FEEDBACK CAPACITY OF DISCRETE ADDITIVE CHANNELS*

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

ABSTRACT

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

1. INTRODUCTION

We consider discrete channels with stationary additive noise. Note that such channels need not be memoryless; in general, they have memory. The Gilbert burst-noise channel [1], as well as the Polya-contagion channel [2], 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 channels.

In earlier related work, Shannon [3] 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 [4] 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's been shown [4] 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 alphabet $A = \{0, 1, \ldots, q-1\}$, 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 as 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) [5]. 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 [5] for more rigorous definitions of causal and historyless channels. We specify these conditions so as to be able to use well-established formulas [5,6] for the non-feedback capacity of the resulting channels.

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

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

and a decoder

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

The encoder represents the message $W \in \{1, 2, \ldots, 2^{nR}\}$ with the codeword $f(W) = X^n = [X_1, X_2, \ldots, X_n]$ 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 $\hat{W} = g(Y^n)$. A decoding error occurs if $\hat{W} \neq W$.

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

$$\begin{aligned} P_e^{(n)} &= \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} Pr\{g(Y^n) \neq W | W = k\} \\ &= Pr\{g(Y^n) \neq W\} \end{aligned}$$

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

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

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

We denote the capacity of the channel with no feedback by C_{NFB}. The objective, of course, is to transmit an arbitrary message W at a high rate and low probability of error. If we define C_{NFB} to be the supremum of all admissable code rates, then C_{NFB} is the *capacity* of the channel.

Because the channel is non-anticipatory and stationary ergodic, the nonfeedback capacity C_{NFB} of this channel is known and is equal to [5]:

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

where

$$
\begin{aligned}
X^n &= (X_1, X_2, \ldots, X_n),\\
Y^n &= (Y_1, Y_2, \ldots, Y_n),\\
Z^n &= (Z_1, Z_2, \ldots, Z_n),
\end{aligned}
$$

$I(X^n;Y^n)$ 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.

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

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

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

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

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

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

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

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

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

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

$$H(W|Y^n) \le 1 + P_e^{(n)} nR \qquad (4)$$

We then have:

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

where R is any admissable rate.

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

$$C_{FB} \le \lim_{n\to\infty} \frac{1}{n} I(W;Y^n) \qquad (5)$$

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

$$I(W;Y^n) = \sum_{i=1}^{n} I(W;Y_i|Y^{i-1}) \qquad (6)$$

but

$$
\begin{aligned}
I(W;Y_i|Y^{i-1}) &= H(Y_i|Y^{i-1}) - H(Y_i|W, Y^{i-1}) & (7)\\
&= H(Y_i|Y^{i-1}) - H(X_i \oplus Z_i|W, Y^{i-1}) & (8)
\end{aligned}
$$

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

$$
\begin{aligned}
H(X_i \oplus Z_i|W, Y^{i-1}) &= H(Z_i|W, Y^{i-1}, X_i) & (9)\\
&= H(Z_i|W, Y^{i-1}, X^i, Z^{i-1}) & (10)\\
&= H(Z_i|Z^{i-1}). & (11)
\end{aligned}
$$

Here,

- Equation (9) follows from the fact that given W and Y^{i-1}, X_i is known deterministically and $H(Z+X|X) = H(Z|X)$.
- Equations (10) follows from the fact that given W 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 (W, Y^{i-1}, X^i) are conditionally independent given Z^{i-1}.

Therefore

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

and

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

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 noise entropy (given in equations (9)-(11)) remain the same *with or without feedback*. This is because addition is invertible; in general $H(X) \leq H(f(X))$ 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 [4] is because for continuous channels we have power constraints on the input, which upon optimization may increases $\lim_{n\to\infty} \frac{1}{n}\, H(Y^n)$ when feedback is used; while for discrete channels this quantity is upperbounded by $\log_2(q)$ and cannot be increased with feedback. 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) [7]. 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 be AMS [8].

Finally, with the result of Proposition 1 in mind, it would be interesting to investigate discrete *non-additive* stationary ergodic 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 with *additive* stationary *ergodic* components.

An averaged channel with stationary ergodic components is defined as follows:

Consider a family of stationary ergodic 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. $W_\theta^{(n)}()$ is the block transition probabilities of the stationary ergodic channels, conditioned on a parameter $\theta \in \Theta$.

Definition 1 We define a channel to be an "averaged" communication channel with stationary ergodic 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 $\{W_\theta^{(n)}(Y^n = y^n|X^n = x^n)\}$ taken with respect to some distribution on θ – i.e., if it's of the form:

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

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 functions as follows: among the (countable or uncountable) stationary ergodic 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 transition probability of the noise) as a mixture of the probabilities of the additive stationary ergodic channels (proposition 2), we need to state first the ergodic decomposition theorem for stationary processes [9].

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), 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 additive stationary ergodic 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 stationary ergodic channels being an additive noise channel:

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

■

The strong capacity of averaged channels does not exist [10], since the strong converse to the channel coding theorem does not hold. However it was shown by Ahlswede [10] that the weak converse holds for these channels. Recalling from Section 2.1 the definitions of a channel block code and the operational (weak) capacity of the channel (the supremum of all achievable rates), we state the formula for the non-feedback operational capacity of an averaged channel [6],[10]:

Lemma 2 (Capacity of Averaged Channel) Consider the averaged channel with stationary ergodic components described by (17), with common input and output alphabet A; input probability space $(A^n, \sigma(A^n), Q^n)$ and general averaging probability distribution G() -i.e. Θ can be either a discrete or continuous parameter space.

Then the non-feedback (weak) capacity of the averaged channel is given by

$$C_{NFB}^{(ac)} = \lim_{\alpha \to 0} C(\alpha) \tag{20}$$

where

$$C(\alpha) = \max_Q \sup_{\{E \in \sigma(\Theta):\ G(E) \geq 1-\alpha\}} \inf_{\theta \in E} i(Q; W_\theta) \tag{21}$$

where the mutual information rate $i(Q; W_\theta)$ is given by

$$i(Q; W_\theta) = \lim_{n \to \infty} \frac{1}{n} I(Q^{(n)}; W_\theta^{(n)})$$

with

$$I(Q^{(n)}; W_\theta^{(n)}) = \sum_{x^n, y^n \in A^n} W_\theta^{(n)}(y^n | x^n) Q^{(n)}(x^n) \log_2 \frac{W_\theta^{(n)}(y^n|x^n)}{q^{(n)}(y^n)}$$

and $q^{(n)}(y^n) \triangleq \sum_{\tilde{x}^n \in A^n} W_\theta^{(n)}(y^n|\tilde{x}^n)\, Q^{(n)}(\tilde{x}^n)$.

Non-Feedback Capacity of the Channel with Additive Noise: As we mentioned earlier, the channel with additive stationary non-ergodic noise is an averaged channel with *additive* stationary *ergodic* components (Proposition 2). Since the channel has an additive noise process that is independent of the input process, we will have that the maximization over the input distribution Q in equation (21) is realized for uniform input distribution (symmetry property). We can therefore interchange the max and the inf in (21) and we get:

$$\max_Q i(Q; W_\theta) = \log(q) - h(W_\theta)$$

The resulting non-feedback capacity of the channel with additive non-ergodic noise is:

$$C_{NFB} = \log_2(q) - \text{ess}_\Theta \sup \; h(W_\theta) \tag{22}$$

where

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

$$h(W_\theta) \stackrel{\triangle}{=} \lim_{n\to\infty} \frac{1}{n} H_n(W_\theta^{(n)}) \tag{23}$$

  with

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

- and the essential supremum is defined by

$$\text{ess}_\Theta \sup \; f(\theta) \stackrel{\triangle}{=} \inf \; [r : dG(f(\theta) \leq r) = 1] \tag{24}$$

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 W 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 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; i.e. for any $R < C_{NFB}$, there exists feedback codes with blocklength n and rate R, such that $\lim_{n\to\infty} \; P_e^{(n)} = 0$.

The additive channel is a mixture of *additive* stationary ergodic channels, thus by Proposition 1, we obtain that for each of these components: $C_{FB}^{(\theta)} = C_{NFB}^{(\theta)}$. Now, examining equation (22), 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 θ's is $\delta > 0$.

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

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

$$P_e^{(n)} \;\; = \sum_{\{y^n : g(y^n) \neq k\}} W^{(n)}(Y^n = y^n | X^n = x_k^n) \tag{25}$$

where x_k^n is the feedback codeword of length n that corresponds to the message $W = k \in \{1, 2, \ldots, 2^{nR}\}$. And using Proposition 2, we can write:

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

where

$$P_e^{(n)}(\theta) = \sum_{\{y^n : g(y^n) \neq k\}} W_\theta^{(n)}(Z^n = y^n - x_k^n \; (mod \; q))$$

Thus we have:

$$\begin{aligned} P_e^{(n)} &= \int_\Theta P_e^{(n)}(\theta) \; dG(\theta) && (27) \\ &\geq \int_{\{\theta \in \Theta : C_{NFB}^{(\theta)} < C_{NFB} + \epsilon\}} P_e^{(n)}(\theta) \; dG(\theta) && (28) \end{aligned}$$

We now recall the weak converse to the channel coding theorem for stationary ergodic channels: 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. This is shown by using Fano's inequality. Note that γ depends only on ϵ' and is independent of the characteristics of the channel.

Therefore, applying the weak converse of the coding theorem for stationary ergodic channels, we get that for $R > C_{NFB} + 2\epsilon > C_{NFB}^{(\theta)} + \epsilon$, 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.

Then

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

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 [10]:

$$C_{NFB}^{(ac)} = \max_{Q^{(1)}} \inf_{\theta\in\Theta} I(Q^{(1)}; W_\theta^{(1)}) \tag{29}$$

Note that:

$$\begin{aligned} C_{NFB}^{(ac)} &\leq \inf_{\theta\in\Theta} \max_{Q^{(1)}} I(Q^{(1)}; W_\theta^{(1)}) \qquad (30) \\ &= \inf_{\theta\in\Theta} C^{(\theta)} \end{aligned}$$

where $C^{(\theta)} = \max_{Q^{(1)}} I(Q^{(1)}; W_\theta^{(1)})$ 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:

$$C_{FB}^{(ac)} = \inf_{\theta\in\Theta} C^{(\theta)} \tag{31}$$

Therefore $C_{FB}^{(ac)} \geq C_{NFB}^{(ac)}$. This result (equation (31)) is equivalent to the result already derived by Ahlswede for the discrete averaged channel with sender informed [11].

Finally, in the case for which the inequality in (30) holds with the *strict* inequality, we obtain that feedback *increases* capacity: $C_{FB}^{(ac)} > C_{NFB}^{(ac)}$. Refer to [12] for an example of a finite collection of DMC's for which (30) holds with the strict inequality.

4. ACKNOWLEDGMENT

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

5. REFERENCES

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

[2] 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.

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

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

[5] 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.

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

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

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

[9] 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.

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

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

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