Capacity of a POST Channel with and without Feedback

Himanshu Asnani
Stanford University
asnani@stanford.edu

Haim H. Permuter
Ben Gurion University
haimp@bgu.ac.il

Tsachy Weissman
Stanford University
tsachy@stanford.edu

***Abstract*—We consider finite state channels where the state of the channel is its previous output. We refer to such channels as POST (Previous Output is the STate) channels. Our focus is on a simple binary POST channel, with binary inputs and outputs where the state determines if the channel behaves as a Z or an S channel (of equal capacities). We show that the non feedback capacity equals the feedback capacity, despite the memory in the channel. The proof of this surprising result is based on showing that the induced output distribution, when maximizing the directed information in the presence of feedback, can also be achieved by an input distribution that is ignorant of the feedback. Indeed, we show that this is a necessary and sufficient condition for the feedback capacity to equal the non feedback capacity for any finite state channel.**

***Index Terms*—Causal conditioning, Convex optimization, Directed information, Feedback capacity, Finite state channel, KKT conditions.**

I. Introduction

The capacity of a memoryless channel is very well understood. There are many simple memoryless channels for which we know the capacity analytically. Such channels include the binary symmetric channel, the erasure channel, the additive Gaussian channel and the Z Channel. Furthermore, using convex optimization tools, such as the Blahut-Arimoto algorithm [1], [2], we can efficiently compute the capacity of any memoryless channel with a finite alphabet. However, in the case of channels with memory, the exact capacities of only a few channels are known, such as additive Gaussian channels (water filling solution) [3], [4] and discrete additive channels with memory [5]. In cases where feedback is allowed there are only a few more cases where the exact capacity is known such as, additive noise channel where the noise is first-order autoregressive moving-average Gaussian [6], the trapdoor channel [7], and the Ising Channel [8].

In this paper we introduce and consider a new family of channels we refer to as "POST channels". These are simple Finite State Channels (FSCs) where the state of the channel is the previous output. In particular, we focus on a simple POST channel that has binary inputs $\{X_i\}_{i\geq 1}$ and binary outputs $\{Y_i\}_{i\geq 1}$ with the following behavior

$$\text{if } X_i = Y_{i-1}, \text{ then } Y_i = X_i, \text{ else } Y_i \sim \text{Bernnouli}\left(\frac{1}{2}\right). \tag{1}$$

We refer to this as a simple point-to-point POST channel and note that it is equivalently described as behaving like a Z channel when the previous channel output is 0 and an S channel when it is 1. Its behavior is depicted in Fig. 1.

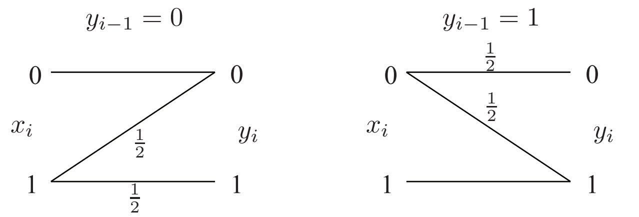

Fig. 1. Simple POST Channel behavior. If $y_{i-1} = 0$ then the channel behaves as a Z channel and If $y_{i-1} = 1$ then it behaves as an S channel.

The simple POST channel is similar to the Ising channel introduced by Berger and Bonomi [9], but rather than the previous input being the state of the channel, here the state of the channel is the previous output. This channel arose in the investigation of controlled feedback in the setting of "to feed or not to feed back" [10]. The POST channel can also be useful in modeling memory affected by past channel outputs, as is the case in flash memory and other storage devices.

In order to gain intuition for investigating the influence of feedback on the simple POST channels, we considered a channel with binary i.i.d. states S_i, distributed Bernoulli($\frac{1}{2}$), where the channel behaves similarly to the simple POST channel. When $S_{i-1} = 0$, then the channel behaves as a Z channel and when $S_{i-1} = 1$, it behaves as an S channel, as shown in Fig. 2. Similar to the POST channel, we assume that the state of the channel is known to the decoder; hence the output of the channel is (Y_i, S_i) (or, equivalently from a capacity standpoint, (Y_i, S_{i-1})).

The non feedback capacity of this channel is simply $C = \max_{P(x)} I(X;Y,S)$ and, because of symmetry, the input that achieves the maximum is Bernoulli($\frac{1}{2}$) , resulting in a capacity of $H_b(\frac{1}{4}) - \frac{1}{2} = 0.3111$, where $H_b(p)$ is the binary entropy function. However, if there is perfect feedback of (Y_i, S_i) to the encoder then the state of the channel is known to the encoder and the capacity is simply the capacity of the Z (or S) channel, which is $H_b(\frac{1}{5}) - \frac{2}{5} = -\log_2 0.8 = 0.3219$. Evidently, feedback increases the capacity of this channel.

The similarity between the channels may seem to suggest that feedback increases the capacity of the POST channel as well. Indeed, our initial interest in this channel was due to this suspicion, in our quest for a channel that would be

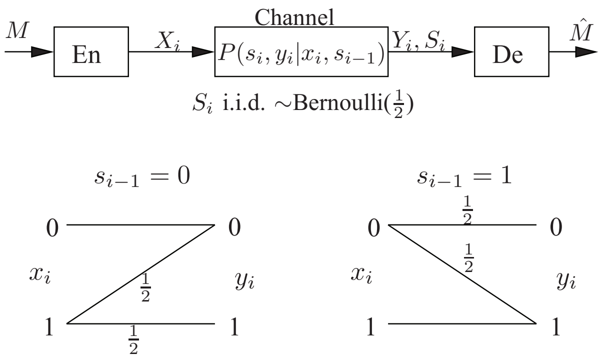

Fig. 2. A channel similar to the simple POST channel, except that the channel state, $\{S_i\}_{i\geq 1}$ is i.i.d. Bernoulli($\frac{1}{2}$) independent of the input.

amenable to analysis under the "to feed or not to feed" framework of constrained feedback in [10], while exhibiting non-trivial dependence of its capacity on the extent to which the feedback is constrained. However, numerical results based on the computational algorithm devised in [10] suggested that feedback does not increase the capacity of the simple POST channel. This paper stemmed from our attempts to make sense of these surprising observations.

In order to prove that feedback does not increase the capacity of the simple POST channel, we look at two convex optimization problems: maximizing the directed information over regular input distributions (non feedback case), i.e., $P(x^n)$ and, secondly, over causal conditioning that is influenced by the feedback i.e., $P(x^n||y^{n-1})$. We show that a necessary and sufficient condition for the solutions of the two optimization problems to achieve the same value is that the induced output distributions $P(y^n)$ by the respective optimal values $P^*(x^n)$ and $P^*(x^n||y^{n-1})$ be the same. This necessary and sufficient condition, we establish in the generality of any finite state channel, follows from the KKT (KarushKuhnTucker) conditions [11, Ch. 5] for convex optimization problems.

The remainder of the paper is organized as follows. In Section II, we briefly present the definitions of the directed information measure and causal conditioning pmfs that we use throughout the paper. In Section III, we show that the optimization problem of maximizing the directed information over causal conditioning pmfs is a convex optimization problem and, using the KKT conditions, we show that if the output distribution induced by the conditional pmfs that achieves the maximum for the feedback case, can also be induced by a regular input distribution that does not use feedback, then feedback does not increase the capacity. In Section IV we compute the feedback capacity of the POST channel. Then we apply the result of Section III to show that it equals the non feedback capacity by establishing the existence of an input distribution without feedback that induces the same output distribution as the capacity achieving one from the feedback case. In Section V, we present conclusions and directions for further research on the family of POST channels. Because of space limitation we omit most of the proofs, however all lemmas and theorems are fully stated and short explanation of the proofs are given.

II. Directed information, causal conditioning and notations

Throughout this paper, we denote random variables by capital letter such as X. The probability $\Pr\{X = x\}$ is denoted by $p(x)$. In this paper we denote the whole vector of probabilities by capital P, i.e., $P(x)$ is the vector probability of random variable X. Some denote the vector probability by P_X but we would like to avoid multiple subscript for the sake of simplicity of notation.

We use the *causal conditioning* notation $(\cdot||\cdot)$ developed by Kramer [12]. We denote by $p(x^n||y^{n-d})$, the probability mass function of $X^n = (X_1, \dots, X_n)$, *causally conditioned* on Y^{n-d} for some integer $d \geq 0$, which is defined as

$$p(x^n||y^{n-d}) := \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-d}). \tag{2}$$

By convention, if $i < d$, then y^{i-d} is set to null, i.e., if $i < d$ then $p(x_i|x^{i-1}, y^{i-d})$ is just $p(x_i|x^{i-1})$. In particular, we use extensively the cases $d = 0, 1$:

$$p(x^n||y^n) := \prod_{i=1}^{n} p(x_i|x^{i-1}, y^i), \tag{3}$$

$$p(x^n||y^{n-1}) := \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-1}). \tag{4}$$

The directed information was defined by Massey [13], inspired by Marko's work [14] on bidirectional communication, as

$$I(X^n \to Y^n) := \sum_{i=1}^{n} I(X^i; Y_i|Y^{i-1}). \tag{5}$$

The directed information can also be rewritten as

$$I(X^n \to Y^n) = \sum_{x^n, y^n} p(x^n||y^{n-1})p(y^n||x^n) \log \frac{p(y^n||x^n)}{\sum_{x^n} p(x^n||y^{n-1})p(y^n||x^n)} \tag{6}$$

This is due to the definition of causal conditioning and the chain rule

$$p(x^n, y^n) = p(x^n||y^{n-1})p(y^n||x^n). \tag{7}$$

It was shown that directed information indeed characterizes the capacity of point-to-point channels with feedback [15]–[18]. In particular, the POST channel has the property that the state is a function of the output. Based on this property, it was shown [7], [15] that the feedback capacity is given by

$$C_{fb} = \lim_{n\to\infty} \frac{1}{n} \max_{P(x^n||y^{n-1})} I(X^n \to Y^n). \tag{8}$$

Furthermore, without feedback the capacity is given by

$$C = \lim_{n\to\infty} \frac{1}{n} \max_{P(x^n)} I(X^n \to Y^n), \tag{9}$$

since the channel is indecomposable [19]. In the case where there is no feedback, namely, the Markov form $X_i - X^{i-1} - Y^{i-1}$ holds, $I(X^n \to Y^n) = I(X^n; Y^n)$, as shown in [13].

III. Maximization of the directed information as a convex optimization problem

In order to show that feedback does not increase the capacity of the simple POST channel, we consider the two optimization problems:

$$\max_{P(x^n||y^{n-1})} I(X^n \to Y^n) \tag{10}$$

and

$$\max_{P(x^n)} I(X^n \to Y^n). \tag{11}$$

In this section, we show that both the problems are convex optimization problems and using, the KKT condition, we state a necessary and sufficient condition in order for the two optimization problems to obtain the same value.

A convex optimization problem, as defined in [11, Ch. 4], is of the form

$$\begin{aligned} &\text{minimize} \quad f_0(x) \\ &\text{subject to} \quad f_i(x) \le b_i \quad i = 1, \cdots, k \\ &\qquad\qquad\quad g_j(x) = 0 \quad j = 1, \cdots, l \end{aligned} \tag{12}$$

where $f_0(x)$ and $\{f_i(x)\}_{i=1}^k$ are convex functions, and $\{g_j(x)\}_{j=1}^l$ are affine.

In order to convert the optimization problem in (10) into a convex optimization problem, as presented in (12), we need to show that the set of conditional pmfs $P(x^n||y^{n-1})$ can be stated using inequalities that contains only convex function and equalities that contains affine function.

Lemma 1 (Causal conditioning is a polyhedron): The set of all causal conditioning distributions of the form $P(x^n||y^{n-1})$ is a polyhedron in $\mathbb{R}^{|\mathcal{X}|^n|\mathcal{Y}|^{n-1}}$ and is given by the following linear equalities and inequalities:

$$\begin{array}{ll} p(x^n||y^{n-1}) \ge 0, & \forall x^n, y^{n-1}, \\ \sum_{x_{i+1}^n} p(x^n||y^{n-1}) = \gamma_{x^i,y^{i-1}}, & \forall x^i, y^{n-1}, i \ge 1, \\ \sum_{x_1^n} p(x^n||y^{n-1}) = 1, & \forall y^{n-1}. \end{array} \tag{13}$$

Note that the two equalities in (13) may be unified to one if we add $i = 0$ to the equality cases and we restrict the corresponding γ to be unity. Furthermore, for $n = 1$ we obtain the regular vector probability, i.e., $p(x) \ge 0$, $\forall x$ and $\sum_x P(x) = 1$.

Note that the optimization problem given in (10) is a convex optimization problem since the set of causal conditioning pmfs is a polyhedron (Lemma 1) and the directed information is concave in $P(x^n||y^{n-1})$ for a fixed $P(y^n||x^n)$ [20, Lemma 2]. Therefore, the KKT conditions [11, Ch 5.5.3] are necessary and sufficient. The next lemma states these conditions explicitly.

Theorem 2: A set of necessary and sufficient conditions for an input probability $P(x^n||y^{n-1})$ to maximize the optimization problem in (9) is that for some numbers $\beta_{y^{n-1}}$

$$\begin{aligned} &\sum_{y_n} p(y^n||x^n) \log \frac{p(y^n||x^n)}{ep(y^n)} = \beta_{y^{n-1}}, \text{ if } p(x^n||y^{n-1}) > 0, \\ &\sum_{y_n} p(y^n||x^n) \log \frac{p(y^n||x^n)}{ep(y^n)} \le \beta_{y^{n-1}},, \text{ if } p(x^n||y^{n-1}) = 0, \end{aligned} \tag{14}$$

where $p(y^n) = \sum_{x^n} p(y^n||x^n)p(x^n||y^{n-1})$. Furthermore, the solution of the optimization problem is

$$\max_{P(x^n||y^{n-1})} I(X^n \to Y^n) = \sum_{y^{n-1}} \beta_{y^{n-1}} + 1. \tag{15}$$

For $n = 1$ we obtain a known result proved by Gallager [19, Theorem 4.5.1] that states that a sufficient and necessary condition for $P^*(x)$ that achieves $\max_{P(x)} I(X;Y)$ is that

$$\sum_y p(y|x) \log \frac{p(y|x)}{p(y)} = C, \quad \forall x \text{ if } p^*(x) > 0, \tag{16}$$

and

$$\sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \le C, \quad \forall x \text{ if } p(x) = 0, \tag{17}$$

and furthermore $C = \max_{P(x)} I(X;Y)$.

The next corollary is the main tool we use in the paper to prove that the feedback capacity and the non feedback capacity of the simple POST channel are equal.

Corollary 3: Let $P^*(x^n||y^{n-1})$ be a pmf that all its elements are positive and that achieves the maximum of $\max_{P(x^n||y^{n-1})} I(X^n \to Y^n)$ and let $P^*(y^n)$ be the output probability induced by $P^*(x^n||y^{n-1})$. If there exists an input probability distribution $P(x^n)$ such that

$$p^*(y^n) = \sum_{x^n} p(y^n||x^n)p(x^n), \tag{18}$$

then the feedback capacity and the nonfeedback capacity are the same.

Proof: Note that the sufficient and necessary condition given in (14) depends only on the channel causal conditioning pmf $P(y^n||x^n)$ and the output pmf $P(y^n)$. Furthermore, note that if (14) is satisfied then

$$\sum_{y^n} p(y^n||x^n) \log \frac{p(y^n||x^n)}{ep(y^n)} = \sum_{y^{n-1}} \beta_{y^{n-1}}, \quad \forall x^n, \tag{19}$$

and since for the non-feedback case $p(y^n||x^n) = p(y^n|x^n)$, $\forall\ (x^n, y^n)$, we obtain

$$\sum_{y^n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{ep(y^n)} = \sum_{y^{n-1}} \beta_{y^{n-1}}, \quad \forall x^n. \tag{20}$$

This means that the KKT conditions of $\max_{P(x^n)} I(X^n;Y^n)$ are satisfied. Furthermore, the maximum value for both optimization problems is $\sum_{y^{n-1}} \beta_{y^{n-1}} + 1$ and, therefore, they are equal. ∎

IV. Capacity of the simple POST channel with and without feedback

Lemma 4 (Feedback capacity): The feedback capacity of the simple POST channel is the same as of the memoryless Z channel, i.e., $-\log_2 0.8$.

Note that the induced Y_i is a Markov chain with transition probability 0.2 (as X_i only depends on past X^{i-1}, Y^{i-1} through only Y_{i-1}). Now, we are interested in writing the conditional pmf of the simple POST channel as a recursive formula. This recursive formula will be used later to find an input distribution that is ignorant of the feedback for the case of POST Channel without feedback and achieves the same output distribution, namely, a Markov chain with transition probability 0.2.

Table I presents the conditional pmf of the simple POST channel when $n = 1$. Let us denote by $P_{n,0}$ and $P_{n,1}$ the conditional matrices of the channel given the respective initial state, i.e., $s_0 = y_0$ is 0 and 1, respectively. Namely,

$$\begin{aligned} P_{n,0} &\triangleq P(y^n \| x^n, s_0 = 0) \\ P_{n,1} &\triangleq P(y^n \| x^n, s_0 = 1). \end{aligned}$$

Y_1 \ X_1	0	1
0	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$

Y_1 \ X_1	0	1
0	$\frac{1}{2}$	0
1	$\frac{1}{2}$	1

TABLE I
CONDITIONAL PROBABILITIES $P(Y_1|X_1, s_0 = 0)$ (ON TOP) AND $P(Y_1|X_1, s_0 = 1)$ (ON THE BOTTOM).

The columns of the matrices $P_{n,0}$ and $P_{n,1}$ are denoted by $x^n = (x_1, x_2, ..., x_n)$ and the rows by $y^n = (y_1, y_2, ..., y_n)$, where x_1 and y_1 are the most significant bits and x_n and y_n are the least significant bits. For instance, the conditional probabilities $P(y_1|x_1, s_0 = 0)$ and $P(y_1|x_1, s_0 = 1)$ are given in Table I. Hence $P_{n,0}$ and $P_{n,1}$ for $n = 1$ are given by

$$P_{1,0} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix} \qquad P_{1,1} = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \tag{21}$$

Furthermore, $P(y^2 \| x^2, s_0 = 0)$ is depicted in Table II and $P_{2,0}$ is given in eq. (22).

$$P_{2,0} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \tag{22}$$

From the channel definition the following recursive relation holds

$$P_{n,0} = \begin{bmatrix} 1 \cdot P_{n-1,0} & \frac{1}{2} \cdot P_{n-1,0} \\ 0 \cdot P_{n-1,1} & \frac{1}{2} \cdot P_{n-1,1} \end{bmatrix} \tag{23}$$

Y_1Y_2 \ X_1X_2	00	01	10	11
00	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
01	0	$\frac{1}{2}$	0	$\frac{1}{4}$
10	0	0	$\frac{1}{4}$	0
11	0	0	$\frac{1}{4}$	$\frac{1}{2}$

TABLE II
CONDITIONAL PROBABILITIES $P(y^2 \| x^2, s_0 = 0)$.

and

$$P_{n,1} = \begin{bmatrix} \frac{1}{2} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\ \frac{1}{2} \cdot P_{n-1,1} & 1 \cdot P_{n-1,1} \end{bmatrix} \tag{24}$$

where $P_{0,0} = P_{0,1} = 1$.

The output process $\{Y_i\}_{i \geq 1}$ induced by the input that achieves the feedback capacity, is a binary symmetric Markov chain with transition probability 0.2. Hence for $n = 1$ we have

$$\begin{aligned} p_0(y_1 = 0) = p_1(y_1 = 1) = 0.8 \\ p_0(y_1 = 1) = p_1(y_1 = 0) = 0.2. \end{aligned} \tag{25}$$

and for $n \geq 2$,

$$\begin{aligned} &\text{if } y_n = y_{n-1}, \text{then } p_0(y_n|y^{n-1}) = p_1(y_n|y^{n-1}) = 0.8 \\ &\text{if } y_n \neq y_{n-1}, \text{then } p_0(y_n|y^{n-1}) = p_1(y_n|y^{n-1}) = 0.2. \end{aligned} \tag{26}$$

Now, we present this Markov process in a recursive way. Recall that $P(y^n)$ denotes a (column) probability vector of size 2^n.

Lemma 5: A process Y^n that is binary symmetric Markov with transition probability 0.2 can be described as follows

$$P_0(y^n) = \begin{bmatrix} 0.8P_0(y^{n-1}) \\ 0.2P_1(y^{n-1}) \end{bmatrix} \qquad P_1(y^n) = \begin{bmatrix} 0.2P_0(y^{n-1}) \\ 0.8P_1(y^{n-1}) \end{bmatrix}, \tag{27}$$

where $P_0(y^0) = P_1(y^0) = 1$, and $\begin{bmatrix} P(y^{n-1}) \\ P(y^{n-1}) \end{bmatrix}$ denotes a concatenation of two columns vectors of size 2^{n-1} into one columns vector of size 2^n.

According to Corollary 3, in order to show that the non-feedback capacity equals the feedback capacity, it suffices to show that there exists an input probability $P(x^n)$ that induces the optimal $P^*(y^n)$ of the feedback case, which is the binary symmetric Markov chain with transition probability 0.2.

We find such an input probability $P(x^n)$ by calculating $P_1(x^n) = P_{n,1}^{-1} P_1(y^n)$ and $P_0(x^n) = P_{n,0}^{-1} P_0(y^n)$. We obtain

$$\begin{aligned} P_0(x^n) &= \begin{bmatrix} 0.8P_0(x^{n-1}) - 0.2P_1(x^{n-1}) \\ 0.4P_1(x^{n-1}) \end{bmatrix}, \\ P_1(x^n) &= \begin{bmatrix} 0.4P_0(x^{n-1}) \\ 0.8P_1(x^{n-1}) - 0.2P_0(x^{n-1}) \end{bmatrix}, \end{aligned} \tag{28}$$

where $P_0(x^0) = P_1(x^0) = 1$.

We need to show that $P_0(x^n)$ and $P_1(x^n)$ that are defined recursively in (28) are, indeed, pmfs, namely, they sum to one and are positive for every x^n. The fact that they sum to 1 is trivial and follows from the initial definition and the coefficients (0.4, 0.8, -0.2) that sums to 1. To show the positivity, we need to show that $0.8P_0(x^{n-1}) - 0.2P_1(x^{n-1})$ and $0.8P_1(x^{n-1}) - 0.2P_0(x^{n-1})$ are positive for all x^{n-1}. We show this by proving,

$$\begin{aligned} 4p_1(x^{n-1}) &\geq p_0(x^{n-1}),\ \forall x^{n-1}, \\ 4p_0(x^{n-1}) &\geq p_1(x^{n-1}),\ \forall x^{n-1}, \end{aligned} \tag{29}$$

as shown in the next lemma (Lemma 6).

Lemma 6: For any $1 + \sqrt{\frac{1}{2}} \leq \beta \leq 2 + \sqrt{2}$, we have the condition

$$\begin{aligned} \beta p_1(x^n) &\geq p_0(x^n),\ \forall x^n, \\ \beta p_0(x^n) &\geq p_1(x^n),\ \forall x^n. \end{aligned} \tag{30}$$

Based on Corollary 3 and the fact that we found a valid input pmf (28) that does not use feedback and induces the same output probability as the optimal conditional pmf that achieves the feedback capacity we conclude the following result.

Corollary 7: Feedback does not increase the capacity of the simple POST channel.

V. Conclusion and further research

In this paper we presented a new channel we termed the "simple POST channel" and showed, somewhat surprisingly, that feedback does not increase its capacity.

In [21] we have also consider the POST channel with the following behavior

$$\begin{aligned} &\text{if } X_i = Y_{i-1}, \text{ then } Y_i = X_i, \\ &\qquad \text{else } Y_i = X_i \oplus Z_i,\ \text{where } Z_i \sim \text{Bernnouli}\,(\alpha)\,. \end{aligned} \tag{31}$$

We call it POST channel with parameter α and its behavior is depicted in Fig. 1. When $y_{i-1} = 0$, then the channel behaves as Z channel with parameter α and when $y_{i-1} = 1$ then it behaves as an S channel with parameter α. We are able

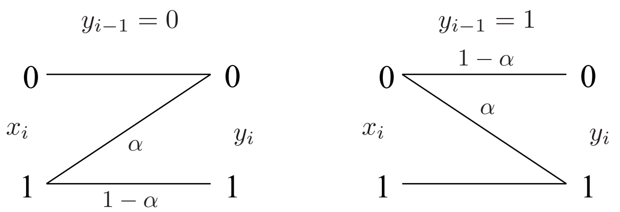

Fig. 3. POST channel with parameter α. If $y_{i-1} = 0$ then the channel behaves as an Z channel with parameter α and if $y_{i-1} = 1$ then it behaves an S channel with parameter α.

to prove, via an analysis similar to that presented here, that feedback does not increase the capacity for any $0 \leq \alpha \leq 1$. Furthermore, we have observed, numerically, that this property holds for more general binary POST channels and we are in the process of characterizing this general family of channels.

Our proof in this paper is based on finding the output probability that is induced by the input causal conditioning pmf that optimizes the directed information when feedback is allowed, and then proving that this output pmf can be achieved by an input distribution without feedback. We suspect there may be a more direct way, that has thus far eluded us, for proving that feedback does not increase the capacity of the Simple POST channel. We hope that the POST channel introduced in this paper will enhance our understanding of capacity of finite state channels with and without feedback, and help us to find simple capacity-achieving codes.

Acknowledgements

The work was supported by NSF center for Science of Information. In addition Haim Permuter was partially supported by ISF grant 684/11 and Himanshu Asnani by The Scott A. and Geraldine D. Macomber Stanford Graduate Fellowship Fund.

References

[1] R. E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory*, 18:460–473, 1972.

[2] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory*, 18:14–20, 1972.

[3] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[4] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Izv. Akad. Nauk, Moskva, 1960. in Russian, translated by A. Feinstein in 1964.

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

[6] Y.-H. Kim. Feedback capacity of stationary Gaussian channels. *IEEE Trans. Inf. Theory.*, 56(1):57–85, 2010.

[7] H. H. Permuter, P. Cuff, B. Van Roy, and T. Weissman. Capacity of the trapdoor channel with feedback. *IEEE Trans. Inf. Theory*, 54(7):3150–3165, 2009.

[8] O. Elishco and H. H. Permuter. Capacity and coding for the Ising channel with feedback. submitted to *IEEE Trans. Inf. Theory.* Available at arxiv.org/abs/1205.4674, 2012.

[9] T. Berger and F. Bonomi. Capacity and zero-error capacity of Ising channels. *IEEE Trans. Inf. Theory*, 36:173–180, 1990.

[10] H. Asnani, H. H. Permuter, and T. Weissman. To feed or not to feed back. 2010. submitted to *IEEE Trans. Inf. Theory.* Available at arxiv.org/abs/1011.1607.

[11] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New-York, 2004.

[12] G. Kramer. *Directed information for channels with feedback*. Ph.D. dissertation, Swiss Federal Institute of Technology (ETH) Zurich, 1998.

[13] J. Massey. Causality, feedback and directed information. *Proc. Int. Symp. Inf. Theory Applic. (ISITA-90)*, pages 303–305, Nov. 1990.

[14] H. Marko. The bidirectional communication theory- a generalization of information theory. *IEEE Trans. on communication*, COM-21:1335–1351, 1973.

[15] J. Chen and T. Berger. The capacity of finite-state Markov channels with feedback. *IEEE Trans. Inf. Theory*, 51:780–789, 2005.

[16] Y.-H. Kim. A coding theorem for a class of stationary channels with feedback. *IEEE Trans. Inf. Theory.*, 25:1488–1499, April, 2008.

[17] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Trans. Inf. Theory*, 55:323–349, 2009.

[18] H. H. Permuter, T. Weissman, and A. J. Goldsmith. Finite state channels with time-invariant deterministic feedback. *IEEE Trans. Inf. Theory*, 55(2):644–662, 2009.

[19] R. G. Gallager. *Information theory and reliable communication*. Wiley, New York, 1968.

[20] I. Naiss and H. H. Permuter. Extension of the Blahut-Arimoto algorithm for maximizing directed information. *IEEE Trans. Inf. Theory*, 59:204–222, 2013.

[21] H. Asnani, H. H. Permuter, and T. Weissman. The capacity of POST channels. In preperation, 2013.