On the Feedback Capacity of a Discrete Non-Binary Noise Channel with Memory

Nevroz Şen, Fady Alajaji and Serdar Yüksel
Department of Mathematics and Statistics
Queen's University
Kingston, ON K7L 3N6, Canada
Email: nsen, fady, yuksel@mast.queensu.ca

***Abstract*— We study the feedback capacity of a discrete binary-input non-binary output channel with memory recently introduced in [5] to model soft-decision demodulated time-correlated fading channels. The channel, whose output process can be explicitly expressed in terms of its binary input process and a non-binary noise process, encompasses modulo-additive noise binary channels as a special case (realized when hard-decision demodulation is used on the underlying fading channel). We show that, even though the channel has memory, feedback does not increase its capacity when the noise process is stationary ergodic. We also note the validity of the result for arbitrary noise processes.**

I. INTRODUCTION

We investigate the feedback capacity of a discrete binary-input 2^q-ary output communication channel, which was recently proposed in [5] to model soft-decision demodulated fading channels with memory. The channel, which we refer to by the non-binary noise discrete channel (NBNDC), is explicitly described in terms of a non-binary noise process that is independent of its input. We show that, in spite of the NBNDC's memory structure, feedback does not help increase its capacity. Although the result is proved under the assumption of stationary ergodic non-binary noise, we remark (without proof) that it also holds for arbitrary (not necessarily stationary ergodic) noise.

This result generalizes in some sense the work in [1], where it is also shown that feedback does not increase capacity for discrete modulo-k additive channels with arbitrary noise with memory. Furthermore, when $q = 1$, the result intersects exactly with the result for $k = 2$ in [1], since the NBNDC reduces to the modulo-2 additive noise channel.

Let us briefly explain the proof used here for the NBNDC in relation to the proof used in [1]. Note that, although both the NBNDC and the modulo-k additive channel of [1] are symmetric in the sense of [4] (this is observed in [5] for the NBNDC) and thus have the property that their non-feedback capacity is achieved by a memoryless uniformly distributed input, the modulo-k channel is "strongly" symmetric in the sense that a memoryless uniformly distributed input yields a memoryless uniformly distributed output; this is not the case for the NBNDC (with $q > 1$). As noted in [1], such output uniformity property is key to showing that feedback does not increase the capacity of the modulo-k channel. For the NBNDC, as the later property does not hold, our proof is based on an intermediate result (see Lemma 1), which states that under feedback, the NBNDC conditional entropy of any output symbol given previous outputs is maximized by a uniform feedback policy.

In previous works, Shannon [7] first showed that feedback does not increase the capacity of discrete memoryless channels. Later, Cover and Pombra [2] considered additive channels with colored Gaussian noise and proved that feedback can either increase their capacity by at most half a bit or at most double it (the later result is originally due to Pinsker [6] and Ebert [3]).

The remainder of the paper is organized as follows. We introduce the NBNDC model in Section II and review its non-feedback capacity in Section III. In Section IV, we investigate the channel's feedback capacity and show that it is identical to its non-feedback capacity when its non-binary noise is a stationary ergodic process. We note that the result also holds for arbitrary noise processes. Finally, we conclude in Section V.

Throughout, random variables will be denoted by upper case letters (X) and their particular realizations by lower case letters (x). Also, we will write the n-tuples $(X_1, X_2, \cdots, X_n)$ and $(x_1, x_2, \cdots, x_n)$ as X^n and x^n, respectively.

II. CHANNEL MODEL

The NBNDC is a discrete binary-input 2^q-ary output communication channel with memory introduced in [5] with the objective of capturing both the statistical memory and the soft-decision information of time-correlated fading channels modulated via binary phase-shift keying (BPSK) and coherently demodulated with an output quantizer of resolution q. Given that the channel has a straightforward structure and useful "invertibility" properties, it can help in the design of coding/decoding schemes for soft-decision demodulated channels with memory that result in superior performance over coding systems that ignore the channel's memory (via interleaving) and/or soft-decision information (via hard demodulation) [5]. Additionally, receivers operating with 1-3 bit quantization have potential applications in wireless communications.

The NBNDC model, which is equivalent to a discrete fading channel composed of a BPSK modulator, a time-correlated flat

This work was supported in part by NSERC of Canada.

fading channel and a q-bit soft-quantized coherent demodulator [5], is explicitly described by the following equation

$$Y_k = (2^q - 1)X_k + (-1)^{X_k} Z_k \quad (1)$$

for $k = 1, 2, \cdots$, where X_k is the input process, Y_k is the output process and Z_k is the noise process. Here the input $X_k \in \mathcal{X} = \{0, 1\}$ is binary, and both noise and output symbols, Z_k and Y_k, take values from the same 2^q-ary alphabet given by $\mathcal{Z} = \mathcal{Y} = \{0, 1, \cdots, 2^q - 1\}$. It is also assumed that the noise and input processes are independent from each other. Note that under hard demodulation, i.e., for $q = 1$, equation (1) reduces to the familiar expression

$$Y_k = X_k \oplus Z_k$$

where $\oplus$ denotes modulo-2 addition. In other words, when $q = 1$, the NBNDC reduces to the binary (modulo-2) additive noise channel.

III. CAPACITY WITHOUT FEEDBACK

Consider the NBNDC given by (1), where the noise process is stationary ergodic. For this information stable channel, the non-feedback capacity is given by [8], [5]

$$C_{NFB} = \lim_{n\to\infty} C^{(n)} \quad (2)$$

where

$$C^{(n)} = \max_{p(x^n)} \frac{1}{n} I(X^n; Y^n),$$

where the maximum is taken with respect to all input distributions $p(x^n)$ and $I(X^n; Y^n)$ is the block mutual information between the channel input tuple X^n and the channel output tuple Y^n. Using (1) and the fact that $\{X_k\}$ and $\{Z_k\}$ are independent from each other, the block mutual information can be rewritten as

$$I(X^n; Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - H(Z^n)$$

Therefore

$$C^{(n)} = \frac{1}{n}\left(\max_{p(x^n)}[H(Y^n)] - H(Z^n)\right). \quad (3)$$

We now point out that the NBNDC, as described by $Y_k = f(X_k, Z_k)$ where $f(\cdot,\cdot)$ is given in (1), satisfies the following "invertibility" properties:

(a) For any fixed input $x \in \mathcal{X}$, $f(x, \cdot) : \mathcal{Z} \to \mathcal{Y}$ is invertible.

(b) Every output symbol is the image of exactly two distinct input-noise pairs; i.e., for any $y \in \mathcal{Y}$, there are exactly two pairs (x_1, z_1) and (x_2, z_2) in $\mathcal{X} \times \mathcal{Z}$ such that $x_1 \neq x_2$, $z_1 \neq z_2$ and $y = f(x_1, z_1) = f(x_2, z_2)$.

Let $\mathbf{Q}^n = [P^n(y^n|x^n)]$ denote the $2^n \times 2^{qn}$ transition probability matrix of the NBNDC, where each row is represented by x^n and each column is represented by y^n. It can be shown using the channel's above properties that $\mathbf{Q}^n$ can be partitioned along its columns into $2^{(q-1)n}$ arrays, where each array is a $2^n \times 2^n$ matrix and has the property that each of its rows (respectively, columns) is a permutation of its other rows (respectively, columns). Thus the channel is symmetric in the sense of Gallager [4, p. 94] and its block mutual information $I(X^n; Y^n)$ is maximized by a uniformly distributed (equally likely) input X^n. Under a uniform input distribution, the value of $H(Y^n)$ is also maximized, satisfying

$$\max_{p(x^n)}[H(Y^n)] = n + H(W^n) \quad (4)$$

where $\{W_k\}$ is a process defined on the alphabet $\mathcal{W} = \{0, 1, \cdots, 2^{q-1} - 1\}$ with n-fold probability distribution

$$Pr(W^n = w^n) = \sum_{x^n \in \mathcal{X}^n} Pr\left(Z^n = \frac{w^n - (2^q-1)x^n}{(-1)^{x^n}}\right) \quad (5)$$

where $Z^n = (w^n - (2^q - 1)x^n)/(-1)^{x^n}$ denotes the n-tuple obtained from component-wise operations, i.e., the n-tuple $(Z_1 = (w_1 - (2^q - 1)x_1)/(-1)^{x_1}, Z_2 = (w_2 - (2^q - 1)x_2)/(-1)^{x_2}, \cdots, Z_n = (w_n - (2^q - 1)x_n)/(-1)^{x_n})$.

Finally, substituting (4) into (3) yields that

$$C^{(n)} = 1 + \frac{1}{n}[H(W^n) - H(Z^n)] \quad (6)$$

and the channel capacity is thus given by

$$\begin{aligned} C_{NFB} &= \lim_{n\to\infty} C^{(n)} \\ &= 1 + \lim_{n\to\infty}\frac{1}{n}[H(W^n) - H(Z^n)] \quad (7) \\ &= 1 + H(\mathcal{W}) - H(\mathcal{Z}) \quad (8) \end{aligned}$$

where $H(\mathcal{W})$ and $H(\mathcal{Z})$ denote the entropy rates of $\{W_n\}$ and $\{Z_n\}$, respectively. A more detailed derivation of the above non-feedback capacity of the NBNDC is available in [5].

IV. CAPACITY WITH FEEDBACK

In this section, we will show that feedback does not increase the capacity of the NBNDC. Without loss of generality, we assume that $q \geq 2$, since for $q = 1$, the NBNDC reduces to the modulo-2 additive noise channel and hence the result trivially holds from [1]. By feedback, we mean that there exists a channel from the receiver to the transmitter which is noiseless, delayless and has large capacity. Thus at any given time, all previously received outputs are unambiguously known by the transmitter and can be used for encoding the message into the next code symbol.

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

$$\psi_i : \{1, 2, ..., 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}$$

for $i = 1, 2, ...n$ and an associated decoding function

$$\phi : \mathcal{Y}^n \to \{1, 2, ..., 2^{nR}\}.$$

Thus when the transmitter wants to send a message, say $V \in \{1, 2, ..., 2^{nR}\}$, it sends the codeword X^n, where $X_1 = \psi_1(V)$ and $X_i = \psi_i(V, Y_1, \cdots, Y_{i-1})$, for $i = 2, \cdots, n$. For a received Y^n at the channel output, the receiver uses the decoding function to estimate the transmitted message as $\hat{V} = \phi(Y^n)$. A decoding error is made when $\hat{V} \neq V$.

We assume that the message V is uniformly distributed over $\{1, 2, ..., 2^{nR}\}$. Therefore, the probability of error is given by

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} P\left\{\phi(Y^n) \neq V | V = k\right\} = P\left\{\phi(Y^n) \neq V\right\}.$$

The capacity with feedback, C_{FB}, is the supremum of all admissible feedback code rates (i.e., all rates for which there exists sequences of feedback codes with asymptotically vanishing probability of error).

From Fano's inequality, we have

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

where the first inequality holds since $h_b(P_e^{(n)}) \leq 1$, where $h_b(\cdot)$ is the binary entropy function. We also know that

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

where R is any admissible rate. Dividing both sides above by n and taking the limit yields

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

where the supremum is taken over all feedback policies $\left\{P(x_i|x^{i-1}, y^{i-1})\right\}_{i=1}^n$. We can write $I(V;Y^n)$ as follows

$$\begin{aligned} I(V;Y^n) = & \sum_{i=1}^n I(V;Y_i|Y^{i-1}) \\ = & \sum_{i=1}^n \left(H(Y_i|Y^{i-1}) - H(Y_i|V, Y^{i-1})\right) \\ = & \sum_{i=1}^n \left(H(Y_i|Y^{i-1}) - H(Y_i|V, Y^{i-1}, X_i, X^{i-1})\right) \end{aligned}$$

where the last equality follows from the fact that $X_k = \psi_k(V, Y_1, Y_2, \ldots, Y_{k-1})$ for $k = 1, \cdots, i$. We also can write

$$\begin{aligned} & H(Y_i|V, Y^{i-1}, X_i, X^{i-1}) \\ &\quad = H(f(X_i, Z_i)|V, Y^{i-1}, X_i, X^{i-1}) \\ &\quad = H(Z_i|V, Y^{i-1}, X_i, X^{i-1}) \\ &\quad = H(Z_i|V, Y^{i-1}, X_i, X^{i-1}, Z^{i-1}) \\ &\quad = H(Z_i|Z^{i-1}) \end{aligned}$$

where the second and third equalities follow from channel property (a) and the last equality holds since Z_i and (V, X_i, Y^{i-1}) are conditionally independent given Z^{i-1}. Therefore, we get that

$$\begin{aligned} I(V;Y^n) &= \sum_{i=1}^n I(V;Y_i|Y^{i-1}) \\ &= \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Z_i|Z^{i-1})\right]. \quad (9) \end{aligned}$$

We next prove that all of the output conditional entropies $H(Y^i|Y^{i-1})$ in (9) are maximized by uniform conditional input distributions $P(X_i|X^{i-1}, Y^{i-1})$ (feedback policies). With this result in hand, we can then directly deduce that feedback does not increase the capacity of the NBNDC as the right hand side of (9) will equal C_{NFB} after normalizing by n and taking the limit.

Lemma 1: For a general noise process $\{Z_k\}$, each conditional output entropies $H(Y_i|Y^{i-1})$, $i = 1, \cdots, n$ in (9) is maximized by a uniform feedback policy:

$$P(X_i = a|X^{i-1} = x^{i-1}, Y^{i-1} = y^{i-1}) = \frac{1}{2}$$

for all $a \in \{0, 1\}$, $x^{i-1} \in \{0, 1\}^{i-1}$ and $y^{i-1} \in \mathcal{Y}^{i-1}$.

Proof: Let us first write the output conditional entropy $H(Y_i|Y^{i-1})$ as

$$H(Y_i|Y^{i-1}) = \sum_{y^{i-1}} P(y^{i-1}) H(Y_i|Y^{i-1} = y^{i-1}) \quad (10)$$

where

$$H(Y_i|Y^{i-1} = y^{i-1}) = -\sum_{y_i} P(y_i|y^{i-1}) \log P(y_i|y^{i-1}). \quad (11)$$

To show that $H(Y_i|Y^{i-1})$ in (10) is maximized by a uniform feedback policy, it is enough to show that such a uniform policy maximizes each of the $H(Y_i|Y^{i-1} = y^{i-1})$ terms.

We now expand $P(y_i|y^{i-1})$ as follows

$$\begin{aligned} & \sum_{x_i} \sum_{x^{i-1}} \sum_{z_i} \sum_{z^{i-1}} P(y_i, x_i, z_i, x^{i-1}, z^{i-1}|y^{i-1}) \\ &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i, x^{i-1}, z^{i-1}, y^{i-1}) \\ &\quad P(x_i, z_i, x^{i-1}, z^{i-1}|y^{i-1}) \quad (12) \\ &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i) \\ &\quad P(x_i, z_i, x^{i-1}, z^{i-1}|y^{i-1}) \quad (13) \\ &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i) P(x_i, x^{i-1}, z^{i-1}|y^{i-1}) \\ &\quad P(z_i|x_i, x^{i-1}, z^{i-1}, y^{i-1}) \quad (14) \\ &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i) P(z_i|z^{i-1}) \\ &\quad P(x_i|x^{i-1}, z^{i-1}, y^{i-1}) P(x^{i-1}, z^{i-1}|y^{i-1}) \quad (15) \\ &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i) P(x_i|x^{i-1}, y^{i-1}) \\ &\quad P(z_i|z^{i-1}) P(x^{i-1}, z^{i-1}|y^{i-1}). \quad (16) \end{aligned}$$

Thus

$$\begin{aligned} P(y_i|y^{i-1}) &= \sum_{x_i} \cdots \sum_{z^{i-1}} P(y_i|x_i, z_i) P(z_i|z^{i-1}) \\ &\quad P(x_i|x^{i-1}, y^{i-1}) P(x^{i-1}, z^{i-1}|y^{i-1}). \quad (17) \end{aligned}$$

Setting $P(X_i = 0|x^{i-1}, y^{i-1}) = p$ and applying channel properties (a) and (b) to (17) yields

$$
\begin{aligned}
&P(Y_i = j|y^{i-1}) + P(Y_i = 2^q - 1 - j|y^{i-1}) \\
&= \sum_{x^{i-1}, z^{i-1}} P(Z_i = j|z^{i-1}) P(x^{i-1}, z^{i-1}|y^{i-1}) \left(p + (1-p)\right) \\
&\quad + \sum_{x^{i-1}, z^{i-1}} P(Z_i = 2^q - 1 - j|z^{i-1}) P(x^{i-1}, z^{i-1}|y^{i-1}) \\
&\qquad \times \left(p + (1-p)\right) \\
&= \sum_{x^{i-1}, z^{i-1}} \left[P(Z_i = j|z^{i-1}) + P(Z_i = 2^q - 1 - j|z^{i-1})\right] \\
&\qquad \times P(x^{i-1}, z^{i-1}|y^{i-1}) \\
&\triangleq k_j \qquad (18)
\end{aligned}
$$

for $j = 0, 1, \cdots, 2^{q-1} - 1$. It should be noted that each k_j in (18) is independent of the feedback policy $P(X_i = 0|x^{i-1}, y^{i-1})$. Using (18), we can write (11) as

$$
\begin{aligned}
H(Y_i|Y^{i-1} = y^{i-1}) \;=\; -\sum_{j=0}^{2^{q-1}-1} &\left[a_j \log a_j \right. \\
&\left. + (k_j - a_j) \log(k_j - a_j)\right] \qquad (19)
\end{aligned}
$$

where

$$a_j = P(Y_i = j|y^{i-1})$$

and

$$k_j - a_j = P(Y_i = 2^q - 1 - j|y^{i-1}).$$

Applying the log-sum inequality on each summand (within brackets) in (19) yields that

$$H(Y_i|Y^{i-1} = y^{i-1}) \leq -\sum_{j=0}^{2^{q-1}-1} k_j \log(k_j/2) \qquad (20)$$

with equality iff $a_j = k_j - a_j$ for $j = 0, 1, ..., 2^{q-1} - 1$. In other words, $H(Y_i|Y^{i-1} = y^{i-1})$ is maximized iff

$$P(Y_i = j|y^{i-1}) = P(Y_i = 2^q - 1 - j|y^{i-1}). \qquad (21)$$

From (17) and using the channel's properties, it can be shown that (21) is satisfied when

$$P(X_i = 0|x^{i-1}, y^{i-1}) = P(X_i = 1|x^{i-1}, y^{i-1}) = \frac{1}{2}. \qquad (22)$$

Hence a uniform feedback policy maximizes the conditional entropy $H(Y_i|Y^{i-1} = y^{i-1})$ for each y^{i-1}; this completes the proof. ■

Lemma 1 directly implies that a uniform feedback policy yields a uniformly distributed input X^n and maximizes the channel's output block entropy $H(Y^n)$, resulting in $H(Y^n) = n + H(W^n)$ as in (4). Substituting the later in (9), normalizing by n and taking the limit yield that

$$C_{FB} \;\leq\; 1 + H(\mathcal{W}) - H(\mathcal{Z}) = C_{NFB} \qquad (23)$$

for a stationary ergodic noise. But by definition of the feedback capacity, we know that $C_{NFB} \leq C_{FB}$. Thus we have shown the following.

Theorem 1: Feedback does not increase the capacity of the NBNDC with stationary ergodic noise:

$$C_{FB} = C_{NFB} = 1 + H\mathcal{W}) - H(\mathcal{Z}).$$

Observation: Remark that, since Lemma 1 holds for arbitrary noise processes, Theorem 1 can be extended for such noise sources (i.e., without requiring them to be stationary ergodic) by using Verdú and Han's non-feedback capacity formula for general channels with memory [8].

V. Conclusions

In this work, we investigated the feedback capacity of a discrete binary-input 2^q-ary output communication channel with memory which was recently proposed in [5] to model BPSK-modulated correlated fading channels used in conjunction with 2^q-ary soft-decision demodulation. We showed that feedback does not increase the capacity of this channel. The result is obtained by first demonstrating that, due to the channel invertibility properties, the best feedback policy is a uniform policy, as in the non-feedback case. Future work may include the study of the channel's capacity-cost function (with and without feedback); i.e., the largest rate for reliably communicating over the channel when cost constraints are imposed on its binary-valued input. It is plausible that in this case feedback can strictly increase the channel's capacity-cost function.

References

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

[2] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inform. Theory,* vol. 35, pp. 37-43, Jan. 1989.

[3] P. Ebert, "The capacity of Gaussian channel with feedback," *Bell Syst. Tech. J.* , pp. 1705-1712, Oct. 1970.

[4] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[5] C. Pimentel and F. Alajaji, "A discrete channel model for capturing memory and soft-decision information: A capacity study," to appear in *Proc. IEEE Int. Conf. Commun.* Dresden, Germany, June 2009.

[6] M. Pinsker, Talk at *the Soviet Information Theory Meeting,* 1969 (no abstract published).

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

[8] S. Verdú and T. H. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory,* vol. 40, pp. 1147-1157, Jul. 1994.