Capacity Theorems for Relay Channels with ISI

Ninoslav Marina, Aleksandar Kavčić and N. Thomas Gaarder
Department of Electrical Engineering
University of Hawai'i at Mānoa
2540 Dole Street, Honolulu, HI 96822

Abstract—**In this paper we initially study degraded relay channels with finite-length intersymbol interference (ISI). For such channels, we show that the decode-and-forward strategy achieves the capacity, and prove a special structure for the capacity achieving distributions of the source and relay signals. We also prove that a general memoryless relay channel used with delayed feedback from the destination node to the relay node is an instance of a degraded relay channel with ISI, and observe that the delayed feedback from the destination node to the relay node does not decrease the capacity compared to instantaneous feedback. In all cases where the channel is used with delayed feedback from the destination node to the relay node the decode-and-forward scheme is optimal and the capacity is not decreased by delaying the feedback from the destination node. We extend these results to general (non-degraded) relay channels with ISI to obtain upper and lower bounds on their capacities.**

Index Terms — **Relay channel, channels with memory, cooperative communications, channel capacity, delayed feedback capacity, feedback capacity, finite-state machine channels, intersymbol interference (ISI).**

I. Introduction

A model for the relay channel was introduced and studied in the pioneering work by van der Meulen [1], [2], [3]. Substantial advances in the theory were made by Cover and El Gamal, who developed two fundamental coding strategies for the relay channel [4]. A combination of these strategies achieves capacities for several classes of degraded memoryless relay channels. Most of the work done so far was related to memoryless relay channels with or without feedback. In this paper we analyze the feedback capacity of the general relay channel with finite length ISI, or equivalently with memory.

First in Section II we introduce the notation and the signal models used in the paper. In Section III we describe the degraded relay channel with memory and derive an expression for its capacity. Section IV establishes the link between the degraded relay channel with ISI and the general memoryless relay channel used with delayed feedback. Section V gives lower and upper bounds on the capacity of a general ISI relay channel. Finally, Section VI concludes the paper.

II. Notation and signal model

In the following text, the uppercase letters represent random variables (or vectors), while lowercase letters represent their realizations. A random variable at discrete time $i \in \mathbb{Z}$ is indexed by i (e.g., X_i). A vector of time-dependent variables is denoted as $X_i^j = [X_i, X_{i+1}, \ldots, X_j]$ and $X^i = X_{-\infty}^i$. $p_X(x)$ is the probability density function of the random variable X, $p_{X,Y}(x,y)$ denotes the joint probability of X and Y, while $p_{Y|X}(y|x)$ stands for the conditional probability of Y given X. Throughout the paper we shall use the notational convenience $p_{X,Y}(x,y) = p(x,y)$ or $p_{Y|X}(y|x) = p(y|x)$, where the dropped subscripts will be obvious from the arguments of the function. To avoid cumbersome distinctions between discrete and continuous random variables, we shall use use $H(X)$ to represent either the entropy of a discrete random variable, or the differential entropy of a continuous random variable. $I(X;Y)$ is the mutual information between the random variables X and Y. All the logarithms are base 2, so the measure units for all information rates are bits.

The general relay channel with ISI of length m is denoted by $(\mathcal{X} \times \mathcal{U}, f(y_i, v_i | x_{i-m}^i, u_{i-m}^i), \mathcal{Y} \times \mathcal{V})$. In order to better distinguish the input and output probabilities from the channel transition probability, we denote the latter by $f(y, v|x, u)$. The channel consists of four sets: $\mathcal{X}, \mathcal{U}, \mathcal{Y}, \mathcal{V}$, and a collection of conditional probability (density) functions $f(y_i, v_i | x_{i-m}^i, u_{i-m}^i)$ on $\mathcal{Y} \times \mathcal{V}$, one for each tuple $(x_{i-m}^i, u_{i-m}^i) \in \mathcal{X}^{m+1} \times \mathcal{U}^{m+1}$. The memoryless channel ($m = 0$) is denoted by $(\mathcal{X} \times \mathcal{U}, f(y_i, v_i | x_i, u_i), \mathcal{Y} \times \mathcal{V})$.

III. Degraded relay channel with ISI

In this section we consider the degraded relay channel with ISI of length m described in (Fig. 1).

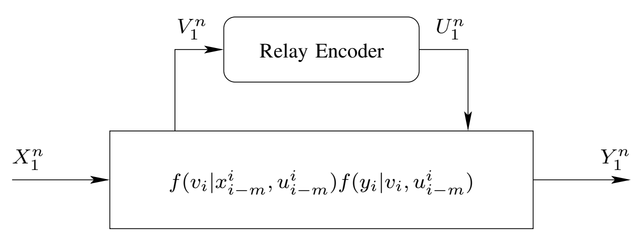

Fig. 1. Degraded relay channel with memory m.

Definition 1: A relay channel with ISI of length m, $(\mathcal{X} \times \mathcal{U}, f(y_i, v_i | x_{i-m}^i, u_{i-m}^i), \mathcal{Y} \times \mathcal{V})$ is said to be *physically degraded* if its channel transition probability $f(y_i, v_i | x_{i-m}^i, u_{i-m}^i)$ can be written in the form

$$f(y_i, v_i | x^i, u^i) = f(v_i | x_{i-m}^i, u_{i-m}^i) f(y_i | v_i, u_{i-m}^i).$$

Equivalently we see that if the relay channel is degraded, then for any i, $f(y_i | v_i, x_{i-m}^i, u_{i-m}^i) = f(y_i | v_i, u_{i-m}^i)$, i.e., $X_{i-m}^i \rightarrow (U_{i-m}^i, V_i) \rightarrow Y_i$ form a Markov chain. We next characterize the capacity of a physically degraded channel with ISI.

Lemma 2: For time-invariant channels the capacity achieving input process is stationary.

Proof: The proof is given in [13].

□

Theorem 3: The capacity of the degraded relay channel with ISI of finite length m, $(\mathcal{X}\times\mathcal{U}, f(y_i, v_i|x_{i-m}^i, u_{i-m}^i), \mathcal{Y}\times\mathcal{V})$ is given by

$$C = \lim_{n\uparrow\infty} \frac{1}{n} \max_{p(x_1^n, u_1^n)} \min\left\{ \begin{array}{l} I(X_1^n, U_1^n; Y_1^n), \\ I(X_1^n; V_1^n|U_1^n) \end{array} \right\}, \tag{1}$$

where the capacity achieving processes X and U are jointly stationary.

Proof: Since the relay channel is degraded and time-invariant, the expression for the feedback capacity follows directly from [4] if we replace $(X_1^n, U_1^n, Y_1^n, V_1^n)$ by $(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{V})$ and then let $n \uparrow \infty$. The proof mimics that of [4] with the only difference being that we need to use *vectors* as signaling symbols, as described in [6]. To avoid a lengthy proof, we only provide a sketch.

The code that achieves the capacity (1) uses the random-binning proof described in [5]. We use the same block-Markov method as in [4], however applied on vectors of length n.

First generate at random 2^{NnR_0} independent and identically distributed (iid) N sequences of vectors of length n, according to $p(\mathbf{u}_1^N) = \prod_{i=1}^N p(\mathbf{u}_i)$ and index them as $\mathbf{u}_1^N(s)$, $s \in [1, 2^{NnR_0}]$. Note that $\mathbf{u}_i$ is a vector of length n and $\mathbf{u}_1^N$ is a vector of length $N \times n$ that represents a codeword (row) in the codebook. For each $\mathbf{u}_1^N(s)$, generate 2^{NnR} conditionally independent N-sequences of vectors of length n, and index them as $\mathbf{x}_1^N(w|s)$, $w \in [1, 2^{NnR}]$ drawn according to $p(\mathbf{x}_1^N|\mathbf{u}_1^N) = \prod_{i=1}^N p(\mathbf{x}_i|\mathbf{u}_i(s))$. We also need a random partition $\mathcal{S} = \{S_1, S_2, \ldots, S_{2^{NnR_0}}\}$ of the message set $\mathcal{W} = \{1, 2, \ldots, 2^{NnR}\}$ into 2^{NnR_0} cells, with $S_i \cap S_j = \emptyset, i \neq j$ and $\bigcup_i^{2^{NnR_0}} S_i = \mathcal{W}$. With a good choice of $\{\mathcal{C}, \mathcal{S}\}$, for N large enough, the relay will know w_i and the receiver will know (w_{i-1}, s_i) at the end of block i with probability of error smaller than some $\varepsilon > 0$. Because of the memory of length m, we transmit the following codewords of length $N(n+m)$ in block b

$$\begin{array}{l} [\boldsymbol{\xi}_0, \mathbf{x}_1(w_b|s_b), \boldsymbol{\xi}_0, \mathbf{x}_2(w_b|s_b), \ldots, \boldsymbol{\xi}_0, \mathbf{x}_N(w_b|s_b)] \\ [\boldsymbol{\eta}_0, \mathbf{u}_1(s_b), \boldsymbol{\eta}_0, \mathbf{u}_2(s_b), \ldots, \boldsymbol{\eta}_0, \mathbf{u}_N(s_b)]. \end{array}$$

The vectors $\boldsymbol{\xi}_0$ and $\boldsymbol{\eta}_0$ are constant vectors of length m and are known as the *initial state* of the channel if the channel is a controllable state machine. In other words, at most m steps are necessary to come back to the initial state. For example $\boldsymbol{\xi}_0$ and $\boldsymbol{\eta}_0$ both could be the all-zero vectors of length m. The relay/channel output in block b are

$$\begin{array}{l} [\mathbf{Y}_1(b), \boldsymbol{\gamma}_1, \mathbf{Y}_2(b), \boldsymbol{\gamma}_2, \ldots, \mathbf{Y}_N(b), \boldsymbol{\gamma}_N] \\ [\mathbf{V}_1(b), \boldsymbol{\nu}_1, \mathbf{V}_2(b), \boldsymbol{\nu}_2, \ldots, \mathbf{V}_N(b), \boldsymbol{\nu}_N] \end{array}$$

received according to

$$\begin{aligned} & f(\mathbf{y}_1^N, \mathbf{v}_1^N, \boldsymbol{\gamma}_1^N, \boldsymbol{\nu}_1^N|\mathbf{x}_1^N(w|s), \mathbf{u}_1^N(s), \boldsymbol{\xi}_0, \boldsymbol{\eta}_0) \\ = & \prod_{i=1}^N f(\mathbf{y}_i, \mathbf{v}_i, \boldsymbol{\gamma}_i, \boldsymbol{\nu}_i|\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\xi}_0, \boldsymbol{\eta}_0), \end{aligned} \tag{2}$$

where the factorization in (2) is possible because the known vectors $\boldsymbol{\xi}_0$ and $\boldsymbol{\eta}_0$ decouple $(\mathbf{y}_i, \boldsymbol{\gamma}_i, \mathbf{v}_i, \boldsymbol{\nu}_i)$ from $(\mathbf{y}_{i+1}, \boldsymbol{\gamma}_{i+1}, \mathbf{v}_{i+1}, \boldsymbol{\nu}_{i+1})$. Then, following the same procedure as in [4], using joint-typicality decoding based on vectors of length n, we can show that as $N \uparrow \infty$, we get that

$$(n+m)R < I(X_1^n; V_1^{n+m}|U_1^n, \boldsymbol{\xi}_0, \boldsymbol{\eta}_0)$$

and

$$(n+m)R < I(X_1^n, U_1^n; Y_1^{n+m}|\boldsymbol{\xi}_0, \boldsymbol{\eta}_0).$$

Since the relay channel is degraded, we may easily prove the converse by applying the method from [4] on vectors of length n.

Now we use the standard procedure where we let $n \to \infty$, so we have

$$C = \lim_{n\uparrow\infty} \frac{1}{n+m} \max_{p(x_1^n, u_1^n)} \min\left\{ \begin{array}{l} I(X_1^n, U_1^n; Y_1^{n+m}|\boldsymbol{\xi}_0, \boldsymbol{\eta}_0), \\ I(X_1^n; V_1^{n+m}|U_1^n, \boldsymbol{\xi}_0, \boldsymbol{\eta}_0) \end{array} \right\}$$

which is equivalent to (1) since m is finite and since the initial state $(\boldsymbol{\xi}_0, \boldsymbol{\eta}_0)$ of finite-length ISI channels does not alter the information rate. Finally, Lemma 2 establishes joint stationarity of X and U.

□

Expression (1) gives the capacity of a degraded relay channel with memory, however it is impractical for computation. Our goal is to further characterize the capacity achieving process and simplify expression (1).

Lemma 4: For the degraded relay channel with ISI

$$I(X^i; V_i|V^{i-1}, U^n) \leq I(X^i; V_i|V^{i-1}, U^i).$$

Proof:

$$\begin{aligned} & I(X^i; V_i|V^{i-1}, U^n) \\ = \; & H(V_i|V^{i-1}, U^n) - H(V_i|X^i, V^{i-1}, U^n) \\ \stackrel{(a)}{=} \; & H(V_i|V^{i-1}, U^n) - H(V_i|X^i, V^{i-1}, U^i) \\ \stackrel{(b)}{\leq} \; & H(V_i|V^{i-1}, U^i) - H(V_i|X^i, V^{i-1}, U^i) \\ = \; & I(X^i; V_i|V^{i-1}, U^i), \end{aligned} \tag{3}$$

where (a) follows from the fact that for all i, V_i is conditionally independent of U_{i+1}^n given V^{i-1} and X^i and (b) due to the fact that conditioning reduces entropy.

□

In order to precisely characterize the capacity achieving source, we will distinguish two types of sources. We call them "type p" and "type q". Type p will stand for a *general* source

$$p(x_1^n, u_1^n) = \prod_{i=1}^n p(x_i|x^{i-1}, u^n) p(u_i|u^{i-1}),$$

while type q is a more *constrained* source derived from p and defined as

$$q(x_1^n, u_1^n) \triangleq \prod_{i=1}^{n} p(x_i|x^{i-1}, u^i)p(u_i|u^{i-1}).$$

We further underline the importance of the difference between constrained sources of type q and general sources of type p. Obviously sources of type q form a subset of all sources of type p. We will show that the capacity achieving source must be of type q. However, before we establish this formally, we give an intuitive interpretation of this claim. For a general source of type p, we have for any i

$$p(x_i, u_i|x^{i-1}, u^{i-1}) = p(x_i|x^{i-1}, u^i)p(u_i|u^{i-1}, x^{i-1}),$$

but for the source of type q, we have for any i

$$q(x_i, u_i|x^{i-1}, u^{i-1}) = q(x_i|x^{i-1}, u^i)q(u_i|u^{i-1}). \quad (4)$$

This reveals that under the probability law q, the process U is allowed to evolve on its own without any dependence on the process X, while the process X may be causally dependent on U. So, under the probability law q, it is as if the source node (X) is capable of observing the relay node signal U. From [4] we know that this is possible by using the block Markov coding strategy. A general source p does not have this property. So, even though in (1) we may optimize over a general source p, the maximum will be achieved by a source of type q which we formally establish next.

In order to distinguish between two different mutual information terms that are induced by two different input probability measures p and q, we denote them by $I_p(\cdot;\cdot)$ and $I_q(\cdot;\cdot)$, respectively. We shall use the same notation for the corresponding entropies, i.e., $H_p(\cdot|\cdot)$ and $H_q(\cdot|\cdot)$.

Lemma 5: For the degraded relay channel with ISI

$$\begin{aligned} I_p(X^i, U^i; Y_i|Y^{i-1}) &= I_q(X^i, U^i; Y_i|Y^{i-1}) && (5)\\ I_p(X^i, V_i|V^{i-1}, U^i) &= I_q(X^i, V_i|V^{i-1}, U^i) && (6)\\ &= I_q(X^i, V_i|V^{i-1}, U^n). && (7) \end{aligned}$$

Proof: From the definition of q, we see that $q(x_i|x^{i-1}, u^n) = q(x_i|x^{i-1}, u^i)$ for every i. This is equivalent to

$$q(x^i|u^n) = q(x^i|u^i) \qquad \text{for every } i.$$

Also, it is easy to see from the construction of the probability measure $q(\cdot,\cdot)$ that $q(x_i|x^{i-1}, u^i) = p(x_i|x^{i-1}, u^i)$, which, by induction, leads to

$$q(x^i|u^i) = p(x^i|u^i) \qquad \text{for every } i.$$

We will use these equalities to prove the lemma.

Equation (5) holds since both conditional input probabilities $p(x^i|u^i)$ and $q(x^i|u^i)$ give rise to the same joint probability measure

$$\begin{aligned} p(y^i, u^i, x^i) &= p(y^i|u^i, x^i)p(x^i, u^i)\\ &= p(y^i|u^i, x^i)p(u^i)p(x^i|u^i)\\ &= p(y^i|u^i, x^i)p(u^i)q(x^i|u^i)\\ &= q(y^i, u^i, x^i). \end{aligned}$$

To prove (6) and (7) note that

$$\begin{aligned} I_p(X^i, V_i|V^{i-1}, U^i) &\overset{(a)}{=} I_q(X^i, V_i|V^{i-1}, U^i)\\ &\overset{(b)}{=} I_q(X^i, V_i|V^{i-1}, U^n), \end{aligned}$$

where (a) holds since both conditional input probabilities $p(x^i|u^i)$ and $q(x^i|u^i)$ induce the same probability measure

$$\begin{aligned} p(v^i, u^i, x^i) &= p(v^i|u^i, x^i)p(x^i, u^i)\\ &= p(v^i|u^i, x^i)p(u^i)p(x^i|u^i)\\ &= p(v^i|u^i, x^i)p(u^i)q(x^i|u^i)\\ &= q(v^i, u^i, x^i). \end{aligned}$$

To prove equality (b) observe that

$$\begin{aligned} &I_q(X^i, V_i|V^{i-1}, U^i)\\ =\ &H_q(V_i|V^{i-1}, U^i) - H_q(V_i|X^i, V^{i-1}, U^i)\\ =\ &H_q(V_i|V^{i-1}, U^i) - H_q(V_i|X^i, V^{i-1}, U^n). \quad (8) \end{aligned}$$

For the second entropy in (8) we use the fact that V_i is conditionally independent of U_{i+1}^n given V^{i-1} and X^i. We now show that $H_q(V_i|V^{i-1}, U^n) = H_q(V_i|V^{i-1}, U^i)$ which concludes the proof of (b).

$$\begin{aligned} &H_q(V_i|V^{i-1}, U^n)\\ =\ &\int\limits_{v^i, u^i, u_{i+1}^n, x^i} f(v^i|x^i, u^n)q(x^i|u^n)p(u^i)p(u_{i+1}^n|u^i)\\ \times\ &\log \frac{p(u^n)\int_{\hat{x}^i} f(v^i|\hat{x}^i, u^n)q(\hat{x}^i|u^n)d\hat{x}^i}{p(u^n)\int_{\hat{x}^{i-1}} f(v^{i-1}|\hat{x}^{i-1}, u^n)q(\hat{x}^{i-1}|u^n)d\hat{x}^{i-1}} d\psi\\ =\ &\int\limits_{v^i, u^i, x^i} f(v^i|x^i, u^i)q(x^i|u^i)p(u^i)\\ \times\ &\log \frac{\int_{\hat{x}^i} f(v^i|\hat{x}^i, u^i)q(\hat{x}^i|u^i)d\hat{x}^i}{\int_{\hat{x}^{i-1}} f(v^{i-1}|\hat{x}^{i-1}, u^i)q(\hat{x}^{i-1}|u^i)d\hat{x}^{i-1}} dv^i du^i dx^i\\ =\ &H_q(V_i|V^{i-1}, U^i), \end{aligned}$$

where for convenience of writing we use $d\psi = dv^i du^n dx^i$.

□

For future use, we combine Lemmas 4 and 5 into one Corollary.

Corollary 6:

$$\begin{aligned} I_p(X^i, U^i; Y_i|Y^{i-1}) &= I_q(X^i, U^i; Y_i|Y^{i-1}),\\ I_p(X^i, V_i|V^{i-1}, U^n) &\leq I_q(X^i, V_i|V^{i-1}, U^n)\\ &= I_q(X^i, V_i|V^{i-1}, U^i). \quad (9) \end{aligned}$$

Theorem 7: The capacity of the degraded relay channel with ISI memory of length m is

$$C = \max_q \min \left\{ \begin{array}{l} I(X_{i-m}^i, U_{i-m}^i; Y_i|Y^{i-1}),\\ I(X_{i-m}^i; V_i|V^{i-1}, U^i) \end{array} \right\}, \quad (10)$$

where the maximization is taken over the input distribution $q = p(x_i|x^{i-1}, u^i)p(u_i|u^{i-1})$.

Proof: Starting from Theorem 3

$$\begin{aligned}
C &= \lim_{n\uparrow\infty}\frac{1}{n}\max_{p(x_1^n,u_1^n)}\min\left\{\begin{array}{l} I(X_1^n,U_1^n;Y_1^n),\\ I(X_1^n;V_1^n|U_1^n)\end{array}\right\}\\
&\overset{(a)}{=} \lim_{n\uparrow\infty}\frac{1}{n}\max_{p}\min\left\{\begin{array}{l}\sum_{i=1}^{n} I(X^i,U^i;Y_i|Y^{i-1}),\\ \sum_{i=1}^{n} I(X^i;V_i|V^{i-1},U^n)\end{array}\right\}\\
&\overset{(b)}{=} \lim_{n\uparrow\infty}\frac{1}{n}\max_{q}\min\left\{\begin{array}{l}\sum_{i=1}^{n} I(X^i,U^i;Y_i|Y^{i-1}),\\ \sum_{i=1}^{n} I(X^i;V_i|V^{i-1},U^n)\end{array}\right\}\\
&\overset{(b')}{=} \lim_{n\uparrow\infty}\frac{1}{n}\max_{q}\min\left\{\begin{array}{l}\sum_{i=1}^{n} I(X^i,U^i;Y_i|Y^{i-1}),\\ \sum_{i=1}^{n} I(X^i;V_i|V^{i-1},U^i)\end{array}\right\}\\
&\overset{(c)}{=} \max_{q}\min\left\{\begin{array}{l} I(X^i,U^i;Y_i|Y^{i-1}),\\ I(X^i;V_i|V^{i-1},U^i)\end{array}\right\}\\
&\overset{(d)}{=} \max_{q}\min\left\{\begin{array}{l} I(X_{i-m}^i,U_{i-m}^i;Y_i|Y^{i-1}),\\ I(X_{i-m}^i;V_i|V^{i-1},U^i)\end{array}\right\}.
\end{aligned}$$

Equality (a) follows from the chain rule for mutual information that is applied to channels used without feedback, equalities (b) and (b') follow from Corollary 6 and the fact that sources of "type q" form a subset of the sources of "type p", (c) from the stationarity of the optimal source (Lemma 2), while (d) follows from the fact that the channel has finite memory m.

□

Comment: The directed information rate between two random processes X and Y is given by [9], [10],

$$\mathcal{I}(X\to Y)=\lim_{n\uparrow\infty}\frac{1}{n}\sum_{i=1}^{n} I(X^i;Y_i|Y^{i-1}),$$

and the causally conditioned directed information rate between X and V given U, by [11]

$$\mathcal{I}(X\to V\|U)=\frac{1}{n}\lim_{n\uparrow\infty}\sum_{i=1}^{n} I(X^i;V_i|U^i,V^{i-1}).$$

Note that

$$\begin{aligned}
I(X_{i-m}^i,U_{i-m}^i;Y_i|Y^{i-1}) &= \mathcal{I}(X,U\to Y)\\
I(X_{i-m}^i;V_i|V^{i-1},U^i) &= \mathcal{I}(X\to V\|U).
\end{aligned}$$

Therefore, the capacity of the degraded relay channel with ISI memory of length m could be written in a more compact form as

$$C=\max_{q}\min\left\{\mathcal{I}(X,U\to Y),\mathcal{I}(X\to V\|U)\right\}. \qquad (11)$$

IV. General memoryless relay channel used with delayed feedback

Consider now the general memoryless relay channel used with noiseless delayed feedback from the destination to the relay node, shown in Fig. 2. Since the feedback is noiseless, we assume that the relay, before emitting the symbol U_i, knows all previous channel output symbols Y^{i-d-1} without an error.

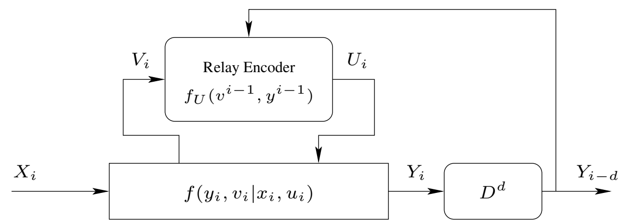

Fig. 2. Relay channel used with feedback of delay d.

First note that any general memoryless relay channel used with feedback of delay d is an instance of a degraded relay channel with ISI of length d. To see this, let $\tilde{Y}_i=Y_{i-d}$ and $\tilde{V}_i=(V_i,Y_{i-d})$. Then, from

$$\begin{aligned}
f(\tilde{y}_i,\tilde{v}_i|x^i,u^i) &= f(y_{i-d},v_i|x_i,x_{i-d},u_i,u_{i-d})\\
&= p(\tilde{y}_i,\tilde{v}_i|x_{i-d}^i,u_{i-d}^i)\\
&= p(\tilde{v}_i|x_{i-d}^i,u_{i-d}^i)\underbrace{p(\tilde{y}_i|\tilde{v}_i,x_{i-d}^i,u_{i-d}^i)}_{1}\\
&= p(\tilde{v}_i|x_{i-d}^i,u_{i-d}^i)p(\tilde{y}_i|\tilde{v}_i,u_{i-d}^i),
\end{aligned}$$

we conclude that the general memoryless relay channel used with feedback of delay d is equivalent to a degraded relay channel with memory d. In that case their capacities are given by Theorem 7, i.e., the capacity of the general memoryless relay channel used with feedback with delay d is

$$\begin{aligned}
C^{FB(d)} &= \max_{q}\min\left\{\begin{array}{l} I(X_{i-d}^i,U_{i-d}^i;\tilde{Y}_i|\tilde{Y}^{i-1}),\\ I(X_{i-d}^i;\tilde{V}_i|\tilde{V}^{i-1},U^i)\end{array}\right\}\\
&= \max_{q}\min\left\{\begin{array}{l} I(X_{i-d}^i,U_{i-d}^i;Y_{i-d}|Y^{i-d-1}),\\ I(X_{i-d}^i;V_i,Y_{i-d}|V^{i-1},Y^{i-d-1},U^i)\end{array}\right\}, \qquad (12)
\end{aligned}$$

where $q=p(x_i|x^{i-1},u^i)p(u_i|u^{i-1})$.

Proposition 8:

$$C^{FB(d)} = C^{FB(0)}.$$

Proof: Since we use a delayed feedback, it is true that

$$C^{FB(d)} \le C^{FB(0)}.$$

Next we show that $C^{FB(d)}\ge C^{FB(0)}$. Let us calculate both mutual information terms in (12) by using a memoryless source $p(x_i|u_i)p(u_i)$. In that case we have

$$\begin{aligned}
I(X_{i-d}^i,U_{i-d}^i;Y_{i-d}|Y^{i-d-1}) &= I(X_{i-d},U_{i-d};Y_{i-d})\\
&= I(X_i,U_i;Y_i)
\end{aligned}$$

and

$$\begin{aligned} & I(X_{i-d}^i; V_i, Y_{i-d} | V^{i-1}, Y^{i-d-1}, U^i) \\ = \; & I(X_i; V_i | U_i) + I(X_{i-d}; Y_{i-d} | V_{i-d}, U_{i-d}) \\ = \; & I(X_i; V_i | U_i) + I(X_i; Y_i | V_i, U_i) \\ = \; & I(X_i; V_i, Y_i | U_i). \end{aligned}$$

Since we use a particular class of memoryless sources $p(x_i, u_i)$, the maximization of these two mutual information terms over this class of sources is a lower bound on $C^{FB(d)}$. At the same time, from [4] we have

$$\max_{p(x_i,u_i)} \min \{I(X_i, U_i; Y_i), I(X_i; V_i, Y_i | U_i)\} = C^{FB(0)}.$$

Hence,

$$C^{FB(0)} \quad \leq \quad C^{FB(d)}.$$

□

Note that similar reasoning could be also straightforwardly applied to general relay channels with ISI. That means that delaying the feedback does not decrease the capacity of any general relay channel used with instantaneous feedback.

V. General relay channels with ISI

We are now in a position to establish lower and upper bounds for the general (non-degraded) relay channel with ISI. It is straightforward that the capacity of a degraded relay channel with ISI is a lower bound for the capacity of the general relay channel with ISI. It is also clear that the feedback capacity is an upper bound for the capacity. Therefore,

$$\begin{aligned} C \quad & \geq \quad \max_q \min \left\{ \begin{array}{l} I(X_{i-m}^i, U_{i-m}^i; Y_i | Y^{i-1}), \\ I(X_{i-m}^i; V_i | V^{i-1}, U^i) \end{array} \right\} \\ C \quad & \leq \quad \max_q \min \left\{ \begin{array}{l} I(X_{i-m}^i, U_{i-m}^i; Y_i | Y^{i-1}), \\ I(X_{i-m}^i; V_i, Y_i | V^{i-1}, Y^{i-1}, U^i) \end{array} \right\}, \end{aligned}$$

where $q = p(x_i | x^{i-1}, u^i) p(u_i | u^{i-1})$. These bounds are very similar to [4], with the distinction that here the maximization is over sources with memory of type q as defined in (4). We can also characterize the bounds using the directed information rate notation

$$\begin{aligned} C \quad & \geq \quad \max_q \min \{\mathcal{I}(X, U \to Y), \mathcal{I}(X \to V \| U)\} \\ C \quad & \leq \quad \max_q \min \{\mathcal{I}(X, U \to Y), \mathcal{I}(X \to Y, V \| U)\}. \end{aligned}$$

Finally, similar to arguments given in Section IV, we can show that for general relay channels with ISI, delaying the feedback from the destination node to the relay node does not reduce the feedback capacity, i.e.,

$$C^{FB(d)} \quad = \quad C^{FB(0)}.$$

VI. Conclusion

We established that the decode-and-forward strategy achieves the capacity of any degraded relay channel with ISI. We also characterized the capacity achieving source distribution to consist of a freely running relay signal U (independent of the source signal X), while the source signal X may be *causally* dependent on the relay signal U. Any general relay channel (with or without ISI) when used with (delayed) feedback from the destination node to the relay node is an instance of a degraded relay channel with ISI, so the (delayed) feedback capacities can readily be established. We also showed that the delayed feedback capacity equals the instantaneous feedback capacity.

Acknowledgment

N. Marina would like to thank Swiss National Science Foundation for supporting his research (Grant PA002–117385).

References

[1] E. C. van der Meulen, "Transmission of information in a $T-$terminal discrete memoryless channel," Ph.D. dissertation, Dep. of Statistics, University of California, Berkeley, 1968.

[2] E. C. van der Meulen, "Three-terminal communication channels,", *Adv. Appl. Prob.*, vol. 3, pp. 120-154, 1971.

[3] E. C. van der Meulen, "A survey of multiway channels in information theory: 1961 - 1976,", *IEEE Trans. Inf. Theory*, 23(1): 1-37, Jan. 1977.

[4] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel", *IEEE Trans. Inf. Theory*, 25(5): 572 - 584, Sep. 1979.

[5] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", *IEEE Trans. Inf. Theory*, 19(4): 471-480, Jul. 1973.

[6] A. Kavčić, X. Ma and M. Mitzenmacher, "Binary Intersymbol Interference Channels: Gallager Codes, Density Evolution, and Code Performance Bounds," *IEEE Trans. Inf. Theory*, 49(7): 1636 - 1652, Jul. 2003.

[7] S. Yang, A. Kavčić and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, 51(3): 799-810, Mar. 2005.

[8] S. Yang, A. Kavčić and S. Tatikonda, "On the Feedback Capacity of Power-Constrained Gaussian Noise Channels With Memory," *IEEE Trans. Inf. Theory*, 53(3): 929-954, Mar. 2007.

[9] J. Massey, "Causality, feedback and directed information," *Proc. Int. Symp. Information Theory and its Applications (ISITA '90)*, pp. 303-305, 1990.

[10] H. Marko, "The bidirectional communication theory - a generalization of information theory," *IEEE Trans. on Communications*, 21(3): 1335-1351, 1973.

[11] G. Kramer, "Directed information for channels with feedback," Swiss Federal Institute of Technology Zürich (ETHZ), Zürich, 1998.

[12] N. Marina, "Successive decoding," Ph.D. Dissertation, School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, 2004.

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

[14] T. M. Cover and J.A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.