a subset $\mathcal{C} = \{c_{(1)}, c_{(2)}, \ldots, c_{(M)}\}$ of $\mathcal{X}^n$ where each $c_{(i)}$ is an n-tuple. The rate of the code is $R = \frac{1}{n}\log_2 M$. The code is β-admissible if $b(c_{(i)}^n) \leq n\beta$ for $i = 1, 2, \ldots, M$. If the encoder wants to transmit message W where W is uniform over $\{1, 2, \ldots, M\}$, he sends the codeword $c_{(W)}$. At the channel output, the decoder receives Y^n and chooses as estimate of the message $\hat{W} = g(Y^n)$, where $g(\cdot)$ is a decoding rule. The probability of decoding error is then $P_e^{(n)} = Pr\{g(Y^n) \neq W\}$.

The *capacity-cost function* $C(\beta)$ is the supremum of all rates R for which there exist sequences of β-admissible block codes with vanishing probability of error as n grows to infinity (achievable codes). In other words, $C(\beta)$ is the maximum amount of information that can be transmitted reliably over the channel, if the channel must be used in such a way that the average cost is $\leq \beta$. If $b(x) = 0$ for every letter $x \in \mathcal{X}$, then $C(\beta)$ is just the channel *capacity* C as we know it.

It is known that for a discrete stationary ergodic channel with no anticipation and no input memory, the formulas of $C(\beta)$ and C are respectively given by [9]:

$$C(\beta) = \sup_n\ C_n(\beta) = \lim_{n\to\infty}\ C_n(\beta) \tag{1}$$

where $C_n(\beta)$ is the *n'th capacity-cost function* given by

$$C_n(\beta) \triangleq \max_{P^{(n)}(X^n) \in \tau_n(\beta)}\ \frac{1}{n}\ I(X^n; Y^n), \tag{2}$$

and

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

where

$$C_n \triangleq \max_{P^{(n)}(X^n)}\ \frac{1}{n}\ I(X^n; Y^n), \tag{4}$$

where $I(X^n; Y^n)$ is the block mutual information between the input vector X^n and the output vector Y^n.
Indeed, the above formulas for capacity hold for a larger class of channels, the class of *information stable* channels [12, 14]. A general formula for the capacity of *arbitrary* channels (not necessarily information stable) was recently derived in [14].

Remark: Note that the maximization in (2) is taken over all n-dimensional *β-admissible* input distributions. However, if the channel is a finite alphabet stationary channel with no input memory and no anticipation (e.g., an additive noise channel), then the maximization in (2) taken with respect to vector distributions $p^{(n)}(x^n)$ of *all* input processes is equal to the maximization taken with respect to vector distributions $p^{(n)}(x^n)$ of stationary ergodic input processes (Lemma 12.4.3 in [9]).

We now state the properties of $C(\beta)$ [11]. We first define respectively β_{min} and β_{max} by $\beta_{min} = \min_{x\in\mathcal{X}}\ b(x)$, and

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

From the definition of β_{min} above, we can see that $\frac{1}{n}E\left[b(X^n)\right] \geq \beta_{min}$ and so $C(\beta)$ is defined only for $\beta \geq \beta_{min}$.

Remark: For a discrete binary channel with additive noise $\{Z_n\}$ and power cost constraints on the input – i.e. $b(x) = x^2$ – we get that $\beta_{min} = 0$, $\beta_{max} = 1/2$, $C(\beta_{min}) = 0$ and $C(\beta_{max}) = C = 1 - H(Z_\infty)$ where $H(Z_\infty)$ is entropy rate of the noise.

Lemma 1 The capacity-cost function $C(\beta)$ defined in (1) is *concave* and *strictly increasing* in β for $\beta_{min} \leq \beta < \beta_{max}$, and is equal to C for $\beta \geq \beta_{max}$.

III. Discrete Channels with Binary Additive Markov Noise

We consider a discrete channel with memory, with common input, noise and output binary alphabet and described by the following equation: $Y_n = X_n \oplus Z_n$, for $n = 1, 2, 3, \ldots$ where $\oplus$ represents the addition operation modulo 2; and X_n, Z_n and Y_n are respectively the input, noise and output symbols of the channel. We assume that the input and noise sequences are independent from each other. The noise process $\{Z_n\}$ is stationary and mixing.
We now turn to the analysis of the capacity-cost function (with no feedback) of this channel. We denote it by $C_{NFB}(\beta)$.

A. An Upper Bound to $C_{NFB}(\beta)$

The channel is non-symmetric for $\beta < \beta_{max} = 1/2$. Thus the formula of $C_{NFB}(\beta)$ given by (1) will not have a closed form. We will then try to derive an upper bound to $C_{NFB}(\beta)$.

In [15], Wyner and Ziv derived a lower bound to the rate-distortion function ($R(D)$) of stationary ergodic sources:

$$R(D) \geq R_1(D) - \mu_1$$

where

- $R_1(D)$ is the rate-distortion function of an "associated" memoryless source with distribution equal to the marginal distribution $P^{(1)}(\cdot)$ of the ergodic source.
- $\mu_1 \triangleq H(X_1) - h(X_\infty)$, is the amount of memory in the source. $H(X_1)$ is the entropy of the associated memoryless source with distribution $P^{(1)}(\cdot)$ and $h(X_\infty)$ is the entropy rate of the original ergodic source.

This lower bound was later tightened by Berger [5]:

$$R(D) \geq R_n(D) - \mu_n \geq R_1(D) - \mu_1 \qquad (5)$$

where $R_n(D)$ is the n'th rate-distortion function of the source, $R_1(D)$ is as defined above and $\mu_n = \frac{1}{n}H(X^n) - h(X_\infty)$.
In light of the striking duality that exists between $R(D)$ and $C(\beta)$, as remarked previously by Shannon [13], we derive an equivalent upper bound to the nonfeedback capacity-cost function of a discrete additive channel.

Proposition 1 Consider a discrete channel with additive stationary mixing noise $\{Z_n\}$. Let $W^{(n)}(\cdot)$ denote the n-fold probability distribution function of the noise process. Then

$$C_{NFB}(\beta) \leq C_n(\beta) + M_n \leq C_1(\beta) + M_1 \qquad (6)$$

where

- $C_n(\beta)$ is the n-fold capacity-cost function without feedback of the channel as defined in (2).
- $C_1(\beta)$ is the capacity-cost function of the associated discrete memoryless channel (DMC) with iid additive noise process whose distribution is equal to the marginal distribution $W^{(1)}(\cdot)$ of the mixing noise process.
- $M_n = \frac{1}{n}H(Z^n) - h(Z_\infty)$, and $M_1 = H(Z_1) - h(Z_\infty)$.

The bound given above is asymptotically tight with n.

Proof 1 The proof uses a dual generalization of Wyner and Ziv's proof of the lower bound to the rate-distortion function. For further details, refer to [3].

B. A Lower Bound and an Upper Bound to $C_{FB}(\beta)$

We now consider the binary additive channel with output feedback. By this we mean that there exists a "return channel" from the receiver to the transmitter; we assume this return channel is noiseless, delayless, and has large capacity. The receiver uses the return channel to inform the transmitter what letters were actually received; these letters are received at the transmitter before the next letter is transmitted, and therefore can be used in choosing the next transmitted letter. We use the same definitions as stated in [1, 7], for feedback channel block code, probability of error, achievability and capacity-cost function with feedback (supremum of all achievable β-admissible feedback code rates). We denote the capacity-cost function of the channel with feedback by $C_{FB}(\beta)$.

We look at a particular class of additive noise channels; the class of channels with additive stationary mixing homogeneous Markov noise. We assume that the Markov noise process is generated by the finite-memory contagion urn scheme derived in [2]. It thus has a marginal distribution given by $W^{(1)}(Z_n = 1) = \rho = 1 - W^{(1)}(Z_n = 0)$, and a transition probability matrix given by

$$Q = [p_{ij}] = \begin{pmatrix} \frac{\sigma+\delta}{1+\delta} & \frac{\rho}{1+\delta} \\ \frac{\sigma}{1+\delta} & \frac{\rho+\delta}{1+\delta} \end{pmatrix} \qquad (7)$$

where $p_{ij} \triangleq Pr(Z_n = j | Z_{n-1} = i)$ and $\sigma = 1 - \rho$. We assume that $\rho < 1/2$ and $\delta > 0$. Note that if $\delta = 0$, the noise process becomes *iid*.
The transition matrix of this Markov model is *general*; it can represent any Markov chain with a positive[1] transition matrix.

We propose the following feedback encoding strategy. Since the encoder knows the previous noise samples, he can find the MAP estimate of the current noise sample. We obtain that if $\delta > 1 - 2\rho$, then the MAP estimate of Z_i is $\hat{Z}_i = Z_{i-1}$. (If $\delta \leq 1 - 2\rho$, then $\hat{Z}_i = 0$, a case of little interest.) Thus the encoder can try to fight the noise Z_i in the channel by adding the MAP estimate $\hat{Z}_i$ to the unencoded input letters. That is if $V^n(W)$ is a n-tuple vector representing message $W \in \{1, 2, \ldots, 2^{nR}\}$, then the encoder sends the feedback codeword $X^n(W) = (X_1, X_2, \ldots, X_n)$ where $X_i = V_i \oplus \hat{Z}_i = V_i \oplus Z_{i-1}$; $i = 1, 2, \ldots, n$ and $Z_0 = 0$ almost surely. The output of the channel is then $Y_i = X_i \oplus Z_i = V_i \oplus U_i$ where $U_i \triangleq Z_i \oplus Z_{i-1}$. Note that since the noise process $\{Z_i\}$ is stationary mixing and U_i is a *time-invariant* function of $\{Z_i\}$, the process $\{U_i\}$ is also stationary and mixing [12].

[1] A positive transition matrix is a matrix whose entries are all strictly positive.

Realize that the encoder transmits the feedback codeword X^n and thus the β-admissibility condition should be satisfied by the components of X^n.

Define $C^{lb}(\beta)$ by

$$C^{lb}(\beta) = \sup_n \ C_n^{lb}(\beta) = \lim_{n\to\infty} \ C_n^{lb}(\beta) \tag{8}$$

where

$$C_n^{lb}(\beta) = \max_{P^{(n)}(V^n)\in\tilde{\tau}_n(\beta)} \ \frac{1}{n} \ I(V^n;Y^n) \tag{9}$$

where

$$\tilde{\tau}_n(\beta) = \left\{ P^{(n)}(V^n) : \frac{1}{n} \ \sum_{i=1}^{n} \ E\left[b(V_i \oplus Z_{i-1})\right] \leq \beta \right\} \tag{10}$$

and $Y_i = V_i \oplus U_i$ where the process $\{U_i\}$ is defined by $U_i = Z_i \oplus Z_{i-1}$.

Definition 3 ([7]) Let (V^n, Y^n) be jointly distributed with distribution $p(v^n, y^n)$, where $V^n = (V_1, V_2, \ldots, V_n)$ and $Y^n = (Y_1, Y_2, \ldots, Y_n)$. Let $\epsilon > 0$, then the set $A_\epsilon^{(n)}$ of *jointly ϵ-typical* (V^n, Y^n) sequences is the set

$$\left\{ (v^n, y^n) \in \mathcal{V}^n \times \mathcal{Y}^n : \left| -\frac{1}{n} \ \log_2 p(v^n) - \frac{1}{n} H(V^n) \right| < \epsilon, \right.$$
$$\left| -\frac{1}{n} \ \log_2 p(y^n) - \frac{1}{n} H(Y^n) \right| < \epsilon,$$
$$\left. \left| -\frac{1}{n} \ \log_2 p(v^n, y^n) - \frac{1}{n} H(V^n, Y^n) \right| < \epsilon \right\},$$

where $p(v^n)$ and $p(y^n)$ are the respective n-fold marginal distributions derived from $p(v^n, y^n)$.

Proposition 2 (Lower Bound) Consider the binary Markov additive-noise channel with feedback. Then for $R < C^{lb}(\beta)$ there exists a sequence of β-admissible feedback codes with blocklength n and rate R such that $P_e^{(n)} \longrightarrow 0$ as $n \longrightarrow \infty$.

Proof 2 The proof uses random coding. Let $V^n(1)$,$V^n(2)$,…,$V^n(2^{nR})$ be independent identically distributed n-vectors, each drawn according to a distribution $p(v^n)$ that achieves $C^{lb}(\beta)$, where $p(v^n)$ is a n-fold distribution of a stationary ergodic process.

Transmission: To send message W, where W is uniform over $\{1, 2, \ldots, 2^{nR}\}$, the transmitter sends $X^n(W) = (X_1, X_2, \ldots, X_n)$ where $X_i = V_i \oplus \hat{Z}_i = V_i \oplus Z_{i-1}$; $i = 1, 2, \ldots, n$. $\{V_i\} \perp \{Z_i\}$ and thus $\{V_i\} \perp \{U_i\}$. We assume that $\delta > 1 - 2\rho$ in order to have $\hat{Z}_i = Z_{i-1}$. Note that the feedback codewords $X^n(W)$ are governed by a stationary ergodic distribution since $V^n(W)$ are selected according to a stationary ergodic $p(v^n)$ (that achieves $C^{lb}(\beta)$) and $\{Z_i\}$ is stationary mixing [9].

Decoding: The receiver receives $Y^n = (Y_1, Y_2, \ldots, Y_n)$ where $Y_i = X_i \oplus Z_i = V_i \oplus Z_{i-1} \oplus Z_i = V_i \oplus U_i$, $i = 1, 2, \ldots, n$.

The receiver declares $\hat{W} \in \{1, 2, \ldots, 2^{nR}\}$ was sent if $\left(V^n(\hat{W}), Y^n\right)$ is the only jointly ϵ-typical pair.

Error: An error is made if there is no ϵ-typical $\left(V^n(\hat{W}), Y^n\right)$ pair, more than one such pair, or $\hat{W} \neq W$. Furthermore, an error is made if the β-admissibility constraint is violated.

We investigate the probability of error $P_e^{(n)}$. We can assume without loss of generality that $W = 1$ was sent. We define the events $E_0 = \{\frac{1}{n}\sum_{i=1}^n b(X_i(1)) > \beta\}$ and

$$E_i = \left\{ (V^n(i), Y^n) \in A_\epsilon^{(n)} \right\}; \quad i = 1, 2, \ldots, 2^{nR}.$$

Let E_i^c be the complement of E_i. Then

$$\begin{aligned} P_e^{(n)} &= Pr(\hat{W} \neq 1 | W = 1) \\ &= Pr\left(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}} | W = 1\right) \\ &\leq Pr(E_0) + Pr(E_1^c | W = 1) + \sum_{i=2}^{2^{nR}} Pr(E_i | W = 1) \\ &= Pr(E_0) + Pr(E_1^c | W = 1) + 2^{nR} Pr(E_2 | W = 1) \end{aligned} \tag{11}$$

Since the codewords are drawn according to a stationary ergodic distribution that achieves $C^{lb}(\beta)$, then by the law of large numbers we get that $Pr(E_0) \leq \epsilon$ for n sufficiently large. We now state the following lemma [8, 9].

Lemma 2 (Joint AEP) Let $\left|A_\epsilon^{(n)}\right|$ denote the cardinality of the set $A_\epsilon^{(n)}$.

Then

$$\left|A_\epsilon^{(n)}\right| \leq 2^{n\left[\frac{1}{n}H(V^n, Y^n) + \epsilon\right]} \tag{12}$$

Furthermore, if the joint process $\{(V_n, Y_n)\}_{n=1}^\infty$ is stationary ergodic, then there exists n_0, such that for all $n > n_0$,

$$Pr\left((V^n, Y^n) \in A_\epsilon^{(n)}\right) \geq 1 - \epsilon. \tag{13}$$

Now, consider $V^n(1)$ as defined above and $Y^n = (Y_1, Y_2, \ldots, Y_n)$ such that $Y_i = V_i(1) \oplus U_i$, where $\{V_i\} \perp \{U_i\}$, $i = 1, 2, \ldots, n$, and the additive noise process $\{U_i\}$ is stationary mixing. Therefore the joint process $\{(V_n(1), Y^n)\}$ is stationary ergodic [9, 12]. By (13) in the above lemma, we thus get that $Pr(E_1^c | W = 1) = Pr\left((V^n(1), Y^n) \notin A_\epsilon^{(n)}\right) \leq \epsilon$ for n sufficiently large. Now,

$$Pr(E_2 | W = 1) = \sum_{(v^n, y^n) \in A_\epsilon^{(n)}} p(v^n) p(y^n),$$

since by our code generation $V^n(1)$ and $V^n(i)$ are independent for $i \neq 1$; thus Y^n and $V^n(i)$ are independent, $i \neq 1$.

$$Pr(E_2 | W = 1) \quad \leq \sum_{(v^n, y^n) \in A_\epsilon^{(n)}} 2^{-n\left[\frac{1}{n}H(V^n) - \epsilon\right]} \ 2^{-n\left[\frac{1}{n}H(Y^n) - \epsilon\right]}$$

$$
\begin{aligned}
&= \left|A_\epsilon^{(n)}\right| \; 2^{n\left[-\frac{1}{n}H(V^n)-\frac{1}{n}H(Y^n)+2\epsilon\right]} \\
&\leq 2^{n\left[-\frac{1}{n}H(V^n)-\frac{1}{n}H(Y^n)+\frac{1}{n}H(V^n,Y^n)+3\epsilon\right]} \\
&= 2^{-n\left[\frac{1}{n}H(Y^n)-\frac{1}{n}H(Y^n|V^n)+3\epsilon\right]} \\
&\leq 2^{-n\left[C^{lb}(\beta)+3\epsilon\right]}
\end{aligned}
$$

where the first inequality follows from Definition 3 and the second inequality above follows from (12).

Therefore (11) yields

$$P_e^{(n)} \;\leq\; 2\epsilon + 2^{-n\left[C^{lb}(\beta)-R+3\epsilon\right]} \;\leq\; 3\epsilon$$

if n is sufficiently large and $R < C^{lb}(\beta)$; and thus $C^{lb}(\beta)$ is achievable. ∎

Proposition 3 (Upper Bound) Any β-admissible sequence of feedback codes with blocklength n and rate R such that $P_e^{(n)} \longrightarrow 0$ as $n \longrightarrow \infty$, must have $R \leq C^{ub}(\beta)$, where

$$C^{ub}(\beta) = \lim_{n\to\infty} \max_{P^{(n)}(X^n)\in\tau_n(\beta)} \frac{1}{n} I(W, Y^n) \qquad (14)$$

where the maximization is taken over all β-admissible feedback codewords X^n of the form $X_i = X_i(W, Y^{i-1})$ for $i = 1, 2, \ldots, n$. $W \in \{1, 2, \ldots, 2^{nR}\}$ is the message, and the output is described by $Y_i = X_i(W, Y^{i-1}) \oplus Z_i$; $i = 1, 2, \ldots, n$.

Proof 3 The proof uses Fano's inequality. For more details, refer to [3].

Observation: Note that since the channel is additive, we can show that $I(W, Y^n) = H(Y^n) - H(Z^n)$ [1]. By comparing Propositions 2 and 3, we can assert that $C_{FB}(\beta) = C^{lb}(\beta) = C^{ub}(\beta)$ if it can be shown that the feedback encoding scheme we propose in Proposition 2 is indeed the optimal encoding scheme; a task we do not address at the present.

IV. Numerical Results

We are interested in determining whether feedback increases the capacity-cost function of our channel. Elucidating this statement would involve comparing $C^{lb}(\beta)$ with $C_n(\beta) + M_n$. This is investigated numerically using Blahut's algorithm for the computation of the capacity-cost function [6]. Using the algorithm of Theorem 10 in [6], with $b(x) = x^2$, we calculate $C_8^{lb}(\beta)$ and $C_8(\beta)$ for different values of the channel parameters δ and ρ (such that $\delta > 1 - 2\rho$). Since the noise is a Markov process, M_n can be expressed analytically in terms of n, δ and ρ; thus $C_8(\beta) + M_8$ is obtained.

Tighter results can be achieved for $n > 8$; however, the tightness improves as $1/n$ while the computation complexity increases exponentially. Some of the results, computed to an accuracy of 10^{-3} bits are plotted in Figures 1 and 2.

The Figures indicate to us that output feedback *does increase* the capacity-cost function of our binary Markov additive channel. We define

$$\Delta_8 = C_8^{lb}(\beta) - (C_8(\beta) + M_8) \qquad \text{bits per channel use,}$$

and

$$G_8 = \frac{\Delta_8}{C_8(\beta) + M_8} \times 100.$$

We therefore obtain that feedback increases the capacity-cost function by *at least* Δ_8 bits per channel use, which results in a gain of *at least* G_8 %. Values of Δ_8 and G_8 are computed in Table 1 for some values of β, ρ and δ. For $\delta = 0.5$, $\rho = 0.26$ and $\beta = 0.05$, we get a feedback gain of at least 0.026 bits, yielding an improvement $\geq 35.62\%$. For the same ρ and δ, at $\beta = 0.1$, $\Delta_8 = 0.023$ bits and $G_8 = 19.00\%$. For $\delta = 0.7$, $\rho = 0.16$ and $\beta = 0.1$, $\Delta_8 = 0.025$ and $G_8 = 11.47\%$.

V. Conclusion

In this paper, we analyzed a binary channel with additive stationary mixing Markov noise. We introduced average cost constraints on the input sequences of the additive channel, rendering it *non-symmetric*. The effect of output feedback on the capacity-cost function of the channel was examined. A tight upper bound to the capacity-cost function of the channel with no feedback ($C_{NFB}(\beta)$) was established. Furthermore, a lower bound to the capacity-cost function with feedback ($C_{FB}(\beta)$) was obtained. The converse to the coding theorem for the channel with feedback was proven; thus providing an upper bound to $C_{FB}(\beta)$. With the help of Blahut's algorithm for the computation of the capacity-cost function, we showed that $C_{FB}(\beta) > C_{NFB}(\beta)$.

In [3], we generalize these results for the case of binary channels with additive stationary mixing Markov noise of order M, where M is a positive constant. Further studies may involve the investigation of the capacity of non-symmetric channels with feedback, like the binary multiplicative channel (the "AND" channel) or the real adder channel. It is conjectured that feedback does cause an increase in capacity for such channels.

REFERENCES

[1] F. Alajaji, "Feedback Does Not Increase the Capacity of Discrete Channels with Additive Noise", submitted to *IEEE Transactions on Information Theory.*

[2] F. Alajaji and T. Fuja, "A Communication Channel Modeled on Contagion", submitted to *IEEE Transactions on Information Theory.*

[3] F. Alajaji and T. Fuja, "The Capacity-Cost Function of Discrete Additive Markov Channels with Feedback", to be submitted to *IEEE Transactions on Information Theory.*

[4] F. Alajaji, "New Results on the Analysis of Discrete Communication Channels with Memory", Ph.D. Dissertation, Department of Electrical Engineering, University of Maryland, College Park, MD 20742, in preparation.

[5] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, N.J., Prentice-Hall, 1971.

[6] R. E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions", *IEEE Transactions on Information Theory*, Vol. 18, No. 4, pp. 460-473, 1972.

[7] T. M. Cover and S. Pombra, "Gaussian Feedback Capacity", *IEEE Transactions on Information Theory*, Vol. 35, pp. 37-43, 1989.

[8] T. M. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, Inc., 1991.

[9] R. M. Gray, *Entropy and Information Theory*, Springer-Verlag New York Inc., 1990.

[10] R. M. Gray and D. S. Ornstein, "Block Coding for Discrete Stationary $\overline{d}$-Continuous Noisy Channels", *IEEE Transactions on Information Theory*, Vol. 25, pp. 292-306, 1979.

[11] R. J. McEliece, *The Theory of Information and Coding: A Mathematical Framework for Communication*, Cambridge University Press, 1984.

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

[13] C. E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion", *IRE Nat. Conv. Rec.*, Pt. 4, pp. 142-163, 1959.

[14] S. Verdú and T. S. Han, "A General Formula for Channel Capacity", *IEEE Transactions on Information Theory*, to appear.

[15] A. Wyner and J. Ziv, "Bounds on the Rate-Distortion Function for Stationary Sources with Memory", *IEEE Transactions on Information Theory*, Vol. 17, pp. 508-513, 1971.

β	δ	ρ	$C_8(\beta) + M_8$	$C_8^{lb}(\beta)$	Δ_8	*Gain* G_8
0.05	0.50	0.26	0.073	0.099	0.026	35.62 %
0.10	0.50	0.26	0.121	0.144	0.023	19.00 %
0.15	0.50	0.26	0.158	0.177	0.019	12.03 %
0.20	0.50	0.26	0.185	0.202	0.017	9.19 %
0.10	0.70	0.16	0.218	0.243	0.025	11.47 %
0.15	0.70	0.16	0.284	0.306	0.022	7.75 %
0.20	0.70	0.16	0.336	0.355	0.019	5.65 %
0.25	0.70	0.16	0.376	0.394	0.018	4.79 %
0.10	0.85	0.10	0.292	0.309	0.017	5.82 %
0.15	0.85	0.10	0.381	0.396	0.015	3.94 %
0.20	0.85	0.10	0.452	0.466	0.014	3.10 %
0.10	0.91	0.05	0.361	0.371	0.010	2.77 %
0.15	0.91	0.05	0.472	0.481	0.009	1.91 %
0.20	0.91	0.05	0.561	0.569	0.008	1.43 %
0.10	1.00	0.10	0.305	0.313	0.008	2.62 %
0.15	1.00	0.10	0.396	0.403	0.007	1.77 %
0.20	1.00	0.10	0.467	0.475	0.008	1.71 %

Table 1. Numerical results of feedback lower bound $C_8(\beta)$ and non-feedback upper bound $C_8(\beta) + M_8$ for the Markov binary channel.

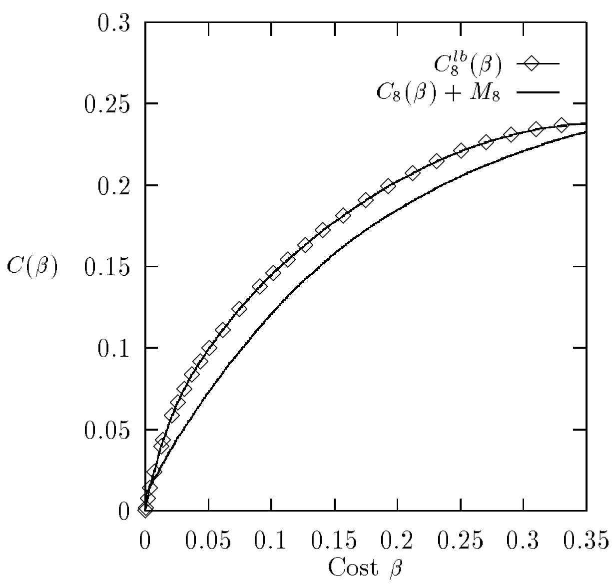

Figure 1. Markov binary channel with $\delta = 0.5$ and $\rho = 0.26$.

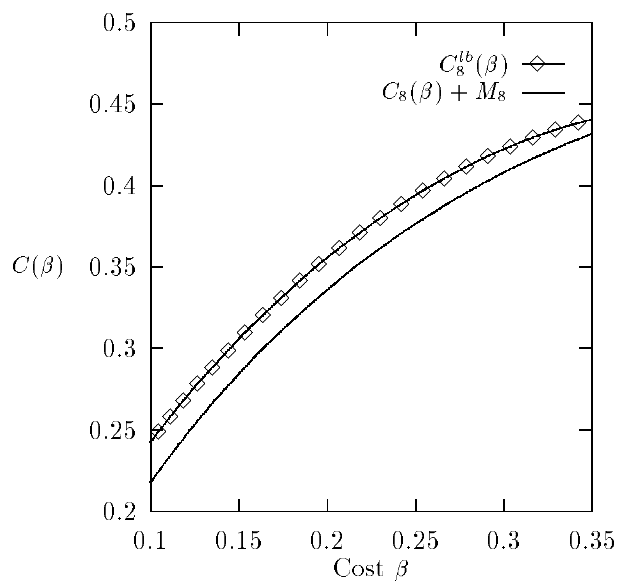

Figure 2. Markov binary channel with $\delta = 0.7$ and $\rho = 0.16$.