INSTITUTE FOR SYSTEMS RESEARCH

TECHNICAL RESEARCH REPORT

Feedback Does Not Increase the Capacity of Discrete Channels with Additive Noise

by F. Alajaji

T.R. 93-79r1

The Institute for Systems Research is supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), the University of Maryland, Harvard University, and Industry

Feedback Does Not Increase the Capacity of Discrete Channels with Additive Noise*

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

Submitted: March 1993
Revised: November 1993

Abstract

We consider discrete channels with additive random noise. We show that output feedback does not increase the capacity of such channels. This is first shown for both ergodic and non-ergodic additive stationary noise processes.

In light of recent results on channel capacity by Verdú and Han, we generalize our result for discrete channels with *arbitrary* non-stationary additive noise.

Keywords: Shannon theory, feedback capacity, discrete channels with additive noise.

0* Supported in part by NSF grant NCR-8957623; also by the NSF Engineering Research Centers Program, CDR-8803012. Parts of this paper were presented at the 1993 Conference on Information Sciences and Systems, March 24-26, 1993, The John Hopkins University, Baltimore, Maryland.

1 Introduction

We consider discrete channels with additive random noise. Note that such channels need not be memoryless; in general, they have memory. The Gilbert burst-noise channel [8], as well as the Polya-contagion channel [3], belong to the class of such channels. We assume that these channels are each accompanied by a noiseless, delayless feedback channel with large capacity. We show that the capacity of the channels with feedback does not exceed their respective capacity without feedback. This is shown for both ergodic and non-ergodic additive stationary noise processes. In light of recent results on a general channel capacity formula by Verdú and Han [17] we then generalize our result for discrete channels with *arbitrary* (non-stationary, non-ergodic in general) additive noise processes.

For these channels, the capacities with and without feedback are equal because additive noise channels are *symmetric* channels. By this we mean that the block mutual information (respectively the *inf-information* rate for the case of arbitrary additive noise) between input and output processes is maximized by equally likely *iid* input process.

In earlier related work, Shannon [16] showed that feedback does not increase the capacity of discrete memoryless channels. The same result was proven to be true for continuous channels with additive white Gaussian noise. Later, Cover and Pombra [7] and others considered continuous channels with additive non-white Gaussian noise and showed that feedback increases their capacity by at most half a bit; similarly, it has been shown [7] that feedback can at most double the capacity of a non-white Gaussian channel.

2 Discrete Channels with Stationary Ergodic Additive Noise

2.1 Capacity with no Feedback

Consider a discrete channel with common input, noise and output q-ary alphabet A where $A = \{0, 1, \ldots, q-1\}$, 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 q.

- The random variables X_n, Z_n and Y_n are respectively the input, noise and output of the channel.

- $\{X_n\} \perp \{Z_n\}$, i.e. the input and noise sequences are independent from each other.

- The noise process $\{Z_n\}_{n=1}^{n=\infty}$ is stationary and ergodic.

Note that additive channels defined above, are "non-anticipatory" channels; where by "non-anticipatory" we mean channels with no input memory (i.e., historyless) and no anticipation (i.e., causal) [12]. A channel is said to have no anticipation 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 if its current output is independent of previous inputs. Refer to [12] for more rigorous definitions of causal and historyless channels.

We furthermore note that discrete additive noise channels are *symmetric* channels. Symmetric channels are channels for which the block mutual information (respectively the inf-information rate for general channels) is maximized by equally likely *iid* input process. This is due to the facts that the input and noise processes of the channel are independent from each other, the addition operation (modulo q) is invertible and the input and output alphabets are *finite* and have the *same cardinality.*

A channel code with blocklength n and rate R consists of an encoder

$$f : \{1, 2, \ldots, 2^{nR}\} \rightarrow A^n$$

and a decoder

$$g : A^n \rightarrow \{1, 2, \ldots, 2^{nR}\}.$$

The encoder represents the message $V \in \{1, 2, \ldots, 2^{nR}\}$ with the codeword $f(V) = X^n = [X_1, X_2, \ldots, X_n]$ which is then transmitted over the channel; at the receiver, the decoder observes the channel output $Y^n = [Y_1, Y_2, \ldots, Y_n]$, and chooses as its estimate of the message $\hat{V} = g(Y^n)$. A decoding error occurs if $\hat{V} \neq V$.

For additive channels, $Y_i = X_i \oplus Z_i$ for all i. We assume that V is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$. The probability of decoding error is thus given by:

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} Pr\{g(Y^n) \neq V | V = k\} = Pr\{g(Y^n) \neq V\}$$

We say that a rate R is *achievable (admissable)* if there exists a sequence of codes with blocklength n and rate R such that

$$\lim_{n\to\infty} P_e^{(n)} = 0.$$

The objective is to find an admissable sequence of codes with as high a rate as possible. The capacity of the channel is defined as the supremum of the rate over all admissable sequences of codes. We denote it by C_{NFB}, to stand for capacity with no feedback.

Because the channel is a discrete channel with additive stationary ergodic noise, the nonfeedback capacity C_{NFB} of this channel is known and is equal to ([17], [14]):

$$\begin{aligned} C_{NFB} &= \lim_{n\to\infty} \sup_{X^n} \frac{1}{n} I(X^n;Y^n) & (1) \\ &= \log_2(q) - \lim_{n\to\infty} \frac{1}{n} H(Z^n) & (2) \end{aligned}$$

where

$$X^n = (X_1, X_2, \ldots, X_n),$$

$$Y^n = (Y_1, Y_2, \ldots, Y_n),$$

$$Z^n = (Z_1, Z_2, \ldots, Z_n),$$

$I(X^n;Y^n)$ is the mutual information between the input vector X^n and the output vector Y^n, and the supremum is taken over the input distributions of X^n. $H(Z^n)$ is the entropy of the noise vector Z^n. The expression in (2) can be shown to be the capacity of the channel using the Shannon-McMillan (AEP) theorem [14], [17].

2.2 Capacity with Feedback

We now consider the corresponding problem for the discrete additive channel with complete 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.

A feedback code with blocklength n and rate R consists of sequence of encoders

$$f_i : \{1, 2, \ldots, 2^{nR}\} \times A^{i-1} \rightarrow A$$

for $i = 1, 2, \ldots, n$, along with a decoding function

$$g : \ A^n \ \rightarrow \{1, 2, \ldots, 2^{nR}\}.$$

The interpretation is simple: If the user wishes to convey message $V \in \{1, 2, \ldots, 2^{nR}\}$ then the first code symbol transmitted is $X_1 = f_1(V)$; the second code symbol transmitted is $X_2 = f_2(V, Y_1)$, where Y_1 is the channel's output due to X_1. The third code symbol transmitted is $X_3 = f_3(V, Y_1, Y_2)$, where Y_2 is the channel's output due to X_2. This process is continued until the encoder transmits $X_n = f_n(V, Y_1, Y_2, \ldots, Y_{n-1})$. At this point the decoder estimates the message to be $g(Y^n)$, where $Y^n = [Y_1, Y_2, \ldots, Y_n]$.

Assuming our additive channel, $Y_i = X_i \oplus Z_i$ where $\{Z_i\}$ is a stationary ergodic noise process. Again, we assume that V is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$, and we define the probability of error and achievability as in Section 2.1.

Note, however, that because of the feedback, X^n and Z^n are no longer independent; X_i may depend on Z^{i-1}.

We will denote the capacity of the channel with feedback by C_{FB}. As before, C_{FB} is the supremum of all admissable feedback code rates.

Proposition 1 Feedback does not increase the capacity of channels with additive stationary ergodic noise:

$$C_{FB} = C_{NFB} = \log_2(q) \ - \lim_{n \rightarrow \infty} \frac{1}{n} \ H(Z^n) \tag{3}$$

Proof 1 Since V is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$, we have that $H(V) = nR$. Furthermore, $H(V) = H(V|Y^n) + I(V; Y^n)$. Now by Fano's inequality,

$$\begin{aligned} H(V|Y^n) &\leq h_b(P_e^{(n)}) + P_e^{(n)} \log_2(2^{nR} - 1) \\ &\leq 1 + P_e^{(n)} \log_2(2^{nR}) \\ &= 1 + P_e^{(n)} nR \end{aligned}$$

since $h_b(P_e^{(n)}) \leq 1$, where $h_b(\cdot)$ is the binary entropy function.

We then have:

$$\begin{aligned} nR &= H(V) \\ &= H(V|Y^n) + I(V;Y^n) \\ &\leq 1 + P_e^{(n)} nR + I(V;Y^n) \end{aligned}$$

where R is any admissable rate.

Dividing both sides by n and taking n to infinity, we get:

$$C_{FB} \leq \lim_{n\to\infty} \frac{1}{n} I(V;Y^n) \tag{4}$$

Let us thus study $I(V;Y^n)$:

$$I(V;Y^n) = \sum_{i=1}^{n} I(V;Y_i|Y^{i-1}) \tag{5}$$

but

$$I(V;Y_i|Y^{i-1}) = H(Y_i|Y^{i-1}) - H(Y_i|V,Y^{i-1}) \tag{6}$$

$$= H(Y_i|Y^{i-1}) - H(X_i \oplus Z_i|V,Y^{i-1}) \tag{7}$$

Now the fact that $X_i = f_i(V, Y_1, \ldots, Y_{i-1})$ implies that

$$H(X_i \oplus Z_i|V,Y^{i-1}) = H(Z_i|V,Y^{i-1},X_i) \tag{8}$$

$$= H(Z_i|V,Y^{i-1},X_i,X^{i-1},Z^{i-1}) \tag{9}$$

$$= H(Z_i|V,Y^{i-1},X^i,Z^{i-1}) \tag{10}$$

$$= H(Z_i|Z^{i-1}). \tag{11}$$

Here,

- Equation (8) follows from the fact that given V and Y^{i-1}, X_i is known deterministically and $H(Z+X|X) = H(Z|X)$.

- Equations (9) and (10) follow from the fact that given V and Y^{i-1}, we know all the previous transmitted letters $X_1, X_2, \ldots, X_{i-1}$ and thus we can recover all the previous noise letters $Z_j = Y_j - X_j \pmod q$ for $j = 1, 2, \ldots, i-1$.

- Equation (11) follows from the fact that Z_i and (V, Y^{i-1}, X^i) are conditionally independent given Z^{i-1}.

Therefore

$$I(V; Y_i|Y^{i-1}) = H(Y_i|Y^{i-1}) - H(Z_i|Z^{i-1}) \tag{12}$$

and

$$\begin{aligned} I(V; Y^n) &= \sum_{i=1}^{n} \left[H(Y_i|Y^{i-1}) - H(Z_i|Z^{i-1}) \right] && (13) \\ &= H(Y^n) - H(Z^n) && (14) \end{aligned}$$

But $H(Y^n) \leq \log_2 q^n$ because the channel is discrete. Therefore, if we divide both sides of (14) by n, and take n to infinity, we obtain that

$$C_{FB} \leq C_{NFB}$$

But by definition of a feedback code, $C_{FB} \geq C_{NFB}$ since a non-feedback code is a special case of a feedback code. Thus we get:

$$C_{FB} = C_{NFB} = \log_2(q) - \lim_{n \to \infty} \frac{1}{n} H(Z^n) \tag{15}$$

■

Observations:

1. It is important to note that for additive channels, the conditional noise entropy (given in equations (8)-(11)) remains the same *with or without feedback.* This is because addition is invertible; in general $H(X) \geq H(f(X))$ with equality holding for invertible functions $f(\cdot)$. This is true for both discrete and continuous alphabet additive channels.

2. The reason why output feedback potentially increases the capacity of additive non-white Gaussian channels [7] is because for continuous channels we have power constraints on the input, which upon optimization may increase $\lim_{n\to\infty} \frac{1}{n} H(Y^n)$ when feedback is used; while for discrete channels this quantity is upperbounded by $\log_2(q)$ and cannot be increased with feedback. In particular for discrete additive channels, the output entropy rate is equal to $\log_2(q)$ without feedback (symmetry property). It is therefore suspected that feedback might increase the capacity of discrete additive channels if we impose power constraints on the input.

3. The result given in Proposition 1 can be easily extended to discrete non-anticipatory channels with additive asymptotically mean stationary (AMS) ergodic noise process. Such class of noise processes include time-homogeneous ergodic Markov chains with arbitrary initial distributions. The proof is identical to that of Proposition 1, since the non-feedback capacity for the channel with AMS ergodic additive noise is still given by equation (2) [9], [17]. A random process has the AMS property (or is an AMS process) if its sample averages converge for a sufficiently large class of measurements (e.g., the indicator functions of all events); furthermore, there exists a stationary measure, called the "stationary mean" of the process, that has the same sample averages. A necessary and sufficient condition for a random process to possess ergodic properties with respect to the class of all bounded measurements is that it is AMS [10].

Finally, with the result of Proposition 1 in mind, it would be interesting to investigate discrete *non-additive* channels with known non-feedback capacities, and see whether output feedback would increase their capacities.

3 Discrete Channels with Stationary Non-Ergodic Additive Noise

3.1 Capacity with no Feedback

Consider a discrete channel similar to the one considered in Section 2 with the exception that the additive noise process $\{Z_n\}$ to the channel is stationary but *non-ergodic*. We will show in Proposition 2 that the resulting channel is an averaged channel whose components are discrete channels with *additive* stationary *ergodic* noise.

An averaged channel with stationary components is defined as follows: Consider a family of stationary channels parameterized by θ:

$$\left\{W_\theta^{(n)}(Y^n = y^n \mid X^n = x^n), \theta \in \Theta\right\}_{n=1}^{\infty} \tag{16}$$

where Y^n and X^n are respectively the input and output blocks of the channel, each of length n. $W_\theta^{(n)}(\cdot)$ is the block transition probabilities of the stationary channels, conditioned on a parameter $\theta \in \Theta$.

Definition 1 We define a channel to be an "averaged" communication channel with stationary components if its block transition probability $W_{ac}^{(n)}(Y^n = y^n \mid X^n = x^n)$ (where "ac" stands for averaged channel) is just the expected value of the block transition probability $\{W_\theta^{(n)}(Y^n = y^n \mid X^n = x^n)\}$ taken with respect to some distribution on θ – i.e., if it is of the form:

$$\begin{aligned} W_{ac}^{(n)}(Y^n = y^n \mid X^n = x^n) &= \int_\Theta W_\theta^{(n)}(Y^n = y^n \mid X^n = x^n)\, dG(\theta) & (17)\\ &= E_\theta[W_\theta^{(n)}(Y^n = y^n \mid X^n = x^n)] & (18)\end{aligned}$$

where $(\Theta, \sigma(\Theta), G)$ is the probability space on which the random variable θ is defined.

Note that the averaged channel has memory and is stationary. The averaged channel could be realized as follows: among the (countable or uncountable) stationary components, nature selects one of these components according to some probability distribution G. This component is then used for the entire transmission. However this selection is unknown to both the encoder and the decoder.

In order to show that we can write the block transition probability of the channel with additive stationary non-ergodic noise (which is equal to the block probability of the noise) as a mixture of the probabilities of stationary channels with additive ergodic noise (Proposition 2), we need to state first the ergodic decomposition theorem for stationary processes [11].

Notation: Consider a discrete time random process with an alphabet D, an event space (σ-field) $\sigma(D^\infty)$ consisting of subsets of the space D^∞ of sequences $u = (u_1, u_2, \ldots)$, $u_i \in D$, a probability measure μ on the space $(D^\infty, \sigma(D^\infty))$ forming a probability space $(D^\infty, \sigma(D^\infty), \mu)$ and a coordinate or sampling function $\mathbf{U}_n : D^\infty \longrightarrow D$ defined by $\mathbf{U}_n(u) = u_n$. The sequence of random variables $\{\mathbf{U}_n; n = 1, 2, \ldots\}$ defined on the probability space $(D^\infty, \sigma(D^\infty), \mu)$ is a discrete time random process. As convenient, random processes will be denoted by either $\{\mathbf{U}_n\}$ (to emphasize the sequence of random variables), or by $[D, \mu, \mathbf{U}]$ (to emphasize alphabet, probability measure, and name of the random variable).

Lemma 1 (Ergodic Decomposition Theorem) Let $[D, \mu, \mathbf{U}]$ be a stationary, discrete time random process. There exists a class of stationary ergodic measures $\{\mu_\theta; \theta \in \Theta\}$ and a probability measure G on a event space of Θ such that for every event $F \subset \sigma(D^\infty)$ we can write:

$$\mu(F) = \int_\Theta \mu_\theta(F)\, dG(\theta) \tag{19}$$

Remark: The ergodic decomposition theorem states that, in an appropriate sense, all stationary nonergodic random processes have the form of equation (19) of being a mixture of stationary ergodic processes; that is if we are viewing a stationary non-ergodic process, we are in reality viewing a stationary ergodic process selected by nature according to some probability measure G. Therefore, by directly applying the ergodic decomposition theorem we get the following result:

Proposition 2 A discrete channel with stationary non-ergodic additive noise process is an averaged channel with stationary channels with additive ergodic noise as components.

Proof 2 Since the additive noise process is independent of the input process, we can write:

$$W^{(n)}(Y^n = y^n \mid X^n = x^n) = W^{(n)}(Z^n = y^n - x^n \ (mod\ q))$$

Now, applying the ergodic decomposition theorem on the non-ergodic noise process $\{Z_n\}$, we get our result with each of the components being a channel with additive stationary ergodic noise:

$$W^{(n)}(Y^n = y^n \mid X^n = x^n) = \int_{\Theta} W_\theta^{(n)}(Z^n = y^n - x^n \ (mod\ q))\ dG(\theta)$$

■

Non-Feedback Capacity of the Channel with Additive Noise: The resulting non-feedback capacity of the channel with additive non-ergodic noise is [13], [15]:

$$C_{NFB} = \log_2(q) - \text{ess}_\Theta \sup\ h(W_\theta) \tag{20}$$

where

- the noise entropy rate $h(W_\theta)$ is given by

$$h(W_\theta) \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n}\ H_n(W_\theta^{(n)}) \tag{21}$$

  with

$$H_n(W_\theta^{(n)}) \stackrel{\text{def}}{=} - \sum_{x^n, y^n \in A^n} W_\theta^{(n)}(y^n|x^n)\ Q^{(n)}(x^n)\ \log_2\ W_\theta^{(n)}(y^n|x^n) \tag{22}$$

  where the input block distribution $Q^{(n)}(x^n) = \frac{1}{q^n}$.

- and the essential supremum is defined by

$$\text{ess}_\Theta \sup\ f(\theta) \stackrel{\text{def}}{=} \inf\ [r : dG(f(\theta) \leq r) = 1] \tag{23}$$

3.2 Capacity with Feedback

As in the previous section, we consider the corresponding problem for the discrete additive channel with complete output feedback. Similarly, we define a feedback code with blocklength n and rate R, as a sequence of encoders

$$f_i : \{1, 2, \ldots, 2^{nR}\} \times A^{i-1} \to A$$

for $i = 1, 2, \ldots, n$, along with a decoding function

$$g :\ A^n\ \to \{1, 2, \ldots, 2^{nR}\}.$$

The interpretation of the functions is identical to those in Section 2.2.

Assuming our additive channel, $Y_i = X_i \oplus Z_i$ where $\{Z_i\}$ is a stationary non-ergodic noise process.

Here again, we assume that V is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$ and we use the same definitions of achievable rates, probability of decoding error and capacity as in Section 2.2.

Because of the feedback, X^n and Z^n are no longer independent; X_i depends causally on Z^{i-1}. We will denote the capacity of the channel *with feedback* by C_{FB}. We now get the following result:

Proposition 3 Feedback does not increase the capacity of channels with additive stationary non-ergodic noise:

$$C_{FB} = C_{NFB} = \log_2(q) - \text{ess}_\Theta \sup \; h(W_\theta)$$

Proof 3 The main idea of the proof is the following. The channel is an averaged channel whose components are stationary channels with additive ergodic noise. Since feedback does not increase the capacity of *each* of these components (as shown in Section 2), it therefore does not increase the capacity of the averaged channel.

To formalize this reasoning, we will show that the (weak) converse to the channel coding theorem still holds with feedback. The coding theorem itself obviously holds since a non-feedback code is a special case of a feedback code, and thus any rate that can be achieved without feedback, can also be achieved with feedback.

The additive channel is a mixture of channels with *additive* stationary ergodic noise, thus by Proposition 1, we obtain that for *each of these components*:

$$C_{FB}^{(\theta)} = C_{NFB}^{(\theta)}.$$

Now, examining equation (20), we have: $h(W_\theta) \leq \text{ess}_\Theta \sup \; h(W_\theta)$ a.e. Then for some small $\epsilon > 0$, there exists components $\theta \in \Theta$ such that:

$$h(W_\theta) > \text{ess}_\Theta \sup \; h(W_\theta) - \epsilon$$

or

$$\log_2(q) - h(W_\theta) < \log_2(q) - \text{ess}_\Theta \sup \; h(W_\theta) + \epsilon$$

or

$$C_{NFB}^{(\theta)} < C_{NFB} + \epsilon$$

And the probability of such components is $\delta > 0$.

By this we mean, that we can find among the stationary components, with probability $\delta > 0$, components with capacity $C_{NFB}^{(\theta)} < C_{NFB} + \epsilon$ for some small $\epsilon > 0$; i.e. $\delta = Pr\{\theta \in \Theta : C_{NFB}^{(\theta)} < C_{NFB} + \epsilon\} > 0$.

With feedback encoder f_i and message $V = k$, we define

$$A(x_k^n) = \left\{ y^n \in A^n : f_i(k, y^{i-1}) = x_k^{(i)} \ (i = 1, 2, \ldots, n) \right\},$$

where $x_k^n = \left(x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)} \right)$. The probability that the feedback codeword for the message k is x_k^n is given by $P_k(x_k^n) = \sum_{y^n \in A(x_k^n)} W^{(n)}(Y^n = y^n | X^n = x_k^n)$. Letting D_k be the decoding set for message k, the probability that the feedback codeword for the message k is x_k^n *and* a decoding error takes place is given by $Pe_k(x_k^n) = \sum_{y^n \in A(x_k^n) \cap D_k^c} W^{(n)}(Y^n = y^n | X^n = x_k^n)$, where D_k^c is the complement of D_k.

Hence, the probability of decoding error when message k was sent is written as

$$\begin{aligned} Pe_k^{(n)} &= Pr\{g(Y^n) \neq V | V = k\} \\ &= \sum_{x_k^n \in A^n} \sum_{y^n \in A(x_k^n) \cap D_k^c} W^{(n)}(Y^n = y^n | X^n = x_k^n) \end{aligned} \tag{24}$$

It should be noted here that $A(x_k^n)$ and D_k^c in the above summation do not depend on the channel $W^{(n)}(\cdot)$. Therefore, the overall average probability of decoding error is

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} \sum_{x_k^n \in A^n} \sum_{y^n \in A(x_k^n) \cap D_k^c} W^{(n)}(Y^n = y^n | X^n = x_k^n) \tag{25}$$

It is evident using Proposition 2, that this probability of error can be expressed as

$$P_e^{(n)} = \int_\Theta P_e^{(n)}(\theta) \, dG(\theta) \tag{26}$$

where

$$P_e^{(n)}(\theta) = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} \sum_{x_k^n \in A^n} \sum_{y^n \in A(x_k^n) \cap D_k^c} W_\theta^{(n)}(Y^n = y^n | X^n = x_k^n) \tag{27}$$

and $P_e^{(n)}(\theta)$ is the average probability of decoding error for the channel component $W_\theta^{(n)}(\cdot)$.

Now, suppose there exists a sequence of feedback codes with blocklength n and rate R, such that $R > C_{NFB} + 2\epsilon$. Thus we have:

$$P_e^{(n)} = \int_\Theta P_e^{(n)}(\theta)\, dG(\theta) \tag{28}$$

$$\geq \int_{\{\theta \in \Theta : C_{NFB}^{(\theta)} < C_{NFB} + \epsilon\}} P_e^{(n)}(\theta)\, dG(\theta) \tag{29}$$

We now recall the weak converse to the nonfeedback channel coding theorem for stationary channels with additive ergodic noise: if $R > C_{NFB} + \epsilon'$, for some small $\epsilon' > 0$, then there exists $\gamma > 0$, such that $P_e^{(n)} > \gamma$ for sufficiently large n. To show this, using Fano's inequality along with the fact that $H(V) = nR$ we have

$$\begin{aligned} nR &\leq 1 + P_e^{(n)} nR + I(V; Y^n) \\ &\leq 1 + P_e^{(n)} nR + I(X^n(V); Y^n) \end{aligned}$$

where the second inequality follows from the data processing theorem with $V \to X^n \to Y^n$ forming a Markov chain. Thus

$$\begin{aligned} P_e^{(n)} &\geq \frac{1}{R}\left(R - \frac{1}{n} I(X^n; Y^n) - \frac{1}{n}\right) \\ &\geq \frac{1}{R}\left(R - C_{NFB} - \frac{1}{n}\right) \\ &> \frac{1}{R}\left(\epsilon' - \frac{1}{n}\right) \end{aligned}$$

and the result is shown. Note that $\gamma \stackrel{\text{def}}{=} \frac{1}{R}\left(\epsilon' - \frac{1}{n}\right)$ is independent of the characteristics of the channel.

Therefore, applying the weak converse of the coding theorem for the stationary channel components with additive ergodic noise, we get that for $R > C_{NFB} + 2\epsilon > C^{(\theta)}_{NFB} + \epsilon$, there exists some small $\gamma > 0$, such that $P_e^{(n)}(\theta) > \gamma$, as $n \to \infty$. As mentioned above, γ is independent of θ and depends only on ϵ and R.

Then

$$\lim_{n\to\infty} P_e^{(n)} > Pr\{\theta \in \Theta : C^{(\theta)}_{NFB} < C_{NFB} + \epsilon\}\gamma = \delta\,\gamma > 0 \tag{30}$$

Therefore the weak converse is proved and $C_{FB} = C_{NFB}$. ∎

Observation: It should be noted that for general averaged channels, i.e. *non-additive* averaged channels, feedback might *increase* capacity. For example, if we consider an averaged channel with a *finite* number of *non-additive* discrete memoryless channels (DMC's), then the non-feedback capacity of the averaged channel is equal to the capacity of the corresponding compound memoryless channel [1]:

$$C^{(ac)}_{NFB} = \max_{Q^{(1)}} \inf_{\theta\in\Theta} I(Q^{(1)}; W^{(1)}_\theta) \tag{31}$$

Note that:

$$\begin{aligned} C^{(ac)}_{NFB} &\leq \inf_{\theta\in\Theta} \max_{Q^{(1)}} I(Q^{(1)}; W^{(1)}_\theta) \\ &= \inf_{\theta\in\Theta} C^{(\theta)} \end{aligned} \tag{32}$$

where $C^{(\theta)} = \max_{Q^{(1)}} I(Q^{(1)}; W^{(1)}_\theta)$ is the non-feedback capacity of each of the DMC components.

Now, if we use output feedback, the encoder knows the previous received outputs, and thus can determine by some statistical means, which one of the DMC components is being used. In the most pessimistic case, the capacity of this DMC component may be equal to $\inf_{\theta\in\Theta} C^{(\theta)}$. Thus the capacity with feedback of the averaged channel will be:

$$C^{(ac)}_{FB} = \inf_{\theta\in\Theta} C^{(\theta)} \tag{33}$$

Therefore $C^{(ac)}_{FB} \geq C^{(ac)}_{NFB}$. This result (equation (33)) is equivalent to the result already derived by Ahlswede for the discrete averaged channel with sender informed [2].

Finally, in the case for which the inequality in (32) holds with the *strict* inequality, we obtain that feedback *increases* capacity: $C_{FB}^{(ac)} > C_{NFB}^{(ac)}$. Refer to Section 2 in [6] for an example of a finite collection of DMC's for which (32) holds with the strict inequality.

4 Discrete Channels with Arbitrary Additive Noise

4.1 Capacity with no Feedback

Consider a discrete channel similar to the one considered in Section 2 with the exception that the additive noise process $\{Z_n\}$ to the channel is an *arbitrary* random process (non-sationary, non-ergodic in general). We again use the same definitions as stated in Section 2.1 for channel block code, probability of error, achievable (or admissable) code rates and operational capacity (the supremum of all achievable rates). We denote the nonfeedback capacity by C_{NFB}.

In [17], Verdú and Han derived a formula for the operational capacity of arbitrary single-users channels (not necessarily stationary, ergodic, information stable, etc.). The (nonfeedback) capacity was shown to equal the supremum, over all input processes, of the input-output *inf-information rate* defined as the liminf in probability of the normalized information density:

$$C_{NFB} = \sup_{X^n} \underline{I}\,(X^n; Y^n) \tag{34}$$

where $X^n = (X_1, X_2, \ldots, X_n)$, for $n = 1, 2, \ldots$, is the block input vector and $Y^n = (Y_1, Y_2, \ldots, Y_n)$ is the corresponding output sequence induced by X^n via the channel $W^{(n)} = P_{Y^n|X^n} : A^n \to B^n$; $n = 1, 2, \ldots$, which is an arbitrary sequence of n-dimentional conditional output distributions from A^n to B^n, where A and B are the input and output alphabets respectively.

The symbol $\underline{I}\,(X^n; Y^n)$ appearing in (34) is the *inf-information rate* between X^n and Y^n and is defined as the *liminf in probability* of the sequence of normalized information densities $\frac{1}{n}\, i_{X^nY^n}(X^n; Y^n)$, where

$$i_{X^nY^n}(a^n;b^n) = \log_2 \frac{P_{Y^n|X^n}(b^n|a^n)}{P_{Y^n}(b^n)} \tag{35}$$

The *liminf in probability* of a sequence of random varaibles is defined as follows: if A_n is a sequence of random variables, then its *liminf in probability* is the supremum of all reals α for which $P(A_n \leq \alpha) \longrightarrow 0$ as $n \longrightarrow \infty$. Similarly, its *limsup in probability* is the infimum of all reals β for which $P(A_n \geq \beta) \longrightarrow 0$ as $n \longrightarrow \infty$. Note that these two quantities are always defined; if they are equal, then the sequence of random variables converges in probability to a constant (which is α).

Using equation (34) as well as the properties of the inf-information rate derived in [17], we obtain that the inf–information rate in (34) is maximized for equiprobable iid X^n (symmetry property), yielding the following expression for the nonfeedback capacity of our discrete channel with arbitrary additive noise:

$$C_{NFB} = \log_2(q) - \overline{H}(Z^n) \tag{36}$$

where $Z^n = (Z_1, Z_2, \ldots, Z_n)$ and $\overline{H}(Z^n)$ is the sup-entropy rate of the additive noise process $\{Z_n\}$, which is defined as the limsup in probability of the normalized noise entropy density

$$\frac{1}{n} \log_2 \frac{1}{P_{Z^n}(Z^n)}.$$

4.2 Capacity with Feedback

As in the previous section, we consider the corresponding problem for the discrete additive channel with complete output feedback. Similarly, we use the same definitions as stated in Section 2.2 for feedback channel block code, probability of error, achievability and operational capacity with feedback (supremum of all achievable feedback code rates). We denote the capacity of the channel with feedback by C_{FB}.

Note again, that because of the feedback, X^n and Z^n are no longer independent; X_i may depend on Z^{i-1}.

We now state the key result (Theorem 4) of [17] which is a new converse approach based on a simple new lower bound on the error probability of an arbitrary channel code as a function of its size.

Lemma 2 Let (n, M, ϵ) represent a channel block code with blocklength n, M codewords and error probability ϵ. Then every (n, M, ϵ) code satisfies

$$\epsilon \geq P\left[\frac{1}{n}\, i_{X^nY^n}(X^n; Y^n) \leq \frac{1}{n}\log_2 M - \gamma\right] - \exp(-\gamma n) \tag{37}$$

for every $\gamma > 0$, where X^n places probability mass $1/M$ on each codeword.

We now obtain the following result:

Proposition 4 Feedback does not increase the capacity of discrete channels with *arbitrary* additive noise:

$$C_{FB} = C_{NFB} = \log_2(q) - \overline{H}(Z^n) \tag{38}$$

Proof 4 We start by noting that the result given in Lemma 2 still holds if we replace the input vector X^n by the message random variable V where V is uniform over the set of messages $\{1, 2, \ldots, M\}$. That is, every (n, M, ϵ) feedback code satisfies

$$\epsilon \geq P\left[\frac{1}{n}\, i_{VY^n}(V; Y^n) \leq \frac{1}{n}\log_2 M - \gamma\right] - \exp(-\gamma n) \tag{39}$$

for every $\gamma > 0$, where V is uniform over $\{1, 2, \ldots, M\}$.

We refer to the sequence (n, M, ϵ_n) of feedback codes with vanishingly small error probability (i.e., $\epsilon_n \to 0$ as $n \to \infty$) as a *reliable feedback code sequence.*

Using equation (39), we first show that

$$C_{FB} \leq \underline{I}(V; Y^n) \tag{40}$$

We prove (40) by contradiction. Assume that for some $\rho > 0$,

$$C_{FB} = \underline{I}(V; Y^n) + 3\rho \tag{41}$$

By definition of capacity, there exists a reliable feedback code sequence with rate

$$R = \frac{1}{n}\log_2 M > C_{FB} - \rho \tag{42}$$

Now using (39) (with $\gamma = \rho$) along with (41) and (42), we obtain that the error probability of the sequence (n, M, ϵ_n) of feedback codes must be lower bounded by

$$\epsilon_n \;\geq\; P\left[\frac{1}{n}\, i_{VY^n}(V;Y^n) \leq \underline{I}(V;Y^n) + \rho\right] \;-\; \exp\left(-\rho n\right) \tag{43}$$

However by definition of $\underline{I}(V;Y^n)$ the probability in the right-hand side of (43) cannot vanish asymptotically; therefore contradicting the fact that $\epsilon_n \to 0$ as $n \to \infty$. Thus (40) is proved.

Now using the properties of the inf-information rate in [17], we can write

$$\begin{aligned} \underline{I}(V;Y^n) \;&\leq\; \overline{H}(Y^n) - \overline{H}(Y^n|V) \\ &\leq\; \log_2(q) - \overline{H}(Y^n|V) \end{aligned} \tag{44}$$

The conditional sup-entropy rate $\overline{H}(Y^n|V)$ is the limsup in probability (according to P_{VY^n}) of $\frac{1}{n}\log_2 \frac{1}{P_{Y^n|V}(Y^n|V)}$. That is $\overline{H}(Y^n|V)$ is the infimum of all reals β such that $Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{Y^n|V}(Y^n|V)} \geq \beta\right\} \longrightarrow 0$, as $n \longrightarrow \infty$. But we can write

$$Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{Y^n|V}(Y^n|V)} \geq \beta\right\} = \sum_v P(V=v) \sum_{y^n : P(Y^n=y^n|V=v) \leq 2^{-n\beta}} P(Y^n = y^n|V=v).$$

Now, letting $f_i \stackrel{\text{def}}{=} f_i(v, y^{i-1})$ and $f^i \stackrel{\text{def}}{=} [f_1(v), f_2(v,y_1), \ldots, f_i(v,y^{i-1})] = [f_1, f_2, \ldots, f_i]$, we have

$$\begin{aligned} P(Y^n = y^n|V=v) = \;&\prod_{i=1}^{n} P(Y_i = y_i|Y^{i-1} = y^{i-1}, V=v) \\ = \;&\prod_{i=1}^{n} P(X_i \oplus Z_i = y_i|Y^{i-1} = y^{i-1}, V=v, X_i = f_i) \end{aligned} \tag{45}$$

$$= \prod_{i=1}^{n} P(Z_i = y_i \oplus f_i | Y^{i-1} = y^{i-1}, V = v, X_i = f_i) \tag{46}$$

$$= \prod_{i=1}^{n} P(Z_i = y_i \oplus f_i | Y^{i-1} = y^{i-1}, V = v, X^i = f^i, Z^{i-1} = y^{i-1} \oplus f^{i-1}) \tag{47}$$

$$= \prod_{i=1}^{n} P(Z_i = y_i \oplus f_i | Z^{i-1} = y^{i-1} \oplus f^{i-1}) \tag{48}$$

$$= P(Z^n = y^n \oplus f^n) \tag{49}$$

Here,

- Equation (45) follows from the fact that $X_i = f_i(V, Y_1, \ldots, Y_{i-1})$ due to feedback.
- Equation (46) holds since $P(Z + X = y | X = x) = P(Z = y - x | X = x)$.
- Equation (47) follows from the fact that given V and Y^{i-1}, we know all the previous transmitted letters $X_1, X_2, \ldots, X_{i-1}$ and thus we can recover all the previous noise letters $Z_j = Y_j - X_j \pmod q$ for $j = 1, 2, \ldots, i-1$.
- Equation (48) follows from the fact that Z_i and (V, Y^{i-1}, X^i) are conditionally independent given Z^{i-1}.

Hence,

$$\begin{aligned} Pr\left\{\frac{1}{n}\log_2 \frac{1}{P_{Y^n|V}(Y^n|V)} \geq \beta\right\} &= \sum_v P(V = v) \sum_{y^n : P(Z^n = y^n \oplus f^n) \leq 2^{-n\beta}} P(Z^n = y^n \oplus f^n) \\ &= \sum_v P(V = v) \sum_{z^n : P(Z^n = z^n) \leq 2^{-n\beta}} P(Z^n = z^n) \\ &= \sum_{z^n : P(Z^n = z^n) \leq 2^{-n\beta}} P(Z^n = z^n) \end{aligned}$$

Therefore we obtain that

$$\overline{H}(Y^n|V) = \overline{H}(Z^n) \tag{50}$$

Thus from (40), (44) and (50) we conclude that

$$C_{FB} \leq \log_2(q) - \overline{H}(Z^n) = C_{NFB} \tag{51}$$

But by definition of a feedback code, $C_{FB} \geq C_{NFB}$ since a non-feedback code is a special case of a feedback code. Thus we get:

$$C_{FB} = C_{NFB} = \log_2(q) \; - \overline{H}(Z^n) \tag{52}$$

■

Observation: Note that if the noise process is stationary, then its sup-entropy rate is equal to the supremum over the entropies of almost every ergodic component of the stationary noise. If the noise process is stationary ergodic, then its sup-entropy rate is equal to the entropy rate of the noise [17].

5 Conclusions

In this paper, we considered a discrete additive noise channel with output feedback. We showed that the capacity of the channel without feedback equals its capacity with feedback. This was first shown for a stationary ergodic and non-ergodic additive noise process. We then generalized the result for discrete channels with arbitrary additive noise.

In [4], we introduce the notion of *symmetric* channels with memory. These channels are obtained by combining an input process with an arbitrary noise process that is independent of the input. These channels have the property that their inf-information rate is maximized when the input process is an equally likely iid process. We show that feedback does not also increase the capacity of these channels. Additive noise channels belong to the class of symmetric channels.

We are currently investigating the effect of feedback on the capacity of additive noise channels that are subject to average cost constraints on their input sequences [5].

6 Acknowledgment

The author wishes to thank Professors Imre Csiszàr and Tom Fuja for their very valuable advice and constructive criticism.

References

[1] R. Ahlswede, "The Weak Capacity of Averaged Channels", *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 11, pp. 61-73, 1968.

[2] R. Ahlswede, "Certain Results in Coding Theory for Compound Channels I", *Proceedings Bolyai Colloquium on Information Theory*, Debrecen, Hungary, pp. 35-60, 1967.

[3] F. Alajaji and T. Fuja, "A Communication Channel Modeled by the Spread of Disease", *Proceedings of the IEEE International Symposium on Information Theory*, San Antonio, TX, January 1993.

[4] F. Alajaji and T. Fuja, "Effect of Feedback on the Capacity of Discrete Additive Channels with Memory", submitted to the *IEEE International Symposium on Information Theory*, Trondheim, Norway, June 1994.

[5] F. Alajaji and T. Fuja, "Feedback Capacity-Cost Function of Discrete Additive Markov Channels", submitted to *Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, March 1994.

[6] D. Blackwell, L. Breiman and A. J. Thomasian, "The Capacity of a Class of Channels", *Annals Math. Stat.*, Vol. 30, pp. 1229-1241, 1959.

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

[8] E. N. Gilbert, "Capacity of Burst-Noise Channels", *Bell Syst. Tech. Journal*, Vol. 39, pp. 1253-1265, 1960.

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

[10] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*, Springer-Verlag New York Inc., 1988.

[11] R. Gray and L. D. Davisson, "The Ergodic Decomposition of Stationary Discrete Random Processes", *IEEE Transactions on Information Theory*, Vol. 20, No. 5, pp. 625-636, 1974.

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

[13] J. C. Kieffer, "A General Formula for the Capacity of Stationary Nonanticipatory Channels", *Information and Control*, Vol. 26, pp. 381-391, 1974.

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

[15] K. R. Parthasarathy, "Effective Entropy Rate and Transmission of Information Through Channels with Additive Random Noise", *Sankhya*, Vol. A(25), pp. 75-84, 1963.

[16] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel", *IRE Transactions on Information Theory*, Vol. 2, pp. 8-19, 1956.

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