On Separation in the Presence of Feedback

Haim Permuter
EE Department
Stanford University
haim1@stanford.edu

Tsachy Weissman
EE Department
Stanford University
tsachy@stanofrd.edu

***Abstract*— We present a framework for characterizing the fundamental limit of communication with feedback for various point-to-point and multi-terminal finite-state channels. Among other benefits, this framework allows to identify joint source-channel coding scenarios for which separation holds. For example, we will characterize a large class of finite-state multiple access channels for which separation holds with, without, or with partial feedback.**

I. Introduction

The source channel separation theorem was stated by Shannon [1] for a stationary source and memoryless channel. The theorem allows us to decompose the problem of communicating sources over channels in two separate steps: In the first step (source coding) we assume that the source is a random process and the channel allows communicating an arbitrary number of bits per use without any error. In the second step (channel coding), we assume that we want to transmit a message, uniformly distributed over an arbitrary finite alphabet, through a channel. The separation theorem claims that if we concatenate the optimum coding scheme of the two steps, we obtain an optimum coding scheme for the source-channel problem.

Somewhat surprisingly, the source-channel separation does not hold for the multi-access channel (MAC). Cover, El-Gamal and Salehi [2] showed that, in general, the source channel separation does not hold even for a memoryless MAC.

In this paper, we consider the source-channel separation theorem in the presence of feedback. The main ingredient for deriving a source-channel separation theorem for the case where feedback is allowed and the channel has memory, is a multi-letter characterization of the feedback capacity. The multi-letter characterization of the feedback capacity uses the directed information and causal conditioning notation which is introduced in Section II. We establish that separation holds for point-to-point channels whose capacity is characterized in terms of directed information. Further, we establish separation for the class of multiplexer MACs (to be defined below).

II. Directed information and causal conditioning notation

Throughout this paper we use superscripts to denote vectors in the following way: $x^i = (x_1 \dots x_i)$ and $x_i^j = (x_i \dots x_j)$ for $i > j$. Moreover, we use lower case to denote sample values and upper case to denote random variables. Probability distributions are denoted by P or Q when the arguments

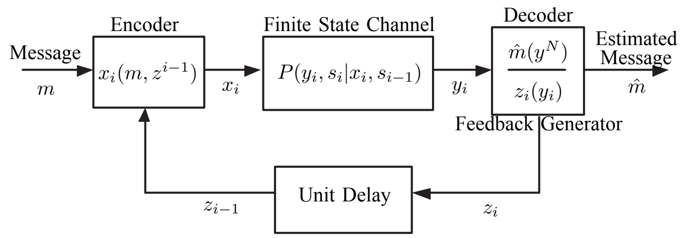

Fig. 1. Channel with feedback that is a time invariant deterministic function of the output

specify the distribution, e.g. $P(x|y) = P(X = x|Y = y)$. We are using the *directed information* $I(X^n \to Y^n)$ and the *causal conditioning* notation $(\cdot||\cdot)$. We denote the probability mass function (pmf) of Y^N causally conditioned on X^{N-d}, for some integer $d \geq 0$, as $P(y^N||x^{N-d})$ which is defined as

$$P(y^N||x^{N-d}) \triangleq \prod_{i=1}^{N} P(y_i|y^{i-1}, x^{i-d}), \qquad (1)$$

(if $i - d \leq 0$ then x^{i-d} is set to null). In particular, we extensively use the cases where $d = 0, 1$:

$$P(y^N||x^N) \triangleq \prod_{i=1}^{N} P(y_i|y^{i-1}, x^i) \qquad (2)$$

$$Q(x^N||y^{N-1}) \triangleq \prod_{i=1}^{N} Q(x_i|x^{i-1}, y^{i-1}), \qquad (3)$$

where the letters Q and P are both used for denoting pmf.

Directed information $I(X^N \to Y^N)$ was defined by Massey [3] as

$$I(X^N \to Y^N) \triangleq \sum_{i=1}^{N} I(X^i; Y_i|Y^{i-1}). \qquad (4)$$

and can also be expressed in terms of causal conditioning as

$$I(X^N \to Y^N) = \sum_{i=1}^{N} I(X^i; Y_i|Y^{i-1}) = \mathbf{E}\left[\log \frac{P(Y^N||X^N)}{P(Y^N)}\right] \qquad (5)$$

where $\mathbf{E}$ denotes expectation. The directed information between X^N and Y^N, causally conditioned on Z^N, is denoted as $I(X^N \to Y^N||Z^N)$ and defined as:

$$I(X^N \to Y^N||Z^N) \triangleq \sum_{i=1}^{N} I(Y_i; X^i|Y^{i-1}, Z^i) \qquad (6)$$

1-4244-1564-0/07/$25.00 ©2007 IEEE

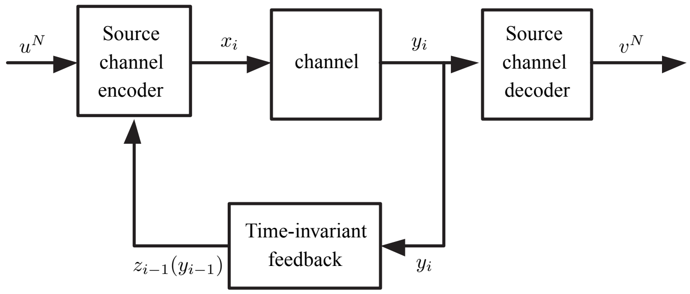

Fig. 2. Source and channel coding, where the channel has time-invariant feedback.

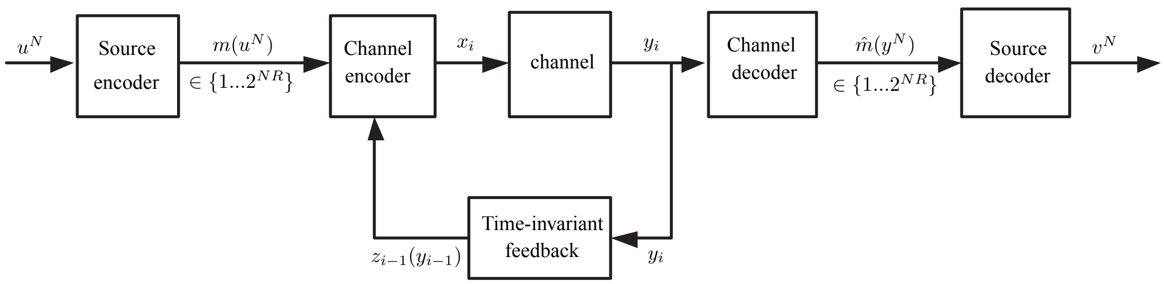

Fig. 3. Source and channel separation.

III. Point-to-point Channel Models and feedback capacity

In this paper, we consider finite state channels (FSCs). The FSCs are a class of channels rich enough to include channels with memory, e.g. channels with intersymbol interference. The input of the channel is denoted by $\{X_1, X_2, \dots\}$, and the output of the channel is denoted by $\{Y_1, Y_2, \dots\}$, both taking values in a finite alphabet. In addition, the channel states take values in a finite set of possible states. The channel is stationary and is characterized by a conditional probability assignment $P(y_i, s_i|x_i, s_{i-1})$ that satisfies $P(y_i, s_i|x^i, s^{i-1}, y^{i-1}) = P(y_i, s_i|x_i, s_{i-1})$. An FSC is said to be without intersymbol interference (ISI) if the input sequence does not effect the evolution of the state sequence, i.e. $P(s_i|s_{i-1}, x_i) = P(s_i|s_{i-1})$.

We assume a communication setting that includes feedback as shown in Fig. 1. The transmitter (encoder) knows at time i the message m and the feedback samples z^{i-1}. The output of the encoder at time i is denoted by x_i and it is a function of the message and the feedback. The channel is an FSC and the output of the channel y_i enters the decoder (receiver). The feedback z_i is a known time-invariant deterministic function of the current output of the channel y_i. For example, z_i could equal y_i (perfect feedback), or a quantized version of y_i, or null (no feedback). The encoders receive the feedback samples with one unit delay.

For some families of channels [4]–[7] we have the capacity as a limit of directed information, i.e.,

$$C = \lim_{N\to\infty} \frac{1}{N} \max_{Q(x^N||z^{N-1})} I(X^N \to Y^N), \tag{7}$$

where the limit can be shown to exist. For instance, in [4], [7] we have that for any FSC of the form

$$P(y_i, s_i|x_i, s_{i-1}) = p(s_i|s_{i-1})P(y_i|x_i, s_{i-1}) \tag{8}$$

where the state process S_i is ergodic and stationary, the capacity is given by (7) and the limit exists. Furthermore, in [5] we have that the capacity is a limit of directed information for the case that the FSC is unifilar (the state is a function of the input, output and the previous state) and there is a positive probability of achieving any state. Kim [6] showed that for any channel of the form $y_i = f(s_{i-m}^i, x_{i-m}^i)$ where f is a deterministic function, m is a constant and s_i is a stationary and ergodic process the capacity is given by (7). In the next section we show that the source-channel separation theorem holds in the presence of feedback for all channels with feedback capacity given by (7).

IV. Point-to-point source-channel separation

In this section, we prove the optimality of source channel separation for the case of an ergodic source that is transmitted through a channel with a deterministic time-invariant feedback where its capacity is given by (7). Namely, we prove that in the communication setting presented in Fig. 3, the number of bits per channel use that can be transmitted and reconstructed within a given distortion is the same as in the communication setting of Fig 2. Here we use the source channel-separation theorem that is restricted to sources that ate stationary and ergodic. However, there exists a more general source-channel separation theorem given by Vembu, Verdú and Steinberg [8] that applies for nonstationary sources and channels.

Let us state the source-channel separation theorem as presented in [9, Chapter 7].

Theorem 1: Let $\epsilon > 0$ and $D \geq 0$ be given. Let $R(\cdot)$ be the rate distortion function of a discrete, stationary, ergodic source with respect to a single letter criterion. Then the source output can be reproduced with fidelity D at the receiving end of any channel if $C > R(D)$. Conversely, fidelity D is unattainable at the receiving end of any channel of capacity $C < R(D)$.

Remark: For the simplicity of the presentation we assumed one channel use per source symbol. Our derivation below extends to the general case where the average number of channel uses per letter is $\frac{\tau_s}{\tau_c}$, analogously as in [10, chapter 9].

The purpose of this section is to prove the theorem for a channel with time-invariant feedback, as shown in Fig. 2, for the cases where its capacity is given by (7). In the case of no feedback the proof of separation optimality is based on the data processing inequality which states that $I(U^N;V^N) \leq I(X^N;Y^N)$ because of the Markov form $U^N - X^N - Y^N - V^N$. However, the regular data processing inequality does not hold for the directed information, as was shown in [4], and therefore an explicit derivation of the inequality $I(U^N;V^N) \leq I(X^N \to Y^N)$ is needed.

Proof: The direct proof, namely that if $C > R(D)$ it is possible to reproduce the source with fidelity D is the same as for the case without feedback [9, Theorem 7.2.6].

For the converse, namely that $R(D)$ has to be less or equal C, we use the fact that for any i, the Markov chain $U^N - X_i(U^N, Y^{i-1}) - Y_i$ holds.

$$\begin{aligned}
NR(D) &\overset{(a)}{\leq} I(U^N;V^N) \\
&\overset{(b)}{\leq} I(U^N;Y^N) \\
&\overset{(c)}{=} \sum_{i=1}^{N} I(U^N;Y_i|Y^{i-1}) \\
&= \sum_{i=1}^{N} H(Y_i|Y^{i-1}) - H(Y_i|U^N,Y^{i-1}) \\
&\overset{(d)}{=} \sum_{i=1}^{N} H(Y_i|Y^{i-1}) - H(Y_i|U^N,Y^{i-1},X^i) \\
&\overset{(e)}{=} \sum_{i=1}^{N} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},X^i) \\
&= \sum_{i=1}^{N} I(Y_i;X^i|Y^{i-1}) \\
&= I(X^N \to Y^N) \\
&\overset{(f)}{\leq} N(C). \qquad (9)
\end{aligned}$$

Inequality (a) follows from the converse for rate distortion [9, Theroem 7.2.5]. Inequality (b) follows from the data processing inequality because $U^N - Y^N - V^N$ form a Markov chain. Equality (c) follows from the chain rule. Inequality (d) follows from the fact that X_i is a deterministic function of (U^N, Y^{i-1}). Inequality (e) follows from the Markov chain $U^N - X_i(U^N,Y^{i-1}) - Y_i$. Finally, inequality (f) follows from the converse of capacity of channels with feedback that their capacity is given by (7). ■

V. The MAC model

The MAC setting consists of two senders and one receiver. Each sender $l \in \{1,2\}$ chooses an index w_l uniformly from the set $\{1,...,2^{nR_l}\}$ and independently of the other sender. The input to the channel from encoder l is denoted by $\{X_{l1}, X_{l2}, X_{l3}, ...\}$, and the output of the channel is denoted by $\{Y_1, Y_2, Y_3, ...\}$.

A code with feedback consists of two encoding functions $g_l : w_l \times \mathcal{Z}_l^{n-1} \to \mathcal{X}_l^n,\ l = 1,2$, where the kth coordinate of $x_l^n \in \mathcal{X}_l^n$ is given by the function

$$x_{lk} = g_{lk}(w_l, z_l^{k-1}), \qquad k = 1,2,\ldots,n, \quad l = 1,2. \quad (10)$$

and a decoding function,

$$g : \mathcal{Y}^n \to \{1,...,2^{nR_1}\} \times ... \times \{1,...,2^{nR_m}\}. \quad (11)$$

The *average probability of error* for $((2^{nR_1}, 2^{nR_2}, n)$ code is defined as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1,w_2} \Pr\{g(Y^n) \neq (w_1,w_2) | (w_1,w_2) \text{ sent}\}. \quad (12)$$

A rate (R_1,R_2) is said to be *achievable* for the MAC if there exists a sequence of $((2^{nR_1}, 2^{nR_2}), n)$ codes with $P_e^{(n)} \to 0$. The *capacity region* of MAC is the closure of the set of achievebale (R_1,R_2) rates.

Here define discrete MAC, called *Multiplexer followed by a point-to-point channel*. This MAC can be decomposed into two components as shown in Fig. 4. The first component is a MAC that can behave as a multiplexer and the second component is a point-to-point channel. The definitions of those components are the following:

Definition 1: A MAC behaves as *a multiplexer* if the inputs and the output have common alphabets and for all $m \in 1,...,M$ there exits a choice of input symbols for all senders except sender m, such that the output is the m input, i.e. $Y = X_m$.

An example of a multiplexer-MAC for the Binary case, is a MAC whose output is one of and/or/xor of the inputs. For a general alphabet q those operations could be max/min/addition-mod-q.

Theorem 2: The capacity region of a multiplexer MAC followed by a point-to-point channel with a time invariant feedback to all encoders, as shown in Fig. 4, is

$$\sum_{m=1}^{M} R_m \leq C \quad (13)$$

where C is the capacity of the point-to-point channel with the time invariant feedback $z_{i-1}(y_{i-1})$.

Proof: The achievability is proved simply by time sharing. At each time, only one selected user sends information

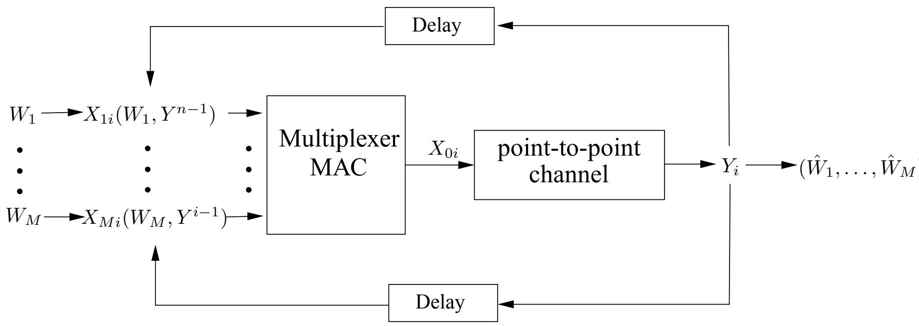

Fig. 4. Discrete MAC that can be decomposed into two parts. The first part is a MAC behaves as a multiplexer and the second part is a point-to-point channel

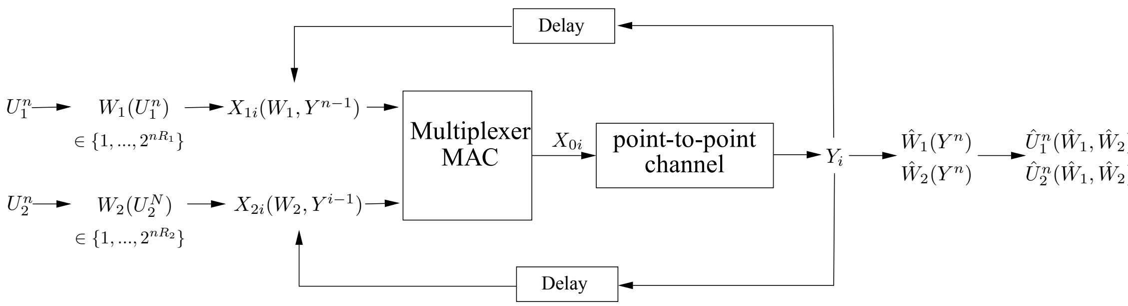

Fig. 5. Source-channel coding separation in a discrete Multiplexer followed by a point-to-point channel.

and the other users send a constant input that insures that the output is the input of the selected user.

The converse is based on the fact that the maximum rate that can be transmitted through the point-to-point channel is C and it is an upper bound sum-rate of multiplexer-MAC. If it wouldn't be an upper bound for the multiplexer-MAC, we could build a fictitious Multiplexer-MAC before the point-to-point channel and achieve by that a higher rate than C, which is its capacity, and hence we reached a contradiction. ■

Theorem 2 will be used in the next section for deriving results on source-channel separation. However, it can also be used for finding capacities of some MACs. For instance, using Theorem 2 and a result by Alajaji [11], [12] that states that feedback does not increase capacity of an additive mod-q channel, where q is the size of the input and output alphabets, we can deduce that feedback does not increase the capacity of an additive mod-q channel.

VI. Source-channel coding separation for multiplexer followed by a point-to-point channel

The source channel separation theorem does not hold for a MAC even for a memoryless channel without feedback [2]. However, here we show that for the case where the MAC is a discrete Multiplexer followed by a channel, it holds.

We want to send the sequence of symbols U_1^n, U_2^n over the MAC, so that the receiver can reconstruct the sequence. To do this we can use a joint source-channel coding scheme where we send through the channel the symbols $x_{1,i}(u_1^n, z^{i-1})$ and $x_{2,i}(u_2^n, z^{i-1})$. The receiver looks at his received sequence Y^n and makes an estimate $\hat{U}_1^n, \hat{U}_2^n$. The receiver makes an error if $\hat{U}_1^n \neq U_1^n$ or if $\hat{U}_2^n \neq U_2^n$, i.e., the probability of error $P_e^{(n)}$ is $P_e^{(n)} = \Pr((\hat{U}_1^n, \hat{U}_2^n) \neq (\hat{U}_1^n, \hat{U}_2^n))$.

Theorem 3: (*Source-channel coding theorem for a Multiplexer followed by a channel.*) Let $(U_1, U_2)_{n\geq 1}$ be a finite alphabet, ergodic and stationary stochastic processes and let the MAC channel be a Multiplexer followed by a point-to-point channel with a time invariant feedback with capacity $C = \lim_{N\to\infty} \frac{1}{N} \max_{Q(x^n||z^{n-1})} I(X^n \to Y^n)$ (e.g. memoryless channel, indecomposable FSC without feedback, stationary and ergodic Markovian channel). For the source and the MAC described above:
(*direct part:*) If $H(\mathcal{U}_1, \mathcal{U}_2) < C$, where $H(\mathcal{U}_1, \mathcal{U}_2)$ is the entropy rate of the sources and C is the capacity of the point-to-point channel with a time-invariant feedback, then there exists a source-channel code with $P_e^{(n)} \to 0$.
(converse part:) If $H(\mathcal{U}_1, \mathcal{U}_2) > C$, then the probability of error is bounded away from zero, and it is not possible to send the process over the channel with arbitrarily low probability of error.

Proof: The achievability is a straightforward consequence of the Slepian-Wolf result for Ergodic and stationary processes [13] and the achievability of the multiplexer followed by a point-to-point channel. First, we encode the sources by using the Sepian-Wolf achievaiblity scheme where we assign every u_1^n to one of 2^{nR_1} bins according to a uniform distribution on $\{1, ..., 2^{nR_1}\}$ and independently we assign every u_2^n to one of 2^{nR_2} bins according to a uniform distribution on $\{1, ..., 2^{nR_2}\}$. Second, we encode the bins as if they were messages, as shown in Fig. 5.

For the converse part, take an arbitrary scheme and consider

$$\begin{aligned}
&H(U_1^n, U_2^n) \\
&\stackrel{(a)}{\leq} \quad I(U_1^n, U_2^n; \hat{U}_1^n, \hat{U}_2^n) + n\epsilon_n \\
&\stackrel{(b)}{\leq} \quad I(U_1^n, U_2^n; Y^n) + n\epsilon_n \\
&= \quad H(Y^n) - H(Y^n|U_1^n, U_2^n) + n\epsilon_n \\
&= \quad \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|U_1^n, U_2^n, Y^{i-1}) + n\epsilon_n \\
&\stackrel{(c)}{=} \quad \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|U_1^n, U_2^n, Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&\stackrel{(d)}{=} \quad \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&= \quad \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\
&= \quad \sum_{i=1}^{n} I(X_1^i, X_2^i; Y_i|Y^{i-1}) + n\epsilon_n \\
&\stackrel{(e)}{\leq} \quad \sum_{i=1}^{n} I(X_0^i; Y_i|Y^{i-1}) + n\epsilon_n \\
&= \quad I(X_0^n \to Y^n) + n\epsilon_n \\
&\leq \quad \max_{Q(x_0^n||z^{n-1})} I(X_0^n \to Y^n) + n\epsilon_n \qquad (14)
\end{aligned}$$

Inequality (a) is due to Fano's inequality where $n\epsilon_n = 1 + nP_e^{(n)}|\mathcal{U}_1||\mathcal{U}_2|$. Inequality (b) follows from the data processing inequality because $(U_1^N, U_2^N) - Y^N - (\hat{U}_1^N, \hat{U}_2^N)$ form a Markov chain. Equality (c) is due to the fact that for a given code X_1^i is a deterministic function of U_1^n, Y^{i-1} and similarly X_2^i is a deterministic function of U_2^n, Y^{i-1}. Equality (d) is due to the Markov chain $(U_1^N, U_2^N) - (X_1^i, X_2^i, Y^{i-1}) - Y_i$. The notation $X_{0,i}$ denotes the output of the multiplexer which is also the input to the point-to-point channel at time i. The inequality in (e) is due to data processing inequality which is a consequence of the fact that given Y^{i-1} we have the Markov chain $X_1^i, X_2^i - X_0^i - Y_i$.

Suppose now that the sequence of codes satisfies $P_e^{(n)} \to 0$, which implies that $\epsilon_n \to \infty$. Then, by dividing both sides of (14) by n, and taking the limit $n \to \infty$, and recalling that $C = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||z^{n-1})} I(X^n; Y^n)$ we have

$$\frac{1}{n} H(U_1^n, U_2^n) \leq C. \qquad (15)$$

■

Acknowledgment

The authors would like to thank Amos Lapidoth and Andrea Goldsmith for helpful discussions.

References

[1] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, 1948.

[2] A. A. El Gamal T. M. Cover and M. Salehi. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inform. Theory*, 26:648–657, 1980.

[3] J. Massey. Causality, feedback and dircted information. *Proc. Int. Symp. Information Theory Application (ISITA-90)*, pages 303–305, 1990.

[4] H. H. Permuter, T. Weissman, and A. J. Goldsmith. Finite state channels with time-invariant deterministic feedback. Submitted to IEEE Trans. Inform. Theory. Availble at arxiv.org/pdf/cs.IT/0608070, Sep 2006.

[5] H. H. Permuter, P. W. Cuff, B. Van-Roy, and T. Weissman. Capacity of the trapdoor channel with feedback. Submitted to IEEE Trans. Inform. Theory. Availble at arxiv.org/cs.IT/0610047, Sep 2006.

[6] Y.H. Kim. A coding theorem for a class of stationary channels with feedback. Submitted to IEEE Trans. Inform. Theory. Availble at arxiv.org/cs.IT/0701041, Jan 2007.

[7] H. H. Permuter and T. Weissman. Capacity region of the finite-state multiple access channel with and without feedback. in preparation, Jun. 2007.

[8] S. Vembu, S. Verd, and Y. Steinberg. The source-channel separation theorem revisited. *IEEE Trans. Inform. Theory*, 41:44, 1995.

[9] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood, NJ, 1971.

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

[11] F. Alajaji and T. Fuja. Effect of feedback on the capacity of discrete additive channels with memory. In *Proceedings ISIT94*, Norway, 1994. IEEE.

[12] F. Alajaji. Feedback does not increase the capacity of discrete channels with additive noise. *IEEE Trans. Inform. Theory*, 41:546–549, 1995.

[13] T. M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Trans. Inform. Theory*, 22:226–228, 1975.