Capacity of the Trapdoor Channel With Feedback

Haim Permuter, *Student Member, IEEE*, Paul Cuff, *Student Member, IEEE*, Benjamin Van Roy, *Member, IEEE*, and Tsachy Weissman, *Senior Member, IEEE*

***Abstract*—We establish that the feedback capacity of the trapdoor channel is the logarithm of the golden ratio and provide a simple communication scheme that achieves capacity. As part of the analysis, we formulate a class of dynamic programs that characterize capacities of unifilar finite-state channels. The trapdoor channel is an instance that admits a simple closed-form solution.**

***Index Terms*—Bellman equation, chemical channel, constrained coding, directed information, feedback capacity, golden-ratio, infinite-horizon dynamic program, trapdoor channel, value iteration.**

I. Introduction

DAVID Blackwell, who has done fundamental work both in information theory and in stochastic dynamic programming, introduced the trapdoor channel in 1961 [1] as a "simple two-state channel." The channel is depicted in Fig. 1, and a detailed discussion of this channel appears in an information theory book by Ash [2], where indeed the channel is shown on the cover of the book.

The channel behaves as follows. Balls labeled "0" or "1" are used to communicate through the channel. The channel starts with a ball already in it. To use the channel, a ball is inserted into the channel by the transmitter, and the receiver receives one of the two balls in the channel with equal probability. The ball that does not exit the channel remains inside for the next channel use.

Another appropriate name for this channel is *chemical channel*. This name suggests a physical system in which the concentrations of chemicals are used to communicate, such as might be the case in some cellular biological systems as shown by Berger [3]. The transmitter adds molecules to the channel, and the receiver samples molecules randomly from the channel. The trapdoor channel is the most basic realization of this type of channel; it has only two types of molecules, and there are only three possible concentrations, $(0, 0.5, 1)$, or, equivalently, only one molecule remains in the channel between uses.

Although the trapdoor channel is very simple to describe, its capacity has been an open problem for over 45 years [1]. The zero-error capacity was found by Ahlswede *et al.* [4], [5] to be 0.5 bits per channel use. More recently, Kobayashi and Morita [6] derived a recursion for the conditional probabilities of output sequences of length n given the input sequences and used it to show that the capacity of this channel is strictly larger than 0.5 bits. Ahlswede and Kaspi [4] considered two modes of the channel called the *permuting jammer* channel and the *permuting relay* channel. In the first mode, there is a jammer in the channel that attempts to frustrate the message sender by selective release of balls in the channel. In the second mode, where the sender is in the channel, a helper supplies balls of a fixed sequence at the input, and the sender is restricted to permuting this sequence. The helper collaborates with the message sender in the channel to increase his ability to transmit distinct messages to the receiver. Ahlswede and Kaspi [4] gave answers for specific cases of both situations, and Kobayashi [7] established the answer to the general permuting relay channel. Additional results for specific cases of the permuting jammer channel can be found in [8], [9].

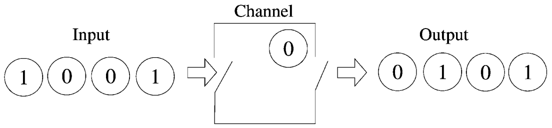


Fig. 1. The trapdoor (chemical) channel.

In this paper, we consider the trapdoor channel with feedback. We derive the feedback capacity of the trapdoor channel by solving an equivalent dynamic programming problem. Our work consists of two main steps. The first step is formulating the feedback capacity of the trapdoor channel as an infinite-horizon dynamic program, and the second step is finding explicitly the exact solution of that program.

Formulating the feedback capacity problem as a dynamic program appeared in Tatikonda's thesis [10] and in work by Yang, Kavčić, and Tatikonda [11] [12], Chen and Berger [13], and recently in a work by Tatikonda and Mitter [14]. Yang *et al.* have shown in [11] that if a channel has a one-to-one mapping between the input and the state, it is possible to formulate feedback capacity as a dynamic programming problem and to find an approximate solution by using the value iteration algorithm [15]. The authors of [11] have also formulated in [12] the feedback capacity of a stationary additive Gaussian-noise channel with a rational noise power spectrum of finite order[1] as a dynamic program. Chen and Berger [13] showed that if the state of the channel is a function of the output, then it is possible to formulate the feedback capacity as a dynamic program with a finite number of states.

Manuscript received October 8, 2006; revised January 25, 2008. This work was supported by the National Science Foundation (NSF) under Grants CCR-0311633, CCF-0515303, IIS-0428868, and the NSF CAREER Grant. The material in this paper was presented in part at the 44th Annual Allerton Conference on Communications, Control, and Computers, Monticello, IL, September 2006, and the IEEE International Symposium on Information Theory, Nice, France, June 2007

H. Permuter, P. Cuff, and B. Van Roy are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: haim1@stanford.edul; pcuff@stanford.edu; bvr@stanford.edu).

T. Weissman is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 3200, Israel He is also with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. (e-mail: tsachy@stanford.edu).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2008.924681

[1] In subsequent work, Kim [16], [17] showed that the optimal input distribution for this family of channels is stationary, which was a long-standing conjecture.

0018-9448/$25.00 © 2008 IEEE

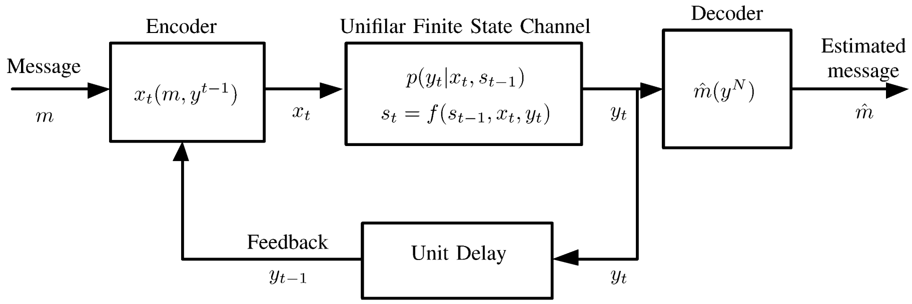


Fig. 2. Unifilar FSC with feedback.

TABLE I
THE PROBABILITY OF THE OUTPUT y_t GIVEN THE INPUT x_t AND THE STATE s_{t-1}

x_t	s_{t-1}	$p(y_t = 0\|x_t, s_{t-1})$	$p(y_t = 1\|x_t, s_{t-1})$
0	0	1	0
0	1	0.5	0.5
1	0	0.5	0.5
1	1	0	1

Our work provides the dynamic programming formulation and a computational algorithm for finding the feedback capacity of a family of channels called unifilar finite state channels (FSCs), which include the channels considered in [11], [13]. We use value iteration [15] to find an approximate solution and to generate a conjecture for the exact solution, and the Bellman equation [18] to verify the optimality of the conjectured solution. As a result, we are able to show that the feedback capacity of the trapdoor channel is $\log \phi$, where ϕ is the golden ratio, $\frac{1+\sqrt{5}}{2}$. In addition, we present a simple encoding/decoding scheme that achieves this capacity.

The remainder of the paper is organized as follows. Section II defines the channel setting and the notation throughout the paper. Section III states the main results of the paper. Section IV presents the capacity of a unifilar FSC in terms of directed information. Section V introduces the dynamic programming framework and shows that the feedback capacity of the unifilar FSC can be characterized as the optimal average reward of a dynamic program. Section VI shows an explicit solution for the capacity of the trapdoor channel by using the dynamic programming formulation. Section VII discusses a simple communication scheme that achieves the capacity of the trapdoor channel with feedback, and Section VIII concludes this work.

II. CHANNEL MODELS AND PRELIMINARIES

We use subscripts and superscripts to denote vectors in the following ways: $x^j = (x_1, \ldots, x_j)$ and $x_i^j = (x_i, \ldots, x_j)$ for $i \leq j$. Moreover, we use lower case x to denote sample values, upper case X to denote random variables, calligraphic letter $\mathcal{X}$ to denote the alphabet, and $|\mathcal{X}|$ to denote the cardinality of the alphabet. The probability distributions are denoted by p when the arguments specify the distribution, e.g., $p(x|y) = p(X = x|Y = y)$. In this paper, we consider only channels for which the input, denoted by $\{X_1, X_2, \ldots\}$, and the output, denoted by $\{Y_1, Y_2, \ldots\}$, are from finite alphabets, $\mathcal{X}$ and $\mathcal{Y}$, respectively. In addition, we consider only the family of FSC known as unifilar channels as discussed by Ziv [19]. An FSC is a channel that, for each time index, has one of a finite number of possible states, s_{t-1}, and has the property that $p(y_t, s_t|x^t, s^{t-1}, y^{t-1}) = p(y_t, s_t|x_t, s_{t-1})$. A *unifilar* FSC also has the property that the state s_t is deterministic given (s_{t-1}, x_t, y_t):

Definition 1: An FSC is called a *unifilar* FSC if there exists a time-invariant function $f(\cdot)$ such that the state evolves according to the equation

$$s_t = f(s_{t-1}, x_t, y_t). \tag{1}$$

We also define a *connected* FSC as follows.

Definition 2: We say that an FSC is connected if for any state s there exists an integer $T(s)$ and an input distribution of the form $\{p(x_t|s_{t-1})\}_{t=1}^{T(s)}$ that may depend on s, such that the probability that the channel reaches s from any starting state s', in less than $T(s)$ time steps, is positive. That is

$$\sum_{t=1}^{T(s)} \Pr(S_t = s|S_0 = s') > 0, \quad \forall s \in \mathcal{S}, \forall s' \in \mathcal{S}. \tag{2}$$

We assume a communication setting that includes feedback as shown in Fig. 2. At time t, the transmitter (encoder) knows the message m and the feedback samples y^{t-1}. The output of the encoder at time t is denoted by x_t and is a function of the message and the feedback. The channel is a unifilar FSC and the output of the channel y_t enters the decoder (receiver). The encoder receives the feedback sample with one unit delay.

A. Trapdoor Channel Is a Unifilar FSC

The state of the trapdoor channel, which is described in the Introduction and shown in Fig. 1, is the ball, 0 or 1, that is in the channel before the transmitter transmits a new ball. Let $x_t \in \{0, 1\}$ be the ball that is transmitted at time t, and $s_{t-1} \in \{0, 1\}$ be the state of the channel when ball x_t is transmitted. The probability of the output y_t given the input x_t and the state of the channel s_{t-1} is shown in Table I.

The trapdoor channel is a unifilar FSC. It has the property that the next state s_t is a deterministic function of the state s_{t-1}, the input x_t, and the output y_t. For a feasible tuple, (x_t, y_t, s_{t-1}), the next state is given by the equation

$$s_t = s_{t-1} \oplus x_t \oplus y_t \tag{3}$$

where $\oplus$ denotes the binary XOR operation.

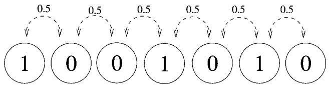


Fig. 3. The trapdoor channel as a permuting channel. Going from left to right, there is a probability of one half that two adjacent bits switch places.

B. Trapdoor Channel Is a Permuting Channel

It is interesting to note, although not consequential in this paper, that the trapdoor channel is a permuting channel [20], where the output is a permutation of the input (Fig. 3). At each time t, a new bit is added to the sequence and the channel switches the new bit with the previous one in the sequence with probability 0.5.

III. Main Results

- The capacity of the trapdoor channel with feedback is
$$C = \log \frac{\sqrt{5}+1}{2}. \tag{4}$$
Furthermore, there exists a simple capacity achieving scheme which will be presented in Section VII.
- The problem of finding the capacity of a connected unifilar channel (Fig. 2) can be formulated as an average-reward dynamic program, where the state of the dynamic program is the probability mass function over the states conditioned on prior outputs, and the action is the stochastic matrix $p(x|s)$. By finding a solution to the average-reward Bellman equation, we find the exact capacity of the channel.
- As a byproduct of our analysis, we also derive a closed-form solution to an infinite horizon average-reward dynamic program with a continuous state-space.

IV. The Capacity Formula for a Unifilar Channel With Feedback

The main goal of this section is to prove the following theorem, which allows us to formulate the problem as a dynamic program.

Theorem 1: The feedback capacity of a connected unifilar FSC when initial state s_0 is known at the encoder and decoder can be expressed as

$$C_{FB} = \sup_{\{p(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} I(X_t,S_{t-1};Y_t|Y^{t-1}) \tag{5}$$

where $\{p(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}$ denotes the set of all distributions such that $p(x_t|y^{t-1},x^{t-1},s^{t-1}) = p(x_t|s_{t-1},y^{t-1})$ for $t = 1,2,\ldots.$

Theorem 1 is a direct consequence of Theorem 3 and (26) in Lemma 4, which are proved in this section.

For any finite-state channel with perfect feedback, as shown in Fig. 2, the capacity was shown in [21], [22] to be bounded as

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1})}\max_{s_0} I(X^N\to Y^N|s_0) &\geq C_{FB}\\ \geq \lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1})}&\min_{s_0} I(X^N\to Y^N|s_0).\end{aligned} \tag{6}$$

The term $I(X^N \to Y^N)$ is the *directed information*[2] defined originally by Massey in [29] as

$$I(X^N\to Y^N) \triangleq \sum_{t=1}^{N} I(X^t;Y_t|Y^{t-1}). \tag{7}$$

The initial state is denoted as s_0 and $p(x^N\|y^{N-1})$ is the causally conditional distribution defined in [21], [26] as

$$p(x^N\|y^{N-1}) \triangleq \prod_{t=1}^{N} p(x_t|x^{t-1},y^{t-1}). \tag{8}$$

The directed information in (6) is under the distribution of $p(x^n,y^n)$ which is uniquely determined by the causal conditioning $p(x^N\|y^{N-1})$ and by the channel.

In our communication setting, we are assuming that the initial state is known both to the decoder and to the encoder. This assumption allows the encoder to know the state of the channel at any time t because s_t is a deterministic function of the previous state, input and output. In order to take into account this assumption, we use a trick of allowing a fictitious time epoch before the first actual use of the channel in which the input does not influence the output nor the state of channel, and the only thing that happens is that the output equals s_0 and is fed back to the encoder such that at time $t=1$ both the encoder and the decoder know the state s_0. Let $t=0$ be the fictitious time before starting the use of the channel. According to the trick, $Y_0 = S_0$, and the input X_0 can be chosen arbitrarily because it does not have any influence whatsoever. For this scenario, the directed information term in (6) becomes

$$I(X_0^N\to Y_0^N|s_0) = I(X^N\to Y^N|s_0). \tag{9}$$

The input distribution becomes

$$p(x_0^N\|\{s_0,y^{N-1}\}) = p(x^N\|y^{N-1},s_0), \tag{10}$$

where $p(x^N\|y^{N-1},s_0)$ is defined as

$$p(x^N\|y^{N-1},s_0) \triangleq \prod_{t=1}^{N} p(x_t|x^{t-1},y^{t-1},s_0).$$

Therefore, the capacity of a unifilar channel with feedback for which the initial state, s_0, is known both at the encoder and the decoder is bounded as

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\max_{s_0} I(X^N\to Y^N|s_0) &\geq C_{FB}\\ \geq \lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}&\min_{s_0} I(X^N\to Y^N|s_0).\end{aligned} \tag{11}$$

[2] In addition to feedback capacity, directed information has recently been used in rate distortion [23], [24], [25], network capacity [26], [27] and computational biology [28].

Lemma 2: If the finite-state channel is connected, then for any input distribution $p_1(x^N\|y^{N-1},s_0)$ and any s_0', there exists an input distribution $p_2(x^N\|y^{N-1},s_0')$ such that

$$\frac{1}{N}|I_{p_1}(X^N\to Y^N|s_0)-I_{p_2}(X^N\to Y^N|s_0')|\le\frac{c}{N} \tag{12}$$

where c is a constant that does not depend on N,s_0,s_0'. The term $I_{p_1}(X^N\to Y^N|s_0)$ denotes the directed information induced by the input distribution $p_1(x^N\|y^{N-1},s_0)$, where s_0 is the initial state. Similarly, the term $I_{p_2}(X^N\to Y^N|s_0')$ denotes the directed information induced by the input distribution $p_2(x^N\|y^{N-1},s_0')$ where s_0' is the initial state.

Proof: Construct $p_2(x^N\|y^N,s_0')$ as follows. Use an input distribution, which has a positive probability of reaching s_0 in T time epochs, until the time that the channel first reaches s_0. Such an input distribution exists because the channel is connected. Denote the first time that the state of the channel equals s_0 by L. After time L, operate exactly as p_1 would (had time started then). Namely, for $t>L$

$$p_2(x_t|x^{t-1},y^{t-1},s_0)=p_1(x_{t-L}|x^{t-L-1},y^{t-L-1},s_0).$$

Then

$$\begin{aligned}
&\frac{1}{N}|I_{p_1}(X^N\to Y^N|s_0)-I_{p_2}(X^N\to Y^N|s_0')|\\
&\overset{(a)}{\le}\ \frac{1}{N}|I_{p_1}(X^N\to Y^N|s_0)\\
&\qquad -I_{p_2}(X^N\to Y^N|L,s_0')|+\frac{1}{N}H(L)\\
&\overset{(b)}{=}\ \frac{1}{N}\left|\sum_{l=1}^{\infty}p(L=l)I_{p_1}(X^N\to Y^N|s_0)\right.\\
&\qquad -\sum_{l=1}^{\infty}p(L=l)\left(I_{p_2}(X_l^N\to Y_l^N|s_l)\right.\\
&\qquad \left.\left.+I_{p_2}(X^{\min(l,N)}\to Y^{\min(l,N)}|s_l,s_0')\right)\right|+\frac{1}{N}H(L)\\
&\overset{(c)}{\le}\ \frac{1}{N}\left|\sum_{l=1}^{\infty}p(L=l)I_{p_1}(X^N\to Y^N|s_0)\right.\\
&\qquad \left.-\sum_{l=1}^{\infty}p(L=l)I_{p_2}(X_l^N\to Y_l^N|s_l)\right|\\
&\qquad +\frac{1}{N}\left|\sum_{l=1}^{\infty}p(L=l)I_{p_2}\right.\\
&\qquad \left.\times\ (X^{\min(l,N)}\to Y^{\min(l,N)}|s_l,s_0')\right|+\frac{1}{N}H(L)\\
&\overset{(d)}{\le}\ \frac{2}{N}\sum_{l=1}^{\infty}p(L=l)l\log|\mathcal{Y}|+\frac{1}{N}H(L)\\
&=\ \frac{1}{N}(2\log|\mathcal{Y}|\mathbb{E}[L]+H(L))
\end{aligned} \tag{13}$$

where

(a) follows from the triangle inequality and Lemma 3 in [21], which claims that for any arbitrary random variables (X^N,Y^N,S), the inequality $|I(X^N\to Y^N)-I(X^N\to Y^N|S)|\le H(S)$ always holds.

(b) follows from using the special structure of $p_2(x^N\|y^N,s_0')$.

(c) follows from the triangle inequality.

(d) follows from the fact that in the first absolute value, $N-l$ terms cancel and therefore only l terms remain where each one is bounded by $I(X^t;Y_t|Y^{t-1})\le\log|\mathcal{Y}|$. In the second absolute value there are $\min(l,N)\le l$ terms, also bounded by $\log|\mathcal{Y}|$.

The proof is completed by noting that $H(L)$ and $E(L)$ are upper-bounded, respectively, by $H(\tilde{L})$ and $E(\tilde{L})$, where $\lfloor\tilde{L}/T\rfloor\sim$ Geometric (p), and p is the minimum probability of reaching s_0 in less than T steps from any state $s\in\mathcal{S}$. Because the random variable $\lfloor\tilde{L}/T\rfloor$ has a geometric distribution, $H(\tilde{L})$ and $\mathbb{E}[\tilde{L}]$ are finite and, consequently, so are $H(L)$ and $E(L)$.□

Theorem 3: The feedback capacity of a connected unifilar FSC, when the initial state is known at the encoder and decoder, is given by

$$C_{FB}=\lim_{N\to\infty}\frac{1}{N}\max_{\{p(x_t|s_{t-1},y^{t-1})\}_{t=1}^N}\sum_{t=1}^{N}I(X_t,S_{t-1};Y_t|Y^{t-1}) \tag{14}$$

Proof: The proof of the theorem contains four main equalities, which are proven separately.

$$\begin{aligned}
C_{FB}&=\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0}I(X^N\to Y^N|s_0) &(15)\\
&=\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}I(X^N\to Y^N|S_0) &(16)\\
&=\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\sum_{t=1}^{N}I(X_t,S_{t-1};Y_t|Y^{t-1}) &(17)\\
&=\lim_{N\to\infty}\frac{1}{N}\max_{\{p(x_t|s_{t-1},y^{t-1})\}_{t=1}^N}\sum_{t=1}^{N}I(X_t,S_{t-1};Y_t|Y^{t-1}). &(18)
\end{aligned}$$

Proof of Equality (15) and (16): As a result of Lemma 2

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}I(X^N\to Y^N|S_0)\\
&\overset{(a)}{=}\ \lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\sum_{s_0}p(s_0)I(X^N\to Y^N|s_0)\\
&\overset{(b)}{=}\ \lim_{N\to\infty}\frac{1}{N}\sum_{s_0}p(s_0)\max_{p(x^N\|y^{N-1},s_0)}I(X^N\to Y^N|s_0)\\
&\overset{(c)}{=}\ \lim_{N\to\infty}\frac{1}{N}\min_{s_0}\max_{p(x^N\|y^{N-1},s_0)}I(X^N\to Y^N|s_0) &(19)\\
&\overset{(d)}{=}\ \lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0}I(X^N\to Y^N|s_0). &(20)
\end{aligned}$$

where

(a) follows from the definition of conditional entropy.

(b) follows from the exchange between the summation and the maximization. The exchange is possible because maximization is over causally conditional distributions that depend on s_0.

(c) follows from Lemma 2.

(d) follows from the observation that the distribution $p^*(x^N\|y^{N-1},s_0)$ that achieves the maximum in (19) and in (20) is the same: $p^*(x^N\|y^{N-1},s_0)=\arg\max_{p(x^N\|y^{N-1},s_0)}I(X^N\to Y^N|s_0)$. This observation allows us to exchange the order of the minimum and the maximum.

Equations (19) and (20) can be repeated also with $\max_{s_0}$ instead of $\min_{s_0}$, and hence we get

$$\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)} I(X^N\to Y^N|S_0)$$
$$=\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\max_{s_0} I(X^N\to Y^N|s_0). \quad (21)$$

By using (20) and (21), we get that the upper bound and the lower bound in (11) are equal, and therefore (15) and (16) hold.

Proof of Equality (17): Using the property that the next state of the channel is a deterministic function of the input, output, and current state, we get

$$\begin{aligned} I(X^N\to Y^N|S_0) &= \sum_{t=1}^{N} I(X^t;Y_t|Y^{t-1},S_0)\\ &= \sum_{t=1}^{N} H(Y_t|Y^{t-1},S_0)-H(Y_t|X^t,Y^{t-1},S_0)\\ &\overset{(a)}{=} \sum_{t=1}^{N} H(Y_t|Y^{t-1},S_0)\\ &\quad - H(Y_t|X^t,Y^{t-1},S_0,S^{t-1}(X^t,Y^{t-1},S_0))\\ &\overset{(b)}{=} \sum_{t=1}^{N} H(Y_t|Y^{t-1},S_0)-H(Y_t|X_t,S_{t-1},Y^{t-1},S_0)\\ &= \sum_{t=1}^{N} I(S_{t-1},X_t;Y_t|Y^{t-1},S_0). \end{aligned} \quad (22)$$

Equality (a) is due to the fact that s^{t-1} is a deterministic function of the tuple (x^t,y^{t-1},s_0). Equality (b) is due to the fact that $p(y_t|x^t,y^{t-1},s^{t-1},s_0)=p(y_t|x_t,y^{t-1},s_{t-1},s_0)$. By combining (16) and (22), we get (17).

Proof of Equality (18): It will suffice to prove by induction that if we have two input distributions $\{p_1(x_t|x^{t-1},y^{t-1},s_0)\}_{t\geq 1}$ and $\{p_2(x_t|x^{t-1},y^{t-1},s_0)\}_{t\geq 1}$ that induce the same distributions $\{p(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}$, then the distributions $\{p(s_{t-1},x_t,y^t)\}_{t\geq 1}$ are equal under both inputs. First, let us verify the equality for $t=1$

$$p(s_0,x_1,y_1)=p(s_0)p(x_1|s_0)p(y_1|s_0,x_1). \quad (23)$$

Since $p(s_0)$ and $p(y_1|s_0,x_1)$ are not influenced by the input distribution, and since $p(x_1|s_0)$ is equal for both input distributions, then $p(s_0,x_1,y_1)$ is also equal for both input distributions. Now, we assume that $p(s_{t-1},x_t,y^t)$ is equal under both input distributions, and we need to prove that $p(s_t,x_{t+1},y^{t+1})$ is also equal under both input distributions. The term $p(s_t,x_{t+1},y^{t+1})$, which can be written as

$$p(s_t,x_{t+1},y^{t+1})=p(s_t,y^t)p(x_{t+1}|s_t,y^t)p(y_{t+1}|x_{t+1},s_t). \quad (24)$$

First we notice that if $p(s_{t-1},x_t,y^t)$ is equal for both cases, then $p(s_{t-1},s_t,x_t,y^t)$ is necessarily equal for both cases because s_t is a deterministic function of the tuple (s_{t-1},x_t,y_t), and therefore both input distributions induce the same $p(s_t,y^t)$. The distribution $p(x_{t+1}|s_t,y^t)$ is the same under both input distributions by assumption, and $p(y_{t+1}|x_{t+1},s_t)$ does not depend on the input distribution. □

The next lemma shows that it is possible to switch between the limit and the maximization in the capacity formula. This is necessary for formulating the problem, as we do in the next section, as an average-reward dynamic program.

Lemma 4: For any FSC, the following equality holds:

$$\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0} I(X^N\to Y^N|s_0)$$
$$=\sup_{\{p(x_t|y^{t-1},x^{t-1},s_0)\}_{t\geq 1}}\liminf_{N\to\infty}\frac{1}{N}\min_{s_0} I(X^N\to Y^N|s_0). \quad (25)$$

And, in particular, for a connected unifilar FSC

$$\lim_{N\to\infty}\frac{1}{N}\max_{\{p(x_t|s_{t-1},y^{t-1})\}_{t=1}^{N}}\sum_{t=1}^{N} I(X_t,S_{t-1};Y_t|Y^{t-1})$$
$$=\sup_{\{p(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}}\liminf_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N} I(X_t,S_{t-1};Y_t|Y^{t-1}) \quad (26)$$

On the left-hand side of the equations appears $\lim$ because, as shown in [22], the limit exists due to the super-additivity property of the sequence.

Proof: We prove (25), which holds for any FSC. For the case of unifilar channel, the left-hand side of (25) is proven to be equal to the left-hand side of (26) in (15)–(18). By the same arguments as in (15)–(18), the right-hand side of (25) and (26) are also equal.

Define

$$\underline{C}_N \triangleq \frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0} I(X^N\to Y^N|s_0). \quad (27)$$

In order to prove that the equality holds, we will use two properties of $\underline{C}_N$ that were proved in [22, Theorem 13].

The first property is that $\underline{C}_N$ is a super additive sequence, namely, for any two positive integers n and l that sums to N

$$N\left[\underline{C}_N-\frac{\log|\mathcal{S}|}{N}\right]\geq n\left[\underline{C}_n-\frac{\log|\mathcal{S}|}{n}\right]+l\left[\underline{C}_l-\frac{\log|\mathcal{S}|}{l}\right]. \quad (28)$$

The second property, which is a result of the first, is that

$$\lim_{N\to\infty}\underline{C}_N=\sup_N \underline{C}_N \quad (29)$$

Now, consider

$$\begin{aligned} &\lim_{N\to\infty}\frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0} I(X^N\to Y^N|s_0)\\ &=\sup_N \underline{C}_N\\ &=\sup_N \frac{1}{N}\max_{p(x^N\|y^{N-1},s_0)}\min_{s_0} I(X^N\to Y^N|s_0)\\ &=\sup_N \frac{1}{N}\sup_{\{p(x^t|y^{t-1},x^{t-1},s_0)\}_{t\geq 1}}\min_{s_0} I(X^N\to Y^N|s_0)\\ &=\sup_{\{p(x^t|y^{t-1},x^{t-1},s_0)\}_{t\geq 1}}\sup_N \frac{1}{N}\min_{s_0} I(X^N\to Y^N|s_0)\\ &\geq \sup_{\{p(x^t|y^{t-1},x^{t-1},s_0)\}_{t\geq 1}}\liminf_N \frac{1}{N}\min_{s_0} I(X^N\to Y^N|s_0) \end{aligned} \quad (30)$$

The limit of the left-hand side of the equation in the lemma implies that, $\forall \epsilon > 0$ there exists $N(\epsilon)$ such that for all $n > N(\epsilon)$

$$\frac{1}{N}\max_{p(x^n\|y^{n-1},s_0)}\min_{s_0} I(X^N \to Y^N|s_0) \geq \sup_N C_N - \epsilon.$$

Let us choose $j > N(\epsilon)$, and let $p^*(x^j\|y^{j-1})$ be the input distribution that attains the maximum. Let us construct

$$\tilde{p}(x^t\|y^{t-1},s_0) = p^*(x_{t-j+1}^t\|y_{t-j+1}^{t-1}, s_{t-j}) \\ p^*(x_{t-2j+1}^{t-j}\|y_{t-2j+1}^{t-j-1}, s_{t-2j})\cdots. \quad (31)$$

Then, we get

$$\sup_{\{p(x^t|y^{t-1},x^{t-1},s_0)\}} \liminf_{N\to\infty}\frac{1}{N}\min_{s_0} I(X^N \to Y^N|s_0) \\ \geq \liminf_{N\to\infty}\frac{1}{N}\min_{s_0} I_{\tilde{p}}(X^N \to Y^N|s_0) \geq \sup_N C_N - \epsilon \quad (32)$$

where $I_{\tilde{p}}(X^N \to Y^N|s_0)$ is the directed information induced by the input $\tilde{p}(x^t\|y^{t-1},s_0)$ and the channel. The left inequality holds because $\tilde{p}(x^t\|y^{t-1},s_0)$ is only one possible input distribution among all $\{p(x^t\|y^{t-1},s_0)\}_{t=1}^{\infty}$. The right inequality holds because the special structure of $\tilde{p}(x^t\|y^{t-1},s_0)$ transforms the whole expression of normalized directed information into an average of infinite sums of terms that each term is directed information between blocks of length j. Because for each block the inequality holds, then it holds also for the average of the blocks. The inequality may not hold on the last block, but because we average over an increasing number of blocks, its influence diminishes. □

V. Feedback Capacity and Dynamic Programming

In this section, we characterize the feedback capacity of the unifilar FSC as the optimal average reward of a dynamic program. Further, we present the Bellman equation, which can be solved to determine this optimal average reward.

A. Dynamic Programs

Here we introduce a formulation for average-reward dynamic programs. Each problem instance is defined by a septuple $(\mathcal{Z},\mathcal{U},\mathcal{W},F,P_z,P_w,g)$. We will explain the roles of the components of this tuple.

We consider a discrete-time dynamic system evolving according to

$$z_t = F(z_{t-1},u_t,w_t), \quad t = 1,2,3,\ldots \quad (33)$$

where each *state* z_t takes values in a Borel space $\mathcal{Z}$, each *action* u_t takes values in a compact subset $\mathcal{U}$ of a Borel space, and each disturbance w_t takes values in a measurable space $\mathcal{W}$. The initial state z_0 is drawn from a distribution P_z. Each disturbance w_t is drawn from a distribution $P_w(\cdot|z_{t-1},u_t)$, which depends only on the state z_{t-1} and action u_t. All functions considered in this paper are assumed to be measurable, though we will not mention this each time we introduce a function or a set of functions.

The *history* $\Phi_t = (z_0,w_0,\ldots,w_{t-1})$ summarizes information available prior to selection of the tth action. The action u_t is selected by a function μ_t, which maps histories to actions. In particular, given a policy $\pi = \{\mu_1,\mu_2,\ldots\}$, actions are generated according to $u_t = \mu_t(\Phi_t)$. Note that given the history Φ_t and a *policy* $\pi = \{\mu_1,\mu_2,\ldots\}$, one can compute past states $z_1,\ldots,z_{t-1}$ and actions $u_1,\ldots,u_{t-1}$. A policy $\pi = \{\mu_1,\mu_2,\ldots\}$ is referred to as stationary if there is a function $\mu : \mathcal{Z} \mapsto \mathcal{U}$ such that $\mu_t(\Phi_t) = \mu(z_{t-1})$ for all t and Φ_t. With some abuse of terminology, we will sometimes refer to such a function μ itself as a stationary policy.

We consider an objective of maximizing average reward, given a bounded reward function $g : \mathcal{Z}\times\mathcal{U} \to \Re$. The average reward for a policy π is defined by

$$\rho_\pi = \liminf_{N\to\infty}\frac{1}{N}\mathbb{E}_\pi\left\{\sum_{t=0}^{N-1} g(Z_t,\mu_{t+1}(\Phi_{t+1}))\right\}$$

where the subscript π indicates that actions are generated by the policy $\pi = (\mu_1,\mu_2,\ldots)$. The optimal average reward is defined by

$$\rho^* = \sup_\pi \rho_\pi.$$

B. The Bellman Equation

An alternative characterization of the optimal average reward is offered by the Bellman equation. This equation offers a mechanism for verifying that a given level of average reward is optimal. It also leads to a characterization of optimal policies. The following result, which we will later use, encapsulates the Bellman equation and its relation to the optimal average reward and optimal policies.

Theorem 5: If $\rho \in \Re$ and a bounded function $h : \mathcal{Z} \mapsto \Re$ satisfy

$$\rho + h(z) = \sup_{u\in\mathcal{U}}\left(g(z,u) + \int P_w(dw|z,u)h(F(z,u,w))\right), \\ \forall z \in \mathcal{Z} \quad (34)$$

then $\rho = \rho^*$. Further, if there is a function $\mu : \mathcal{Z} \mapsto \mathcal{U}$ such that $\mu(z)$ attains the supremum for each z, then $\rho_\pi = \rho^*$ for $\pi = (\mu_0,\mu_1,\ldots)$ with $\mu_t(\Phi_t) = \mu(z_{t-1})$ for each t.

This result follows immediately from Theorem 6.1 of [18]. It shows that if there exists a (ρ,μ,h) that satisfies the Bellman equation, then ρ is the optimal average reward, μ is a stationary policy that achieves the optimum, and the policy depends on the history Φ_t only through the state z_{t-1}. It is convenient to define a dynamic programming operator T by

$$(Th)(z) = \sup_{u\in\mathcal{U}}\left(g(z,u) + \int P_w(dw|z,u)h(F(z,u,w))\right)$$

for all functions h. Then, Bellman's equation can be written as $\rho\mathbf{1}+h = Th$. It is also useful to define for each stationary policy μ an operator

$$(T_\mu h)(z) = g(z,\mu(z)) + \int P_w(dw|z,\mu(z))h(F(z,\mu(z),w)).$$

The operators T and T_μ obey some well-known properties. First, they are monotonic, i.e., for bounded functions h and $\overline{h}$ such that $h \leq \overline{h}, Th \leq T\overline{h}$ and $T_\mu h \leq T_\mu\overline{h}$. Second, they are nonexpansive with respect to the sup-norm, i.e., for bounded functions h and $\overline{h}, \|Th - T\overline{h}\|_\infty \leq \|h - \overline{h}\|_\infty$ and

$\|T_\mu h - T_\mu \overline{h}\|_\infty \leq \|h - \overline{h}\|_\infty$. Third, as a consequence of nonexpansiveness, T is continuous with respect to the sup-norm.[3]

C. Feedback Capacity as a Dynamic Program

We will now formulate a dynamic program such that the optimal average reward equals the feedback capacity of a unifilar channel as presented in Theorem 1. This entails defining the septuple $(\mathcal{Z}, \mathcal{U}, \mathcal{W}, F, P_z, P_w, g)$ based on properties of the unifilar channel and then verifying that the optimal average reward is equal to the capacity of the channel.

Let β_t denote the $|\mathcal{S}|$-dimensional vector of channel state probabilities given information available to the decoder at time t. In particular, each component corresponds to a channel state s_t and is given by $\beta_t(s_t) \triangleq p(s_t|y^t)$. We consider the states of the dynamic program to be $z_t = \beta_t$. Hence, the state space $\mathcal{Z}$ is the $|\mathcal{S}|$-dimensional unit simplex. Each action u_t is taken to be the matrix of conditional probabilities of the input x_t given the previous state s_{t-1} of the channel. Hence, the action space $\mathcal{U}$ is the set of stochastic matrices of dimension $|\mathcal{S}| \times |\mathcal{X}|$. The disturbance w_t is taken to be the channel output y_t. The disturbance space $\mathcal{W}$ is the output alphabet Y.

The initial state distribution P_z is concentrated at the prior distribution of the initial channel state s_0. Note that the channel state s_t is conditionally independent of the past given the previous channel state s_{t-1}, the input probabilities u_t, and the current output y_t. Hence, $\beta_t(s_t) = p(s_t|y^t) = p(s_t|\beta_{t-1}, u_t, y_t)$. More concretely, given a policy $\pi = (\mu_1, \mu_2, \ldots)$, $\beta_t(s_t)$ is given in (35) at the bottom of the page, where $\mathbf{1}(\cdot)$ is the indicator function. Note that $p(y_t|s_{t-1}, x_t)$ is given by the channel model. Hence, β_t is determined by β_{t-1}, u_t, and y_t, and, therefore, there is a function F such that $z_t = F(z_{t-1}, u_t, w_t)$.

The distribution of the disturbance w_t is

$$p(w_t|z^{t-1}, w^{t-1}, u^t) = p(w_t|z_{t-1}, u_t).$$

Conditional independence from z^{t-2} and w^{t-1} given z_{t-1} is due to the fact that the channel output is determined by the previous channel state and current input. More concretely

$$\begin{aligned}
p(w_t&|z^{t-1}, w^{t-1}, u^t)\\
&= p(y_t|\beta^{t-1}, y^{t-1}, u^t)\\
&= \sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|\beta^{t-1}, y^{t-1}, u^t)\\
&= \sum_{x_t, s_{t-1}} p(s_{t-1}|\beta_{t-1}, u_t)\\
&\quad \times p(x_t|s_{t-1}, \beta_{t-1}, u_t) p(y_t|x_t, s_{t-1}, \beta_{t-1}, u_t)\\
&= \sum_{x_t, s_{t-1}} p(s_{t-1}, x_t, y_t|\beta_{t-1}, u_t)\\
&= p(y_t|\beta_{t-1}, u_t)\\
&= p(w_t|z_{t-1}, u_t). \qquad (36)
\end{aligned}$$

Hence, there is a disturbance distribution $P_w(\cdot|z_{t-1}, u_t)$ that depends only on z_{t-1} and u_t.

We consider a reward of $I(Y_t; X_t, S_{t-1}|y^{t-1})$. Note that the reward depends only on the probabilities $p(x_t, y_t, s_{t-1}|y^{t-1})$ for all x_t, y_t and s_{t-1}. Further

$$\begin{aligned}
p(x_t&, y_t, s_{t-1}|y^{t-1})\\
&= p(s_{t-1}|y^{t-1}) p(x_t|s_{t-1}, y^{t-1}) p(y_t|x_t, s_{t-1})\\
&= \beta_{t-1}(s_{t-1}) u_t(s_{t-1}, x_t) p(y_t|x_t, s_{t-1}). \qquad (37)
\end{aligned}$$

Recall that $p(y_t|x_t, s_{t-1})$ is given by the channel model. Hence, the reward depends only on β_{t-1} and u_t.

Given an initial state z_0 and a policy $\pi = (\mu_1, \mu_2, \ldots)$, u_t and β_{t-1} are determined by y^{t-1}. Further, (X_t, S_{t-1}, Y_t) is conditionally independent of y^{t-1} given β_{t-1} and u_t as shown in (37). Hence

$$\begin{aligned}
g(z_{t-1}, u_t) &= I(Y_t; X_t, S_{t-1}|y^{t-1})\\
&= I(X_t, S_{t-1}; Y_t|\beta_{t-1}, u_t). \qquad (38)
\end{aligned}$$

It follows that the optimal average reward is

$$\rho^* = \sup_\pi \liminf_{N\to\infty} \frac{1}{N} \mathbb{E}_\pi \left[\sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t|Y^{t-1}) \right] = C_{FB}.$$

Table II summarizes the dynamic programming formulation. The table identifies the septuple $(\mathcal{Z}, \mathcal{U}, \mathcal{W}, F, P_z, P_w, g)$ for the feedback capacity problem, where the channel is a unifilar channel. In this formulation, the action u_t depends on the whole

$$\begin{aligned}
\beta_t(s_t) &= p(s_t|y^t)\\
&= \sum_{x_t, s_{t-1}} p(s_t, s_{t-1}, x_t|y^t)\\
&= \sum_{x_t, s_{t-1}} \frac{p(s_t, s_{t-1}, x_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}\\
&= \sum_{x_t, s_{t-1}} \frac{p(s_{t-1}|y^{t-1}) p(x_t|s_{t-1}, y^{t-1}) p(y_t|s_{t-1}, x_t) p(s_t|s_{t-1}, x_t, y_t)}{p(y_t|y^{t-1})}\\
&= \frac{\sum_{x_t, s_{t-1}} \beta_{t-1}(s_{t-1}) p(x_t|s_{t-1}, y^{t-1}) p(y_t|s_{t-1}, x_t) p(s_t|s_{t-1}, x_t, y_t)}{\sum_{x_t, s_t, s_{t-1}} \beta_{t-1}(s_{t-1}) p(x_t|s_{t-1}, y^{t-1}) p(y_t|s_{t-1}, x_t) p(s_t|s_{t-1}, x_t, y_t)}\\
&= \frac{\sum_{x_t, s_{t-1}} \beta_{t-1}(s_{t-1}) u_t(s_{t-1}, x_t) p(y_t|s_{t-1}, x_t) \mathbf{1}(s_t = f(s_{t-1}, x_t, y_t))}{\sum_{x_t, s_t, s_{t-1}} \beta_{t-1}(s_{t-1}) u_t(s_{t-1}, x_t) p(y_t|s_{t-1}, x_t) \mathbf{1}(s_t = f(s_{t-1}, x_t, y_t))} \qquad (35)
\end{aligned}$$

[3] The proof of the properties of T are entirely analogous to the proofs of Propositions 1.2.1 and 1.2.4 in [15, vol. II].

TABLE II
THE DYNAMIC PROGRAMMING FORMULATION THAT CORRESPONDS TO THE FEEDBACK CAPACITY PROBLEM OF A UNIFILAR CHANNEL

Dynamic Programming	Feedback capacity of unifilar channel
z_t, the state of DP	$\beta_t = p(s_t \vert y^t)$, channel state probabilities given the output up to time t.
w_t, the disturbance	y_t, the output of the channel
u_t, the action at time t	$p(x_t \vert s_{t-1})$, the matrix of conditional probabilities of the input given the state
$z_t = F(z_{t-1}, u_{t-1}, w_{t-1})$	eq. (35)
$p(z_0)$, initial state distribution	$p(s_0)$ initial channel state distribution
$p(w_t \vert z_{t-1}, u_t)$, the disturbance distribution	$p(y_t \vert \beta_{t-1}, u_t)$
$g(z_{t-1}, u_t)$, the reward at time t	$I(X_t, S_{t-1}; Y_t \vert \beta_{t-1}, u_t)$

history Φ_t; however, for the trapdoor channel, we will be able to restrict the dependency to only z_{t-1}, and show that it takes one of only four possible values with probability one. This is done in Section VI-D by utilizing Theorem 5.

The dynamic programming formulation that is presented here is an extension of the formulation presented in [11] by Yang, Kavčić, and Tatikonda. In [11], the formulation is for channels with the property that the state is deterministically identified by the previous inputs, and here we allow the state to be determined by the previous outputs and inputs.

VI. SOLUTION FOR THE TRAPDOOR CHANNEL

The trapdoor channel presented in Section II is a simple example of a unifilar FSC. In this section, we present an explicit solution to the associated dynamic program, which yields the feedback capacity of the trapdoor channel as well as an optimal encoder–decoder pair. The analysis begins with a computational study using numerical dynamic programming techniques. The results give rise to conjectures about the average reward, the differential value function, and an optimal policy. These conjectures are proved to be true through verifying that they satisfy Bellman's equation.

A. The Dynamic Program

In Section V-C, we formulated a class of dynamic programs associated with unifilar channels. From here on, we will focus on the particular instance from this class that represents the trapdoor channel.

Using the same notation as in Section V-C, the state z_{t-1} would be the vector of channel state probabilities $[p(s_{t-1} = 0 \vert y^{t-1}), p(s_{t-1} = 1 \vert y^{t-1})]$. However, to simplify notation, we will consider the state z_{t-1} to be the first component; that is, $z_{t-1} \triangleq p(s_{t-1} = 0 \vert y^{t-1})$. This comes with no loss of generality—the second component can be derived from the first since the pair sums to one. The action is a 2 × 2 stochastic matrix

$$u_t = \begin{bmatrix} p(x_t = 0 \vert s_{t-1} = 0) & p(x_t = 1 \vert s_{t-1} = 0) \\ p(x_t = 0 \vert s_{t-1} = 1) & p(x_t = 1 \vert s_{t-1} = 1) \end{bmatrix}. \qquad (39)$$

The disturbance w_t is the channel output y_t.

The state evolves according to $z_t = F(z_{t-1}, u_t, w_t)$, where we obtain the function F explicitly using relations from (3), (35), and Table I as shown in the equation at the bottom of the page.These expressions can be simplified by defining

$$\gamma_t \triangleq (1 - z_{t-1}) u_t(2,2) \qquad (40)$$

$$\delta_t \triangleq z_{t-1} u_t(1,1) \qquad (41)$$

so that

$$z_t = \begin{cases} \frac{2\delta_t}{1+\delta_t-\gamma_t}, & \text{if } w_t = 0 \\ 1 - \frac{2\gamma_t}{1-\delta_t+\gamma_t}, & \text{if } w_t = 1. \end{cases}$$

Note that, given z_{t-1}, the action u_t defines the pair (γ_t, δ_t) and vice versa. From here on, we will represent the action in terms of γ_t and δ_t. Because u_t is required to be a stochastic matrix, δ_t and γ_t are constrained by $0 \le \delta_t \le z_t$ and $0 \le \gamma_t \le 1 - z_t$.

Recall from (38) that the reward function is given by

$$g(z_{t-1}, u_t) = I(X_t, S_{t-1}; Y_t \vert \beta_{t-1}, u_t).$$

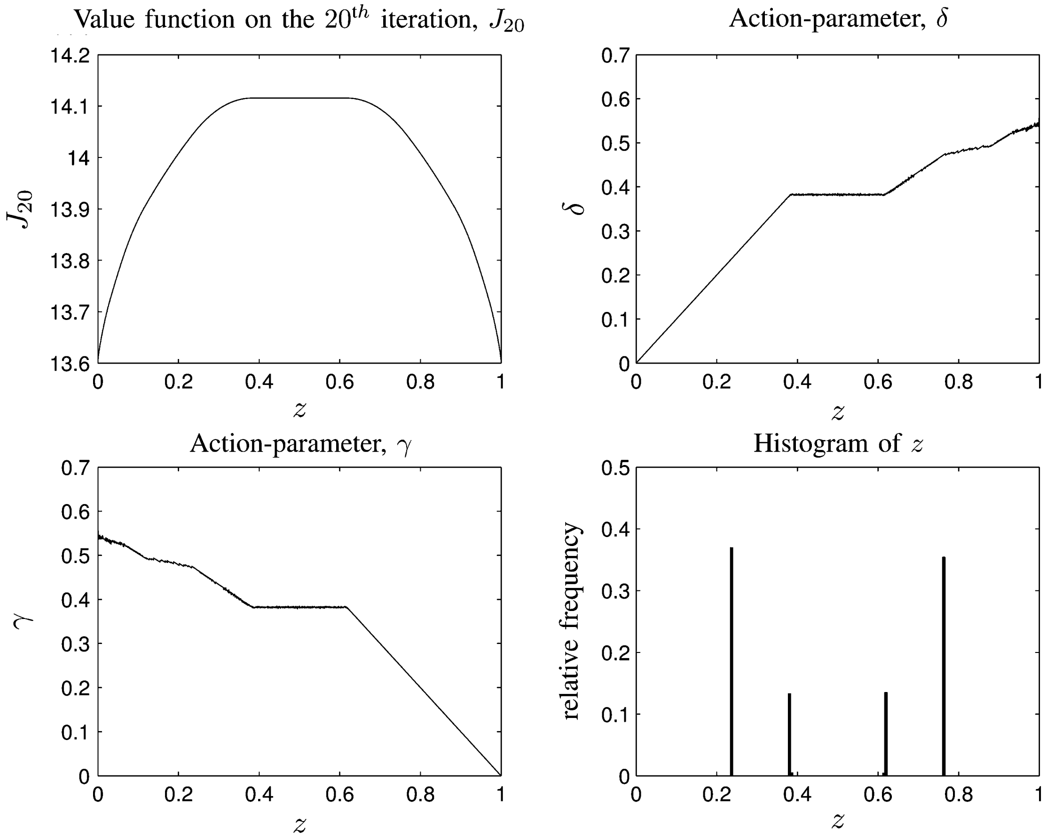
This reward can be computed from the conditional probabilities $p(x_t, s_{t-1}, y_t \vert \beta_{t-1}, u_t)$. Using the expressions for these conditional probabilities provided in Table III, we obtain

$$\begin{aligned} g(z_{t-1}, u_t) &= I(X_t, S_{t-1}; Y_t \vert \beta_{t-1}, u_t) \\ &= H(Y_t \vert u_t, \beta_{t-1}) - H(Y_t \vert X_t, S_{t-1}, \beta_{t-1}, u_t) \\ &= H\left(z_{t-1} u_t(1,1) + \frac{z_{t-1} u_t(1,2)}{2} \right. \\ &\qquad \left. + \frac{(1 - z_{t-1}) u_t(2,1)}{2} \right) \\ &\quad - z_{t-1} u_t(1,2) - (1 - z_{t-1}) u_t(2,1) \\ &= H\left(\frac{1}{2} + \frac{\delta_t - \gamma_t}{2} \right) + \delta_t + \gamma_t - 1 \end{aligned}$$

where, with some abuse of notation, we use H to denote the binary entropy function: $H(q) = -q \ln q - (1-q) \ln(1-q)$.

We now have a dynamic program—the objective is to maximize over all policies π the average reward ρ_π. The capacity of the trapdoor channel is the maximum of the average reward ρ^*.

$$z_t = \begin{cases} \dfrac{z_{t-1} u_t(1,1)}{z_{t-1} u_t(1,1) + 0.5 z_{t-1} u_t(1,2) + 0.5(1 - z_{t-1}) u_t(2,1)}, & \text{if } w_t = 0 \\ \dfrac{0.5(1 - z_{t-1}) u_t(2,1) + 0.5 z_{t-1} u_t(1,2)}{0.5(1 - z_{t-1}) u_t(2,1) + 0.5 z_{t-1} u_t(1,2) + (1 - z_{t-1}) u_t(2,2)}, & \text{if } w_t = 1. \end{cases}$$


Value function on the 20^{th} iteration, J_{20}

Action-parameter, δ

Action-parameter, γ

Histogram of z

Fig. 4. Results from 20 value iterations. On the top-left side, the value function J_{20} is plotted. On the top-right and bottom-left, the optimal action-parameters δ and γ with respect to the 20th iteration are plotted. On the bottom-right, the relative state frequencies of the associated Markov process of z with the policy that is optimal with respect to J_{20} is plotted.

TABLE III
THE CONDITIONAL DISTRIBUTION $p(x_t, s_{t-1}, y_t|\beta_{t-1}, u_t)$

x_t	s_{t-1}	$y_t = 0$	$y_t = 1$
0	0	$\beta_{t-1}u_t(1,1)$	0
0	1	$0.5(1-\beta_{t-1})u_t(2,1)$	$0.5(1-\beta_{t-1})u_t(2,1)$
1	0	$0.5\beta_{t-1}u_t(1,2)$	$0.5\beta_{t-1}u_t(1,2)$
1	1	0	$(1-\beta_{t-1})u_t(2,2)$

In the context of the trapdoor channel, the dynamic programming operator takes the form

$$(Th)(z) = \sup_{0\le\delta\le z, 0\le\gamma\le 1-z} \left(H\left(\frac{1}{2} + \frac{\delta-\gamma}{2}\right) + \delta + \gamma - 1 + \frac{1+\delta-\gamma}{2} h\left(\frac{2\delta}{1+\delta-\gamma}\right) + \frac{1-\delta+\gamma}{2} h\left(1 - \frac{2\gamma}{1-\delta+\gamma}\right)\right). \quad (42)$$

By Theorem 5, if we identify a scalar ρ and bounded function h that satisfy Bellman's equation, $\rho\mathbf{1} + Th = h$, then ρ is the optimal average reward. Further, if for each $z, T_\mu h = Th$, then the stationary policy μ is an optimal policy.

B. Computational Study

We carried out computations to develop an understanding of solutions to Bellman's equation. For this purpose, we used the *value iteration* algorithm, which in our context generates a sequence of iterates according to

$$J_{k+1} = TJ_k \quad (43)$$

initialized with $J_0 = 0$. For each k and z, $J_k(z)$ is the maximal expected reward over k time periods given that the system starts in state z. Since rewards are positive, for each $z, J_k(z)$ grows with k. For each k, we define a differential reward function $h_k(z) \triangleq J_k(z) - J_k(0)$. These functions capture differences among values $J_k(z)$ for different states x. Under certain conditions, such as those presented in [30], the sequence h_k converges uniformly to a function that solves Bellman's equation. We will neither discuss such conditions nor verify that they hold. Rather, we will use the algorithm heuristically in order to develop intuition and conjectures.

Value iteration as described above cannot be implemented on a computer because it requires storing and updating a function with infinite domain and optimizing over an infinite number of actions. To address this, we discretize the state and action spaces, approximating the state space using a uniform grid with 2000 points in the unit interval and restricting actions δ and γ to values in a uniform grid with 4000 points in the unit interval.

We executed 20 value iterations. Fig. 4 plots the function J_{20} and actions that maximize the right-hand side of (43) with $k = 20$. We also simulated the system, selecting actions δ_t and γ_t in each time period to maximize this expression. This led to an average reward of approximately 0.694. In the right-bottom side of Fig. 4, we plot the relative state frequencies of the associated Markov process. Note that the distribution concentrates around four points which are approximately $0.236, 0.382, 0.613$, and 0.764.

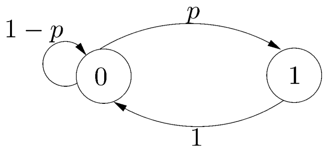


Fig. 5. The Markov chain of question 1.

C. Conjectures

The results obtained from value iteration were, amazingly, close to the answers of two questions posed in an information theory class at Stanford taught by Prof. Thomas Cover. Here is a simplified version of the questions given to the class.

1) *Entropy rate*. Find the maximum entropy rate of the two-state Markov chain (Fig. 5) with transition matrix

$$P = \begin{bmatrix} 1-p & p \\ 1 & 0 \end{bmatrix} \tag{44}$$

where $0 \leq p \leq 1$ is the free parameter to maximize over.

2) *Number of sequences*. To first order in the exponent, what is the number of binary sequences of length n with no two 1's in a row?

The entropy rate of the Markov chain of question 1 is given by $\frac{H(p)}{1+p}$, and when maximizing over $0 \leq p \leq 1$, we get that $p = \frac{3-\sqrt{5}}{2}$ and the entropy rate is 0.6942. It can be shown that the number of sequences of length $n-1$ that do not have two 1's in a row is the nth number in the Fibonacci sequence. This can be proved by induction in the following way. Let us denote (N_n^0, N_n^1) the number of sequences of length n with the condition of not having two 1's in a row that are ending with "0" and with "1," respectively. For the sequences that end with "0" we can either add a next bit "1" or "0," and for the sequences that end with :1," we can add only "0." Hence, $N_{n+1}^0 = N_n^0 + N_n^1$ and $N_{n+1}^1 = N_n^0$. By repeating this logic, we get that N_n^0 behaves as a Fibonacci sequence. To first order in the exponent, the Fibonacci number behaves as

$$\lim_{n\to\infty} \frac{1}{N} \log f_n = \log \frac{1+\sqrt{5}}{2} = 0.6942$$

where the number $\frac{1+\sqrt{5}}{2}$ is called the golden ratio. The golden ratio is also known to be a positive number that solves the equation $\frac{1}{\phi} = 1 - \phi$, and it appears in many math, science, and even artistic contexts [31]. As these problems illustrate, the number of typical sequences created by the Markov process given in question 1 is, to first order in the exponent, equal to the number of binary sequences that do not have two 1's in a row.

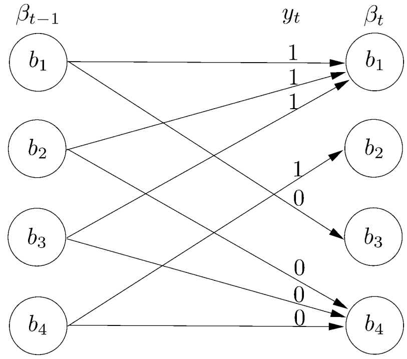


Fig. 6. The transition between β_{t-1} and β_t, under the policy $\tilde{\delta}, \tilde{\gamma}$.

Let us consider a policy for the dynamic program associated with a binary random process that is created by the Markov chain from question 1 (see Fig. 5) and inspired by the communication scheme introduced in Section VII. Let the state of the Markov process indicate if the input to the channel will be the same or different from the state of the channel. In other words, if at time t the binary Markov sequence is "0," then the input to the channel is equal to the state of the channel, i.e., $x_t = s_{t-1}$. Otherwise, the input to the channel complements the state of the channel, i.e., $x_t = s_{t-1} \oplus 1$. This scheme uniquely defines the distribution $p(x_t|s_{t-1}, y^{t-1})$

$$p(X_t = s_{t-1}|s_{t-1}, y_{t-1}) = \begin{cases} 1-p, & \text{if } s_{t-1} = y_{t-1} \\ 1, & \text{if } s_{t-1} \neq y_{t-1}. \end{cases} \tag{45}$$

This distribution is derived from the fact that for the trapdoor channel the state evolves according to (3) which can be written as

$$s_{t-1} \oplus y_{t-1} = x_{t-1} \oplus s_{t-2}. \tag{46}$$

Hence, if $s_{t-1} \neq y_{t-1}$ then necessarily also $x_{t-1} \neq s_{t-2}$. This means that the tuple (s_{t-1}, y_{t-1}) defines the state of the Markov chain at time $t-1$ and the tuple (x_t, s_{t-1}) defines the state of the Markov chain at time t. Having the distribution $p(x_t|s_{t-1}, y^{t-1})$, for the following four values of z $\{b_1 \triangleq \sqrt{5}-2, b_2 \triangleq \frac{3-\sqrt{5}}{2}, b_3 \triangleq \frac{\sqrt{5}-1}{2}, b_4 \triangleq 3-\sqrt{5}\}$, the corresponding actions $\tilde{\gamma}(z)$ and $\tilde{\delta}(z)$, which are defined in (40), (41), are

z	$\tilde{\gamma}(z)$	$\tilde{\delta}(z)$
b_1 or b_2	$\frac{\sqrt{5}-1}{2}(1-z)$	z
b_3 or b_4	$1-z$	$\frac{\sqrt{5}-1}{2}z$

It can be verified, by using (35), that the only values of z ever reached are

$$z \in \left\{ b_1 \triangleq \sqrt{5}-2, b_2 \triangleq \frac{3-\sqrt{5}}{2}, b_3 \triangleq \frac{\sqrt{5}-1}{2}, b_4 \triangleq 3-\sqrt{5} \right\} \tag{47}$$

and the transitions are a function of y_t, shown graphically in Fig. 6. Our goal is to prove that an extension of this policy is indeed optimal. Based on the answer to question 1, we conjectured that the entropy rate is the average reward, i.e.,

$$\tilde{\rho} = \frac{H\left(\frac{3-\sqrt{5}}{2}\right)}{1+\frac{3-\sqrt{5}}{2}} = \log \frac{\sqrt{5}+1}{2} \approx 0.6942. \tag{48}$$

It is interesting to notice that all the numbers appearing above can be written in terms of the golden ratio $\phi = \frac{\sqrt{5}+1}{2}$. In particular, $\tilde{\rho} = \log \phi, b_1 = 2\phi - 3, b_2 = 2 - \phi, b_3 = \phi - 1$, and $b_4 = 4 - 2\phi$.

By inspection of Fig. 4, we let $\tilde{\gamma}$ and $\tilde{\delta}$ be linear over the intervals $[b_1, b_2]$, $[b_2, b_3]$, and $[b_3, b_4]$, and we get the form presented in Table IV.

TABLE IV
CONJECTURED POLICY WHICH IN THE NEXT SECTION WILL BE PROVEN TO BE TRUE

z	$\tilde{\gamma}(z)$	$\tilde{\delta}(z)$
$b_1 \le z \le b_2$	$\frac{\sqrt{5}-1}{2}(1-z)$	z
$b_2 \le z \le b_3$	$\frac{3-\sqrt{5}}{2}$	$\frac{3-\sqrt{5}}{2}$
$b_3 \le z \le b_4$	$1-z$	$\frac{\sqrt{5}-1}{2}z$

We now propose differential values $\tilde{h}(z)$ for $z \in [b_1, b_4]$. If we assume that $\tilde{\delta}$ and $\tilde{\gamma}$ maximize the right-hand side of the Bellman equation (34) for $z \in [b_1, b_4]$ with $h = \tilde{h}$ and $\rho = \tilde{\rho}$, we obtain

$$\tilde{h}(z) = H\left(\frac{1}{2}\right) - (\sqrt{5}-2) - \tilde{\rho} + \tilde{h}(3-\sqrt{5}), \quad b_2 \le z \le b_3 \tag{49}$$

$$\begin{aligned}\tilde{h}(z) = H\left(\frac{\sqrt{5}+1}{4}z\right) - \frac{3-\sqrt{5}}{2}z - \tilde{\rho} \\ + \frac{\sqrt{5}+1}{4}z\tilde{h}(3-\sqrt{5}) + \left(1 - \frac{\sqrt{5}+1}{4}z\right) \\ \times \tilde{h}\left(\frac{1-z}{1-\frac{\sqrt{5}+1}{4}z}\right), \quad b_3 \le z \le b_4.\end{aligned} \tag{50}$$

The equation for the range $b_1 \le z \le b_2$ is implied by the symmetry relation: $\tilde{h}(z) = \tilde{h}(1-z)$.

If a scalar ρ and function h solve Bellman's equation, so do ρ and $h + c\mathbf{1}$ for any scalar c. Therefore, there is no loss of generality in setting $\tilde{h}(1/2) = 1$. From (49) we have that

$$\tilde{h}(z) = 1, \quad b_2 \le z \le b_3. \tag{51}$$

In addition, by symmetry considerations we can deduce that $\tilde{h}(\sqrt{5}-2) = \tilde{h}(3-\sqrt{5})$, and from (49) we obtain

$$\tilde{h}(\sqrt{5}-2) = \tilde{h}(3-\sqrt{5}) = \tilde{\rho} - 2 + \sqrt{5} \approx 0.9303. \tag{52}$$

The argument of the last term in (50), which we denote here as $l(z) \triangleq \frac{1-z}{1-\frac{\sqrt{5}+1}{4}z}$, is in $[b_3, b_4]$ for $z \in [b_3, b_4]$. Hence, we can apply (50) twice. Namely, we substitute the last term in (50) with the identity given in (50), and by using simple algebra, such as $l(l(z)) = z$, we obtain[4]

$$\tilde{h}(z) = H(z) + \tilde{\rho}z + c_1, \quad b_3 \le z \le b_4, \tag{53}$$

where $c_1 = \log(3-\sqrt{5})$. By symmetry, we obtain

$$\tilde{h}(z) = H(z) - \tilde{\rho}z + c_2, \quad b_1 \le z \le b_2 \tag{54}$$

where $c_2 = \log(\sqrt{5}-1)$.

The conjectured policy $(\tilde{\gamma}, \tilde{\delta})$, which is given in Table IV, and the conjectured differential value $\tilde{h}$, which is given in (51)–(54), are plotted in Fig. 7.

D. Verification

In this section, we verify that the conjectures made in the previous section are correct. Our verification process proceeds as follows. First, we establish that if a function $h : [0,1] \mapsto \Re$ is concave, so is Th. In other words, value iteration retains concavity. We then consider a version of value iteration involving an iteration $h_{k+1} = Th_k - \tilde{\rho}\mathbf{1}$. Since subtracting a constant does not affect concavity, this iteration also retains concavity. We prove that if a function h_0 is the pointwise maximum among concave functions that are equal to $\tilde{h}$ in the interval $[b_1, b_4]$, then each iterate h_k is also concave and equal to $\tilde{h}$ in this interval. Further, the sequence is pointwise nonincreasing. These properties of the sequence imply that it converges to a function h^* that again is concave and equal to $\tilde{h}$ in the interval $[b_1, b_4]$. This function h^* together with $\tilde{\rho}$ satisfies Bellman's equation. Given this, Theorem 5 verifies our conjectures.

We begin with a lemma that will be useful in showing that value iteration retains concavity.

Lemma 6: Let $\zeta : [0,1] \times [0,1] \mapsto \Re$ be concave on $[0,z] \times [0, 1-z]$ for all $z \in [0,1]$ and

$$\psi(z) = \sup_{\delta \in [0,z], \gamma \in [0,1-z]} \zeta(\delta, \gamma).$$

Then $\psi : [0,1] \mapsto \Re$ is concave.

The proof of Lemma 6 is given in the Appendix.

Lemma 7: The operator T, defined in (42), retains concavity and continuity. Namely

- if h is concave then Th is concave,
- if h is continuous then Th is continuous.

Proof (Concavity): It is well known that the binary entropy function H is concave, so the reward function

$$H\left(\frac{1}{2} + \frac{\delta - \gamma}{2}\right) + \delta + \gamma - 1$$

is concave in (δ, γ).

Next, we show that if $h(z)$ is concave, then $\frac{1+\delta-\gamma}{2}h(\frac{2\delta}{1+\delta-\gamma})$ is concave in (δ, γ). Let $\xi_1 = \frac{1+\delta_1-\gamma_1}{2}$ and $\xi_2 = \frac{1+\delta_2-\gamma_2}{2}$. We will show that, for any $\alpha \in (0,1)$

$$\begin{aligned}\alpha\xi_1 h\left(\frac{\delta_1}{\xi_1}\right) + (1-\alpha)\xi_2 h\left(\frac{\delta_2}{\xi_2}\right) \\ \ge (\alpha\xi_1 + (1-\alpha)\xi_2) h\left(\frac{\alpha\delta_1 + (1-\alpha)\delta_2}{\alpha\xi_1 + (1-\alpha)\xi_2}\right).\end{aligned} \tag{55}$$

Dividing both sides by $(\alpha\xi_1 + (1-\alpha)\xi_2)$, we get

$$\begin{aligned}\frac{\alpha\xi_1}{\alpha\xi_1 + (1-\alpha\xi_2)} h\left(\frac{\delta_1}{\xi_1}\right) + \frac{(1-\alpha)\xi_2}{\alpha\xi_1 + (1-\alpha\xi_2)} \\ h\left(\frac{\delta_2}{\xi_2}\right) \ge h\left(\frac{\alpha\delta_1 + (1-\alpha)\delta_2}{\alpha\xi_1 + (1-\alpha)\xi_2}\right).\end{aligned} \tag{56}$$

[4] It is also possible to verify that $\tilde{h}$ defined in (53) satisfies (50) by substituting $\tilde{h}$, collecting terms, and using the following algebraic facts: $H(\frac{\sqrt{5}+1}{4}z) + (1 - \frac{\sqrt{5}+1}{4}z)H(l(z)) = H(z) + zH(\frac{\sqrt{5}+1}{4})$ and $H(\frac{\sqrt{5}+1}{4}) = -\frac{3-\sqrt{5}}{2} + \frac{\sqrt{5}+1}{4}(\tilde{\rho} - 2 + \sqrt{5}) - c_1\frac{\sqrt{5}+1}{4}$.

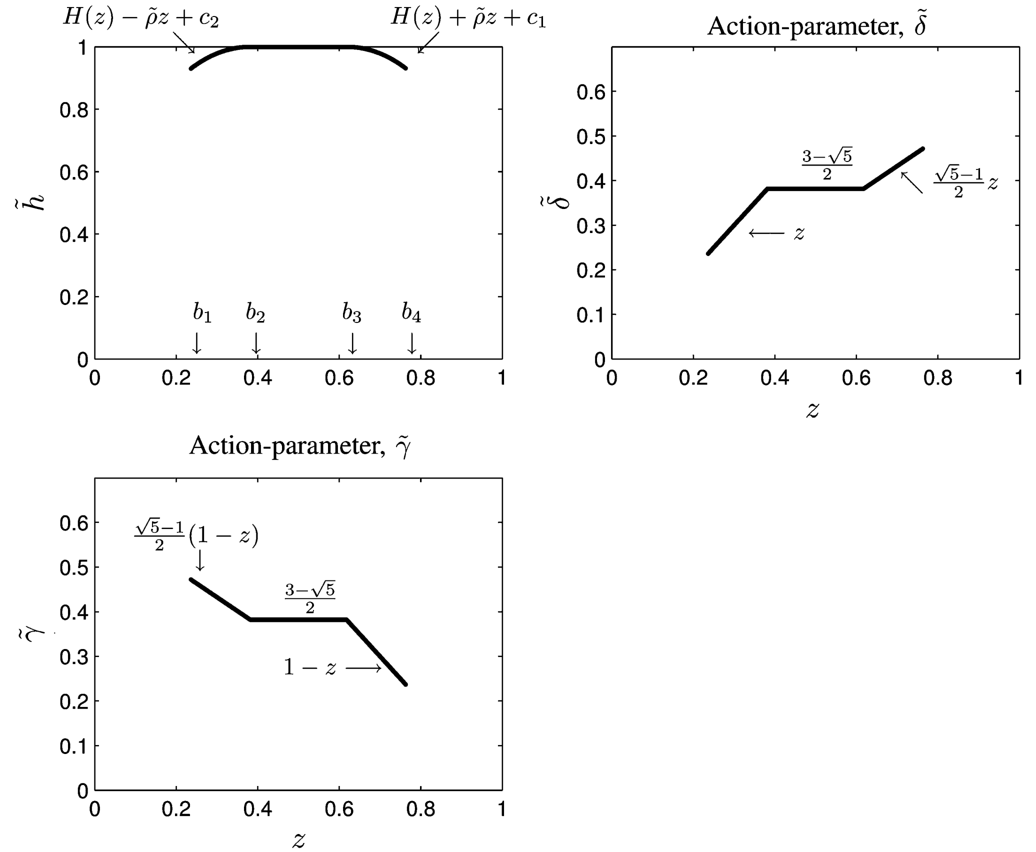


Fig. 7. A conjecture about the optimal solution based on the 20th value iteration of the dynamic program which is shown in Fig. 4 and on the communication scheme derived from the questions given by Prof. Cover. On the top-left, the conjectured differential value $\tilde{h}(z)$ is plotted for $z \in [b_1, b_4]$. On the top-right side and bottom-left, the conjectured policy $(\tilde{\delta}(z), \tilde{\gamma}(z))$ is plotted for $z \in [b_1, b_4]$.

Note that the last inequality is true because of the concavity of h. It follows that

$$f(\delta,\gamma) \triangleq H\left(\frac{1}{2}+\frac{\delta-\gamma}{2}\right) + \delta+\gamma-1+\frac{1+\delta-\gamma}{2}h\left(\frac{2\delta}{1+\delta-\gamma}\right) + \frac{1-\delta+\gamma}{2}h\left(1-\frac{2\gamma}{1-\delta+\gamma}\right) \tag{57}$$

is concave in (δ, γ). Since

$$(Th)(z) = \sup_{\delta\in[0,z],\gamma\in[0,1-z]} f(\delta,\gamma),$$

it is concave by Lemma 6. □

Proof (Continuity): Note that the binary entropy function H is continuous. Further, $h(\frac{2\delta}{1+\delta-\gamma})$ and $h(1-\frac{2\gamma}{1-\delta+\gamma})$ are continuous over the region $\{(\delta,\gamma)|\delta \geq 0, \gamma \geq 0, \delta+\gamma \leq 1\}$. It follows that $f(\delta,\gamma)$ is continuous over the region $\{(\delta,\gamma)|\delta \geq 0, \gamma \geq 0, \delta+\gamma \leq 1\}$. Hence

$$(Th)(z) = \sup_{\delta\in[0,z],\gamma\in[0,1-z]} f(\delta,\gamma)$$

is continuous over $[0,1]$. □

Let us construct *value iteration function* $h_k(z)$ as follows. Let $h_0(z)$ be the pointwise maximum among concave functions satisfying $h_0(z) = \tilde{h}(z)$ for $z \in [b_1, b_4]$, where $\tilde{h}(z)$ is defined in (51)–(54). Note that $h_0(z)$ is concave and that for $z \notin [b_1, b_4]$, $h_0(z)$ is a linear extrapolation from the boundary of $[b_1, b_4]$. Let

$$h_{k+1}(z) = (Th_k)(z) - \tilde{\rho} \tag{58}$$

and

$$h^*(z) \triangleq \limsup_{k\to\infty} h_k(z). \tag{59}$$

The following lemma shows several properties of the sequence of functions $h_k(z)$ including the uniform convergence. The uniform convergence is needed for verifying the conjecture, while the other properties are intermediate steps in proving the uniform convergence.

Lemma 8: The following properties hold:

8.1 for all $k \geq 0, h_k(z)$ is concave and continuous in z;

8.2 for all $k \geq 0, h_k(z)$ is symmetric around $\frac{1}{2}$, i.e.,

$$h_k(z) = h_k(1-z) \tag{60}$$

8.3 for all $k \geq 0$ and $z \in [b_1, b_4], h_k(z)$ is a fixed point, i.e.,

$$h_k(z) = \tilde{h}(z), \quad z \in [b_1, b_4] \tag{61}$$

and the stationary policy $\mu(z) = (\tilde{\delta}(z), \tilde{\gamma}(z))$, where $(\tilde{\delta}(z), \tilde{\gamma}(z))$ are defined in Table IV, satisfies $(T_\mu h_k)(z) = (Th_k)(z)$

8.4 $h_k(z)$ is uniformly bounded in k and z, i.e.,

$$\sup_k \sup_{z\in[0,1]} |h_k(z)| < \infty \tag{62}$$

8.5 $h_k(z)$ is monotonically nonincreasing in k, i.e.,

$$\lim_{k\to\infty} h_k(z) = h^*(z) \tag{63}$$

8.6 $h_k(z)$ converges uniformly to $h^*(z)$

Proof of 8.1: Since $h_0(z)$ is concave and continuous, and since the operator T retains continuity and concavity (see Lemma 7), it follows that $h_k(z)$ is concave and continuous for every k. □

Proof of 8.2: We prove this property by induction. First notice that $h_0(z)$ is symmetric and satisfies $h_0(z) = h_0(1-z)$. Now let us show that if it holds for h_k, then it holds for h_{k+1}.

Let $f_k(\delta,\gamma)$ denote the expression maximized to obtain $(Th_k)(z)$, i.e.,

$$\begin{aligned} f_k(\delta,\gamma) \triangleq\ & H\left(\frac{1}{2}+\frac{\delta-\gamma}{2}\right) \\ & + \delta+\gamma-1+\frac{1+\delta-\gamma}{2}h_k\left(\frac{2\delta}{1+\delta-\gamma}\right) \\ & + \frac{1-\delta+\gamma}{2}h_k\left(1-\frac{2\gamma}{1-\delta+\gamma}\right). \end{aligned} \tag{64}$$

Notice that $f_k(\delta,\gamma) = f_k(\gamma,\delta)$. Also observe that replacing the argument z with $1-z$ in Th_k yields the same result as exchanging between γ and δ. From those two observations it follows that $Th_k(z) = Th_k(1-z)$ and from the definition of h_{k+1} given in (58) it follows that $h_{k+1}(z) = h_{k+1}(1-z)$. □

Proof of 8.3: We prove this property by induction. Notice that h_0 satisfies $h_0(z) = \tilde{h}(z)$ for $z\in[b_1,b_4]$. We assume that h_k satisfies $h_k(z) = \tilde{h}(z)$, and then we will prove the property for h_{k+1}. Later in this proof, we will show that for $z\in[b_1,b_4]$

$$(T_\mu h_k)(z) = (Th_k)(z). \tag{65}$$

Since $(T_\mu h_k)(z)-\tilde{\rho} = \tilde{h}(z)$ for all $z\in[b_1,b_4]$ (see (49)–(54)), it follows that $h_{k+1}(z) = \tilde{h}(z)$ for all $z\in[b_1,b_4]$.

Now, let us show that (65) holds. Recall that in the proof of Lemma 7, (64), we showed that $f_k(\delta,\gamma)$ is concave in (δ,γ). The derivative with respect to δ is

$$\begin{aligned} \frac{\partial f_k(\delta,\gamma)}{\partial\delta} =\ & \frac{1}{2}\log\frac{1-\delta+\gamma}{1+\delta-\gamma}+1 \\ & + \frac{1}{2}h_k\left(\frac{2\delta}{1+\delta-\gamma}\right)-\frac{1}{2}h_k\left(\frac{2\gamma}{1-\delta+\gamma}\right) \\ & + \frac{1-\gamma}{1+\delta-\gamma}h'_k\left(\frac{2\delta}{1+\delta-\gamma}\right) \\ & + \frac{\gamma}{1-\delta+\gamma}h'_k\left(\frac{2\gamma}{1-\delta+\gamma}\right). \end{aligned} \tag{66}$$

The derivative with respect to γ is entirely analogous and can be obtained by mutually exchanging γ and δ.

For $z\in[b_2,b_3]$, the action $\tilde{\gamma}(z) = \tilde{\delta}(z) = \frac{3-\sqrt{5}}{2}$ is feasible and

$$\frac{2\tilde{\gamma}(z)}{1-\tilde{\delta}(z)+\tilde{\gamma}(z)} = \frac{2\tilde{\delta}(z)}{1+\tilde{\delta}(z)-\tilde{\gamma}(z)} = b_4.$$

Moreover, it is straightforward to check that the derivatives of f_k are zero at $(\tilde{\gamma}(z),\tilde{\delta}(z))$, and since f_k is concave, $(\tilde{\gamma}(z),\tilde{\delta}(z))$ attains the maximum. Hence, $(T_\mu h)(z) = (Th)(z)$ for $z\in[b_2,b_3]$.

For $z\in[b_3,b_4], \tilde{\gamma}(z) = 1-z$ and $\tilde{\delta}(z) = \frac{\sqrt{5}-1}{2}z$. Note that $\frac{2\tilde{\gamma}(z)}{1-\tilde{\delta}(z)+\tilde{\gamma}(z)}$ and $\frac{2\tilde{\delta}(z)}{1+\tilde{\delta}(z)-\tilde{\gamma}(z)}$ are in $[b_1,b_2]\cup[b_3,b_4]$. Using expressions for $\tilde{h}(z)$ given in (53) and (54), we can write derivatives of f at $(\tilde{\delta}(z),\tilde{\gamma}(z))$ as

$$\left.\frac{\partial f(\delta,\gamma)}{\partial\delta}\right|_{(\tilde{\delta}(z),\tilde{\gamma}(z))} = \log\frac{1-\tilde{\delta}(z)-\tilde{\gamma}(z)}{2\tilde{\delta}(z)}+1+\tilde{\rho} = 0 \tag{67}$$

$$\left.\frac{\partial f(\delta,\gamma)}{\partial\gamma}\right|_{(\tilde{\delta}(z),\tilde{\gamma}(z))} = \log\frac{1-\tilde{\delta}(z)-\tilde{\gamma}(z)}{2\tilde{\gamma}(z)}+1+\tilde{\rho} \geq 0. \tag{68}$$

Note that $\tilde{\gamma}(z)$ is the maximum of the feasible set $[0,1-z]$ and that the derivative of f_k with respect to γ at $(\tilde{\delta}(z),\tilde{\gamma}(z))$ is positive. In addition, $\tilde{\delta}(z)$ is in the interior of the feasible set $[0,z]$ and the derivative of f_k with respect to δ at $(\tilde{\delta}(z),\tilde{\gamma}(z))$ is zero. Since f_k is concave, any feasible change in $(\tilde{\gamma}(z),\tilde{\delta}(z))$ will decrease the value of the function. Hence, $(T_\mu h_k)(z) = (Th_k)(z)$ for $z\in[b_3,b_4]$. The situation for $z\in[b_1,b_2]$ is completely analogous. □

Proof of 8.4: From Propositions 8.1–8.3, it follows that the maximum over z of $h_k(z)$ is attained at $z=1/2$ and $h_k(1/2) = 1$ for all k. Furthermore, because of concavity and symmetry the minimumm of $h_k(z)$ is attained at $z=0$ and $z=1$. Hence, it is enough to show that $h_k(0)$ is uniformly bounded from below for all k.

For $z=0$, let us consider the action $\gamma(0) = \frac{1-b_2}{1+b_2}$ and $\delta(0) = 0$ and for $b_1\leq z\leq b_4$ the action $\tilde{\gamma}(z),\tilde{\delta}(z)$. Now let us prove that under this policy $h'_k(0)$, which is less than or equal to the optimal value, is uniformly bounded.

Under this policy, $h'_{k+1}(0) = (Th'_k)(0)-\tilde{\rho}$ becomes

$$h'_k(0) = c+\alpha h'_{k-1}(0)+(1-\alpha)1-\tilde{\rho} \tag{69}$$

where c and α are constant: $c = H(\frac{b_2}{1+b_2})+\frac{1-b_2}{1+b_2}-1, \alpha = \frac{b_2}{1+b_2}$.

Iterating (69) $k-1$ times, we get

$$h'_k(0) = (c+1-\alpha-\tilde{\rho})\sum_{i=0}^{k-1}\alpha^i+\alpha^k h'_0(0). \tag{70}$$

Since $\alpha<1, h'_k(0)$ is uniformly bounded for all k. Finally, since $h_k(0)\geq h'_k(0)$, then $h_k(0)$ is also bounded from below. □

Proof of 8.5: By Proposition 8.1, h_k is concave for each k and by Proposition 8.3, $h_k(z) = \tilde{h}(z)$ for $z\in[b_1,b_4]$. Since h_0 is the pointwise maximum of functions satisfying this condition, we must have $h_0\geq h_1$. It is easy to see that T is a monotonic operator. As such, $h_k\geq h_{k+1}$ for all k. Proposition 8.4 establishes that the sequence is bounded below, and therefore it converges pointwise. □

Proof of 8.6: By Proposition 8.1, each h_k is concave and continuous. Further, by Proposition 8.5, the sequence has a pointwise limit h^* which is concave. Concavity of h^* implies continuity [32, Theorem 10.1] over $(0, 1)$. Let $h^\dagger$ be the continuous extension of h^* from $(0, 1)$ to $[0, 1]$. Since h^* is concave, $h^\dagger \geq h^*$.

By Proposition 8.5, $h_k \geq h^*$. It follows from continuity of h_k that $h_k \geq h^\dagger$. Hence, $h^*(z) = \lim_k h_k(z) \geq h^\dagger(z)$ for $z \in [0, 1]$. Recalling that $h^* \leq h^\dagger$, we have $h^* = h^\dagger$.

Since the iterates h_k are continuous and monotonically nonincreasing and their pointwise limit h^* is continuous, h_k converges uniformly by Dini's theorem [33]. □

The following theorem verifies our conjectures.

Theorem 9: The function h^* and scalar $\tilde{\rho}$ satisfy $\tilde{\rho}\mathbf{1} + h^* = Th^*$. Further, $\tilde{\rho}$ is the optimal average reward and there is an optimal policy that takes actions $\delta_t = \tilde{\delta}(z_{t-1})$ and $\gamma_t = \tilde{\gamma}(z_{t-1})$ whenever $z_{t-1} \in [b_1, b_4]$.

Proof: Since the sequence $h_{k+1} = Th_k - \tilde{\rho}\mathbf{1}$ converges uniformly and T is sup-norm continuous, $h^* = Th^* - \tilde{\rho}\mathbf{1}$. It follows from Theorem 5 that $\tilde{\rho}$ is the optimal average reward. Together with Proposition 8.3, this implies existence of an optimal policy that takes actions $\delta_t = \tilde{\delta}(z_{t-1})$ and $\gamma_t = \tilde{\gamma}(z_{t-1})$ whenever $z_{t-1} \in [b_1, b_4]$. □

VII. A Capacity-Achieving Scheme

In this section, we describe a simple encoder and decoder pair that provides error-free communication through the trapdoor channel with feedback and known initial state. We then show that the rates achievable with this encoding scheme are arbitrarily close to capacity.

It will be helpful to discuss the input and output of the channel in different terms. Recall that the state of the channel is known to the transmitter because it is a deterministic function of the previous state, input, and output, and the initial state is known. Let the input action $\tilde{x}$ be one of the following:

$$\tilde{x} = \begin{cases} 0, & \text{input ball is same as state,} \\ 1, & \text{input ball is opposite of state.} \end{cases}$$

Also, let the output be recorded differentially as

$$\tilde{y} = \begin{cases} 0, & \text{received ball is same as previous,} \\ 1, & \text{received ball is opposite of previous,} \end{cases}$$

where $\tilde{y}_1$ is undefined and irrelevant for our scheme.

A. Encode/Decode Scheme

Encoding: Each message is mapped to a unique binary sequence of N actions $\tilde{x}^N$ that ends with 0 and has no occurrences of two 1's in a row. The input to the channel is derived from the action and the state as $x_k = \tilde{x}_k \oplus s_{k-1}$.

Decoding: The channel outputs are recorded differentially as $\tilde{y}_k = y_k \oplus y_{k-1}$ for $k = 2, \ldots, N$. Decoding of the action sequence is accomplished in reverse order, beginning with $\tilde{x}_N = 0$ by construction.

TABLE V
DECODING EXAMPLE

Variable	Value	Reason
y_n	1011010001	Channel output
$\tilde{y}_n$	*110111001	Differential output
$\tilde{x}_n$	0	Given
	10	Case 2
	010	Case 1
	0010	Case 2
	10010	Case 2
	010010	Case 1
	1010010	Case 2
	01010010	Case 1
	101010010	Case 2
	0101010010	Case 1

Lemma 10: If $\tilde{x}_{k+1}$ is known to the decoder, $\tilde{x}_k$ can be correctly decoded. Furthermore, the decoding rule is

$$\begin{aligned} &\text{Case 1}: \quad \text{If } \tilde{x}_{k+1} = 1, \text{ then } \tilde{x}_k = 0. \\ &\text{Case 2}: \quad \text{If } \tilde{x}_{k+1} = 0, \text{ then } \tilde{x}_k = \tilde{y}_{k+1}. \end{aligned}$$

Proof of Case 1: By construction there are never two 1's in a row.

Proof of Case 2: Recall that the trapdoor channel has the property that for all k,

$$x_{k+1} \oplus s_k = y_{k+1} \oplus s_{k+1}. \tag{71}$$

Hence, if $\tilde{x}_{k+1} = 0$, then

$$x_{k+1} = s_k = y_{k+1} = s_{k+1}. \tag{72}$$

Finally, this implies that

$$\begin{aligned} \tilde{x}_k &\triangleq x_k \oplus s_{k-1} \\ &\overset{(a)}{=} y_k \oplus s_k \\ &\overset{(b)}{=} y_k \oplus y_{k+1} \\ &\triangleq \tilde{y}_{k+1} \end{aligned} \tag{73}$$

where (a) is due to (71) and (b) to (72). □

Decoding Example: Table V shows an example of decoding a sequence of actions for $N = 10$.

B. Rate

Under this encoding scheme, the number of admissible unique action sequences is the number of binary sequences of length $N - 1$ without any repeating 1's. This is known to be exponentially equivalent to ϕ^{N-1}, where ϕ is the golden ratio (see question 2 in Section VI-C). Since $\lim_{N\to\infty} \frac{N-1}{N} \log\phi = \log\phi$, rates arbitrary close to $\log\phi$ are achievable.

C. Remarks

Early Decoding: Decoding can often begin before the entire block is received. From the decoding rule, it is easy to see that we can decode $\tilde{x}_k$ without knowledge of $\tilde{x}_{k+1}$ for any k such that $\tilde{y}_{k+1} = 0$. Decoding can begin from any such point and work backward.

Preparing the Channel: This communication scheme can still be implemented even if the initial state of the channel is not known as long as some channel uses are expended to prepare the channel for communication. The repeating sequence $010101\cdots$ can be used to flush the channel until the state becomes evident. As soon as the output of the channel is different from the input, both the transmitter (through feedback) and the receiver know that the state is the previous input. At that point, zero-error communication can begin as described above.

This flushing method requires a random and unbounded number of channel uses. However, it only needs to be performed once, after which multiple blocks of communication can be accomplished. The expected number of required channel uses is 3.5, which is derived by conditioning on the initial state and noticing that the number of pairs of channel uses needed is geometrically distributed. For a detailed finite block-length zero-error communication scheme over the trapdoor channel see [37].

Permuting Relay Channel Similarity: The permuting relay channel described in [4] has the same capacity as the trapdoor channel with feedback. A connection can be made using the communication scheme described in this section.

The permuting relay channel supposes that the transmitter chooses an input distribution to the channel that is independent of the message to be sent. The transmitter lives inside the trapdoor channel and chooses which of the two balls will be released to the receiver in order to send the message. Without proof here, let us assume that the deterministic input $010101\cdots$ is optimal. Now we count how many distinguishable outputs are possible.

It is helpful to view this as a permutation channel as described in Section II, but now the permuting is not done randomly but deliberately. Notice that for this input sequence, after each time that a pair of different numbers is permuted, the next pair of numbers will be the same, and the associated action will have no consequence. Therefore, the number of distinguishable permutations can be easily shown to be related to the number of unique binary sequences without two 1's in a row.

Three Channels Have the Same Feedback Capacity: The encoding/decoding scheme in this section allows zero-error communication. Therefore, this scheme could also be used to communicate with feedback through the permuting jammer channel from [4], which assumes that the trapdoor channel behavior is not random but is the worst possible to make communication difficult.

In the permuting relay channel [4], all information (input and output) is available to the transmitter, so feedback is irrelevant. Thus, we find that the feedback capacity is the same for the trapdoor, permuting jammer, and permuting relay channels.

Constrained Coding: The capacity-achieving scheme requires uniquely mapping a message to a sequence with the constraint of having no two 1's in a row. A practical way of accomplishing this is by a technique called *enumeration* [34]. The technique translates the message into codewords and vice versa by invoking an algorithmic procedure rather then using a lookup table. Vast literature on coding a source word into a constrained sequence can be found in [35] and [36].

VIII. Conclusion and Further Work

This paper gives an information-theoretic formulation for the feedback capacity of a connected unifilar finite-state channel and it shows that the feedback capacity expression can be formulated as an average-reward dynamic program. For the trapdoor channel, we were able to solve explicitly the dynamic programming problem and to show that the capacity of the channel is the log of the golden ratio. Furthermore, we were able to find a simple encoding/decoding scheme that achieves this capacity.

There are several directions in which this work can be extended.

- *Generalization*: Extend the trapdoor channel definition. It is possible to add parameters to the channel and make it more general. For instance, there could be a parameter that determines which ball from the two has the higher probability of being the output of the channel. Other parameters might include the number of balls that can be in the channel at the same time or the number of different types of balls that are used. These tie in nicely with viewing the trapdoor channel as a chemical channel.
- *Unifilar FSC Problems*: Find connected unifilar FSCs that can be solved, similar to the way we solved the trapdoor channel.
- *Dynamic Programming*: Classify a family of average-reward dynamic programs that have analytic solutions.

Appendix

Proof of Lemma 6: For any $z_1, z_2 \in [0,1]$ and $\theta \in (0,1)$

$$
\begin{aligned}
&\psi(\theta z_1 + (1-\theta)z_2)\\
&\quad= \sup_{\delta\in[0,\theta z_1+(1-\theta)z_2]} \sup_{\gamma\in[0,1-(\theta z_1+(1-\theta)z_2)]} \zeta(\delta,\gamma)\\
&\quad= \sup_{\delta_1\in[0,\theta z_1]} \sup_{\delta_2\in[0,(1-\theta)z_2]} \sup_{\gamma_1\in[0,\theta(1-z_1)]}\\
&\qquad\times \sup_{\gamma_2\in[0,(1-\theta)(1-z_2)]} \zeta(\delta_1+\delta_2,\gamma_1+\gamma_2)\\
&\overset{(a)}{=} \sup_{\delta_1'\in[0,z_1]} \sup_{\delta_2'\in[0,z_2]} \sup_{\gamma_1'\in[0,1-z_1]} \sup_{\gamma_2'\in[0,1-z_2]}\\
&\qquad\times \zeta(\theta\delta_1' + (1-\theta)\delta_2', \theta\gamma_1' + (1-\theta)\gamma_2')\\
&\overset{(b)}{\geq} \sup_{\delta_1'\in[0,z_1]} \sup_{\delta_2'\in[0,z_2]} \sup_{\gamma_1'\in[0,1-z_1]} \sup_{\gamma_2'\in[0,1-z_2]}\\
&\qquad\times \theta\zeta(\delta_1',\gamma_1') + (1-\theta)\zeta(\delta_2',\gamma_2')\\
&\quad= \sup_{\delta_1'\in[0,z_1]} \sup_{\gamma_1'\in[0,1-z_1]} \theta\zeta(\delta_1',\gamma_1')\\
&\qquad+ \sup_{\delta_2'\in[0,z_2]} \sup_{\gamma_2'\in[0,1-z_2]} (1-\theta)\zeta(\delta_2',\gamma_2')\\
&\quad= \theta\psi(z_1) + (1-\theta)\psi(z_2). \qquad (74)
\end{aligned}
$$

Step (a) is a change of variable $(\theta\delta_1' = \delta_1, (1-\theta)\delta_2' = \delta_2, \theta\gamma_1' = \gamma_1, (1-\theta)\gamma_2' = \gamma_2)$. Step (b) is due to concavity of ζ. □

Acknowledgment

The authors would like to thank Prof. Tom Cover, who introduced the trapdoor channel to H. Permuter and P. Cuff and asked them the two questions that appear in Section VI-C; this eventually led to the solution of the dynamic programming and to the simple scheme that achieves the feedback capacity. The final version of the manuscript benefited from comments of an anonymous referee.

REFERENCES

[1] D. Blackwell, "Information theory," *Modern Mathematics for the Engineer: Second Series*, pp. 183–193, 1961.

[2] R. Ash, *Information Theory*. New York: Wiley, 1965.

[3] T. Berger, "The generalized Shannon-Blackwell billiard ball channel," in *Lecture 2 of CSL Distinguished Visiting Professorship*. Urbana, IL: Univ. Illinois, Apr. 22, 2002, "Information Theory Invades Biology".

[4] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 3, pp. 310–314, May 1987.

[5] R. Ahlswede, N. Cai, and Z. Zhang, "Zero-error capacity for models with memory and the enlightened dictator channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1250–1252, May 1998.

[6] K. Kobayashi and H. Morita, "An input/output recursion for the trapdoor channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT2002)*, Lausanne, Switzerland, Jun./Jul. 2002, p. 423.

[7] K. Kobayashi, "Combinatorial structure and capacity of the permuting relay channel," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 813–826, Nov. 1987.

[8] P. Piret, "Two results on the permuting mailbox channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 4, pp. 888–892, Jul. 1989.

[9] W. K. Chan, "Coding strategies for the permuting jammer channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT2000)*, Sorrento, Italy, Jun. 1993, p. 211.

[10] S. C. Tatikonda, "Control Under Communication Constraints," Ph.D. disertation, MIT, Cambridge, MA, 2000.

[11] S. Yang, A. Kavčić, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.

[12] S. Yang, A. Kavčić, and S. Tatikonda, "On the feedback capacity of power constrained Gaussian channels with memory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 929–954, Mar. 2007.

[13] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780–789, Mar 2005.

[14] S. C. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory* [Online]. Available: arxiv.org/cs.IT/0609139, submitted for publication

[15] D. P. Bertsekas, *Dynamic Programming and Optimal Control: Vols 1 and 2*, 3rd ed. Belmont, MA: AthenaScientific, 2005.

[16] Y.-H. Kim, "Feedback capacity of the first-order moving average gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3063–3079, Jul. 2006.

[17] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels with feedback," *IEEE Trans. Inf. Theory* [Online]. Available: arxiv.org/cs.IT/0701041, submitted for publication

[18] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. Marcus, "Discrete time controlled Markov processes with average cost criterion—A survey," *SIAM J. Contr. Optimiz.*, vol. 31, no. 2, pp. 282–344, 1993.

[19] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 4, pp. 453–460, Jul. 1985.

[20] T. W. Benjamin, "Coding for a Noisy Channel With Permutation Errors," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1975.

[21] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Capacity of finite-state channels with time-invariant deterministic feedback," in *Proc. IEEE Int. Symp. Information Theory (ISIT2006)*, Seattle, WA, Jul. 2006, pp. 64–68.

[22] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory* [Online]. Available: arxiv.org/pdf/cs.IT/0608070, submitted for publication

[23] S. S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proceedings 2004 International Symposium on Information Theory*, 2004, p. 212.

[24] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion function for general sources," in *Proc. IEEE Information Theory Workshop (ITW)*, San Antonio, TX, Oct. 2004.

[25] R. Zamir, Y. Kochman, and U. Erez, "Achieving the gaussian rate-distortion function by prediction," *IEEE Trans. Inf. Theory*, submitted for publication.

[26] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[27] G. Kramer, "Directed Information for Channels With Feedback," Ph.D. dissertation, Swiss Federal Institute of Technology (ETHZ), Zurich, Switzerland, 1998.

[28] A. Rao, A. O. Hero, D. J. States, and J. D. Engel, "Inference of biologically relevant gene influence networks using the directed information criterion," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing* , Toulouse, France, May 2006.

[29] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Information Theory and Its Applications (ISITA-90)*, Honolulu, HI, Nov. 1990, pp. 303–305.

[30] Q. Zhu and X. Guo, "Value iteration for average cost Markov decision processes in Borel spaces," *AMRX Applied Mathematics Research eXpress*, vol. 2, pp. 61–76, 2005.

[31] L. Mario, *The Golden Ratio : The Story of Phi, the World's Most Astonishing Number*. New York: Broadway Books, 2002.

[32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[33] J. E. Marsden and M. J. Hoffman, *Elementary Classical Analysis*, 2nd ed. New York: W. H. Freeman and Company, 1993.

[34] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. IT–19, no. 1, pp. 73–77, Jan. 1973.

[35] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. S. Pless and W. C. Hu, Eds. Amsterdam, The Netherlands: Elsevier, 1998, pp. 1635–1764.

[36] K. A. S. Immink, *Codes for Mass Data Storage Systems*. Rotterdam, The Netherlands: Shannon Foundation, 2004.

[37] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity and zero-error capacity of the chemical channel with feedback," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1866–1870.