

Capacity of a POST Channel With and Without Feedback

Haim Henri Permuter, *Senior Member, IEEE*, Himanshu Asnani, and Tsachy Weissman, *Fellow, IEEE*

Abstract—We consider finite state channels, where the state of the channel is its previous output. We refer to these as Previous Output is the State (POST) channels. We first focus on POST(α) channels. These channels have binary inputs and outputs, where the state determines if the channel behaves as a Z or an S channel, both with parameter α . We show that the nonfeedback capacity of the POST(α) channel equals its feedback capacity, despite the memory of 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 does not utilize the feedback. We show that this is a sufficient condition for the feedback capacity to equal the nonfeedback capacity for any finite state channel. We show that the result carries over from the POST(α) channel to a binary POST channel, where the previous output determines whether the current channel will be binary with parameters (a, b) or (b, a) . Finally, we show that, in general, feedback may increase the capacity of a POST channel.

Index Terms—Causal conditioning, convex optimization, channels with memory, directed information, feedback capacity, finite state channel, KKT conditions, POST channel.

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. These 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 are known

Manuscript received September 21, 2013; revised March 3, 2014; accepted July 15, 2014. Date of publication July 25, 2014; date of current version September 11, 2014. This work was supported in part by the Center for Science of Information under Grant and in part by the National Science Foundation under Grant 1157849-1-QAZCC. H. Permuter was supported in part by the European Research Council and in part by the Israel Science Foundation under Grant 684/11. H. Asnani was supported in part by the Scott A. and Geraldine D. Macomber Stanford and in part by the Graduate Fellowship Fund. This paper was presented at the 2013 IEEE International Symposium on Information Theory.

H. H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, BeerSheva 84105, Israel (e-mail: haimp@bgu.ac.il).

H. Asnani is with the Research and Development, Ericsson, Silicon Valley, San Jose, CA 95134 USA (e-mail: himanshu.asnani@ericsson.com).

T. Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: tsachy@stanford.edu).

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

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2014.2343232

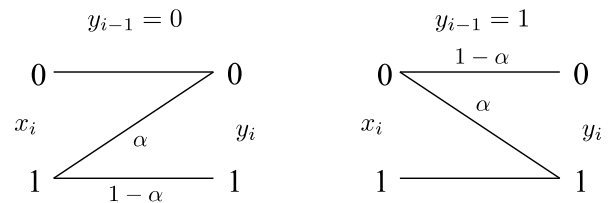


Fig. 1. POST(α): If $y_{i-1} = 0$ then the channel behaves as a Z channel with parameter α and if $y_{i-1} = 1$ then it behaves as an S channel with parameter α .

for only a few channels, 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 the modulo-additive noise channel, the additive noise channel where the noise is a first-order autoregressive moving-average Gaussian process [6], the trapdoor channel [7], and the Ising Channel [8]. If the state is known at the decoder, then knowledge of the state at the encoder can be considered as partial feedback, as considered and solved in [9] and [10].

In this paper we introduce and study a new family of channels that 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 family of POST channels that have binary inputs $\{X_i\}_{i \geq 1}$ and binary outputs $\{Y_i\}_{i \geq 1}$ related as follows:

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

We call these channels POST(α) and their behavior is depicted in Fig. 1. When $y_{i-1} = 0$, the current channel behaves as a Z channel with parameter α and when $y_{i-1} = 1$, it behaves as an S channel with parameter α . We refer to POST($\frac{1}{2}$) as the *simple* POST channel.

The simple POST channel is similar to the Ising channel introduced by Berger and Bonomi [11], 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” [12]. 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 channel, let us first consider a channel with binary i.i.d. states S_i , distributed Bernoulli($\frac{1}{2}$),

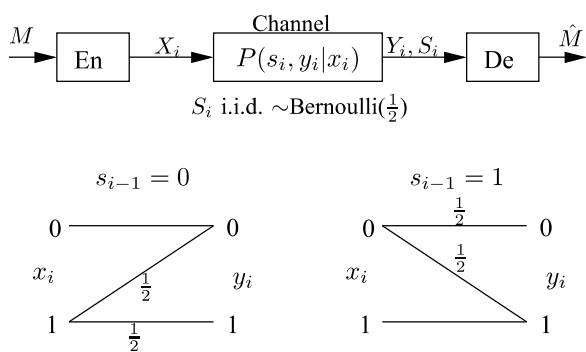


Fig. 2. A channel similar to the simple POST channel, except that the channel state, $\{S_i\}_{i \geq 1}$ is i.i.d. Bernoulli($\frac{1}{2}$) independent of the input.

where the channel behaves similarly to the simple POST channel. When $S_{i-1} = 0$, then the current 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 hint that feedback increases the capacity of the POST(α) channel as well. Indeed, our initial interest in this channel was due to this belief, in our quest for a channel with memory that would be amenable to analysis under the “to feed or not to feed” framework of constrained feedback in [12], 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 [12] suggested that feedback does not increase the capacity of the simple POST channel. This paper stemmed from our attempts to make sense of these observations.

In order to prove that feedback does not increase the capacity of some families of POST channels, 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 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 $P^*(x^n || y^{n-1})$ are the same. This sufficient condition that we establish, in the generality of any finite state channel, follows from the KKT conditions [13, Ch. 5] for convex optimization problems.

The remainder of the paper is organized as follows. In Section II, we briefly present the definitions of directed information 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 convex. Additionally, using the KKT conditions, we show that if the output distribution induced by the conditional pmfs that achieve the maximum in the presence of feedback can also be induced by an 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(α) 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 for the feedback case. In Section V we consider a binary POST(a, b) channel with two states; in each state there is a binary channel and the channels have opposite parameters. The binary POST(a, b) channel generalizes the POST(α) channel and we show that feedback does not increase capacity for this considerably larger class of channels as well. In Section VI, we show that, unlike the POST(a, b), feedback may increase the capacity of POST channels in general. In Section VII, we conclude and suggest some directions for further research on the family of POST channels.

II. DIRECTED INFORMATION, CAUSAL CONDITIONING AND NOTATIONS

Throughout this paper, we denote random variables by capital letters such as X . The probability $\Pr\{X = x\}$ is denoted by $p(x)$. We denote the whole vector of probabilities by capital P , i.e., $P(x)$ is the probability vector of the random variable X .

We use the *causal conditioning* notation $(\cdot || \cdot)$ developed by Kramer [14]. 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

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

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 $d = 0, 1$:

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

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

The directed information was defined by Massey [15], inspired by Marko's work [16] on bidirectional communication, as

$$I(X^n \rightarrow Y^n) := \sum_{i=1}^n I(X^i; Y_i | Y^{i-1}). \quad (4)$$

The directed information can also be rewritten as

$$I(X^n \rightarrow Y^n) = \sum_{x^n, y^n} p(x^n \| y^{n-1}) p(y^n \| x^n) \log \frac{p(y^n \| x^n)}{\sum_{x^n} p(x^n \| y^{n-1}) p(y^n \| x^n)}. \quad (5)$$

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

$$p(x^n, y^n) = p(x^n \| y^{n-1}) p(y^n \| x^n). \quad (6)$$

We will make use the fact that directed information $I(X^n \rightarrow Y^n)$ is concave in $P(x^n \| y^{n-1})$ for a fixed $P(y^n \| x^n)$, which is proved in Lemma 15 in Appendix A.

Directed information characterizes the capacity of point-to-point channels with feedback [10], [17]–[19]. For channels where the state is a function of the output, of which the POST channel is a special case, it was shown [7], [10] that the feedback capacity is given by

$$C_{fb} = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{P(x^n \| y^{n-1})} I(X^n \rightarrow Y^n). \quad (7)$$

On the other hand, without feedback the capacity is given by

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{P(x^n)} I(X^n \rightarrow Y^n), \quad (8)$$

since the channel is indecomposable¹ [20]. In the case where there is no feedback, namely, the Markov form $X_i - X^{i-1} - Y^{i-1}$ holds, $I(X^n \rightarrow Y^n) = I(X^n; Y^n)$, as shown in [15].

III. MAXIMIZATION OF THE DIRECTED INFORMATION AS A CONVEX OPTIMIZATION PROBLEM

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

$$\max_{P(x^n \| y^{n-1})} I(X^n \rightarrow Y^n) \quad (9)$$

and

$$\max_{P(x^n)} I(X^n \rightarrow Y^n). \quad (10)$$

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

A convex optimization problem, as defined in [13, Ch. 4], is a problem of the form

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq b_i \quad i = 1, \dots, k \\ & && g_j(x) = 0 \quad j = 1, \dots, l \end{aligned} \quad (11)$$

¹An indecomposable FSC is a FSC for which the influence of the initial state on the conditional probability dies away with time [20, Ch. 4]. A sufficient and necessary condition [20, Th. 4.6.3] that the channel is indecomposable is that for a fixed n , and for each x^n , there exists an s_n that may depend on x^n such that $P(s_n | x^n, s_0) > 0$ for all s_0 . Trivially, for the channels that are investigated in this paper (the POST(α) where $\alpha > 0$ or POST(a, b) where $a > 0$ and $b > 0$), if we choose $n = 1$ and $s_1 = x_1$ we obtain that the condition holds for all s_0 .

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

In order to convert the optimization problem in (9) into a convex optimization problem, as presented in (11), we need to show that the set of conditional pmfs $P(x^n \| y^{n-1})$ can be expressed using inequalities that contains only convex functions and equalities that contains affine functions.

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 a set of linear equalities and inequalities of the form:

$$\begin{aligned} p(x^n \| y^{n-1}) &\geq 0, && \forall x^n, y^{n-1}, \\ \sum_{x_{i+1}^n} p(x^n \| y^{n-1}) &= \gamma_{x^i, y^{i-1}}, && \forall x^i, y^{i-1}, i \geq 1, \\ \sum_{x_1^n} p(x^n \| y^{n-1}) &= 1, && \forall y^{n-1}. \end{aligned} \quad (12)$$

Note that the two equalities in (12) may be unified into 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., $p(x) \geq 0, \forall x$ and $\sum_x P(x) = 1$.

Proof: It is straightforward to see that every causal conditioning $P(x^n \| y^n)$ satisfies these equalities and inequalities. Now, we need to show that if an element in $\mathbb{R}^{|\mathcal{X}^n| |\mathcal{Y}^{n-1}|}$, call it $P(x^n \| y^{n-1})$, satisfies (12), then $P(x^n \| y^{n-1})$ is a causal conditioning pmf, namely, there exists a sequence of regular conditioning $\{P(x_i | x^{i-1}, y^{i-1})\}_{i=1}^n$ such that $p(x^n \| y^{n-1}) = \prod_{i=1}^n p(x_i | x^{i-1}, y^{i-1})$.

Let us define for all x^i, y^{i-1}

$$\begin{aligned} p(x_i | x^{i-1}, y^{i-1}) &\triangleq \frac{\sum_{x_{i+1}^n} p(x^n \| y^{n-1})}{\sum_{x_i} p(x^n \| y^{n-1})} \\ &= \frac{\gamma_{x^i, y^{i-1}}}{\gamma_{x^{i-1}, y^{i-2}}}. \end{aligned} \quad (13)$$

when $\gamma_{x^{i-1}, y^{i-2}} > 0$ and zero otherwise. Now, note that the vector probability $P(x_i | x^{i-1}, y^{i-1})$ defined in (13) satisfies $p(x_i | x^{i-1}, y^{i-1}) \geq 0$, and $\sum_{x_i} p(x_i | x^{i-1}, y^{i-1}) = 1$. Furthermore, observe that

$$\begin{aligned} \prod_{i=1}^n p(x_i | x^{i-1}, y^{i-1}) &= \prod_{i=1}^n \frac{\gamma_{x^i, y^{i-1}}}{\gamma_{x^{i-1}, y^{i-2}}} \\ &= \frac{\gamma_{x^n, y^{n-1}}}{1} \\ &= \gamma_{x^n, y^{n-1}} \\ &= p(x^n \| y^{n-1}), \end{aligned} \quad (14)$$

when $\gamma_{x^i, y^{i-1}} > 0$ for $i = 1, \dots, n$, and if for some i , $\gamma_{x^i, y^{i-1}} = 0$, it implies that $p(x_i | x^{i-1}, y^{i-1}) = 0$ and $p(x^n \| y^{n-1}) = 0$. ■

Note that the optimization problem given in (9) is a convex optimization one 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 $P(y^n \| x^n)$ [21, Lemma 2]. Therefore, the KKT conditions [13, Ch. 5.5.3] are necessary and sufficient. The next theorem states these conditions explicitly for our setting.

Theorem 2 (Conditions for Max. the Directed Information): A set of necessary and sufficient conditions for an input probability $P(x^n \| y^{n-1})$ to achieve the maximum in (9) is

that for some numbers $\beta_{y^{n-1}}, \forall x^n, y^{n-1}$

$$\begin{aligned} \sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} &= \beta_{y^{n-1}}, \quad \text{if } p(x^n || y^{n-1}) > 0, \\ \sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} &\leq \beta_{y^{n-1}}, \quad \text{if } p(x^n || y^{n-1}) = 0, \end{aligned} \quad (15)$$

where $p(y^n) = \sum_{x^n} p(y^n || x^n) p(x^n || y^{n-1})$ and $e \approx 2.718$ is Euler's number. Furthermore, the maximum is given by

$$\max_{P(x^n || y^{n-1})} I(X^n \rightarrow Y^n) = \sum_{y^{n-1}} \beta_{y^{n-1}} + 1. \quad (16)$$

For $n = 1$ we obtain a known result proved by Gallager [20, Th. 4.5.1] that states that a sufficient and necessary condition for $P^*(x)$ to achieve $\max_{P(x)} I(X; Y)$ is that

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

and

$$\sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \leq C, \quad \forall x \text{ if } p^*(x) = 0, \quad (18)$$

for some C . Furthermore, $C = \max_{P(x)} I(X; Y)$.

Proof: Using the fact that a causal conditioning pmf is a polyhedron (Lemma 1), we can write the maximization of the directed information as a standard convex optimization problem:

minimize

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

subject to

$$-p(x^n || y^{n-1}) \leq 0, \quad \forall x^n, y^{n-1}, \quad (20)$$

$$\sum_{x_{i+1}^n} p(x^n || y^{n-1}) - \gamma_{x^i, y^{i-1}} = 0, \quad \forall x^i, y^{i-1}, i \geq 1 \quad (21)$$

$$\sum_{x^n} p(x^n || y^{n-1}) = 1, \quad \forall y^{n-1}. \quad (22)$$

The Lagrangian is defined as

$$\begin{aligned} L = & -\sum_{x^n, y^n} p(x^n || y^{n-1}) p(y^n || x^n) \log \frac{p(y^n || x^n)}{\sum_{x^n} p(x^n || y^{n-1}) p(y^n || x^n)} \\ & - \sum_{x^n, y^{n-1}} \lambda_{x^n, y^{n-1}} p(x^n || y^{n-1}) \\ & + \sum_{i=1}^n \sum_{x^i, y^{i-1}} v_{x^i, y^{i-1}} \left(\sum_{x_{i+1}^n} p(x^n || y^{n-1}) - \gamma_{x^i, y^{i-1}} \right) \\ & + \sum_{y^{n-1}} v_{y^{n-1}} \left(\sum_{x^n} p(x^n || y^{n-1}) - 1 \right), \end{aligned} \quad (23)$$

where $\lambda_{x^n, y^{n-1}} \geq 0$. The KKT conditions for this problem

are

$$\begin{aligned} \frac{\partial}{\partial p(x^n || y^{n-1})} &= 0, \quad \forall x^n, y^{n-1}, \\ \frac{\partial}{\partial \gamma_{x^i, y^{i-1}}} &= 0, \quad \forall x^i, y^{i-1}, \\ \lambda_{x^n, y^{n-1}} &\geq 0, \quad \forall x^n, y^{n-1}, \\ \lambda_{x^n, y^{n-1}} p(x^n || y^{n-1}) &= 0, \quad \forall x^n, y^{n-1} \end{aligned} \quad (24)$$

and that $P(x^n || y^{n-1})$ is a valid causal conditioning pmf, namely, satisfies the constraint of (19). Now, we need to show that the KKT conditions described above are equivalent to (15). Let us compute the derivatives in order to write the conditions in (24) more explicitly:

$$\begin{aligned} \frac{\partial L}{\partial p(x^n || y^{n-1})} &= -\sum_{y^n} p(y^n || x^n) \log P(y^n || x^n) \\ &\quad - \sum_{y^n} p(y^n || x^n) \log \frac{1}{\sum_{x^n} p(x^n || y^{n-1}) P(y^n || x^n)} \\ &\quad + \sum_{x^n, y^n} p(x^n || y^{n-1}) P(y^n || x^n) \frac{p(y^n || x^n)}{\sum_{\tilde{x}^n} p(\tilde{x}^n || y^{n-1}) P(y^n || \tilde{x}^n)} \\ &\quad - \lambda_{x^n, y^{n-1}} + \sum_{i=0}^n v_{x^i, y^{i-1}} + v_{y^{n-1}} \\ &= -\sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{\sum_{x^n} p(x^n || y^{n-1}) p(y^n || x^n)} \\ &\quad + \sum_{y^n} p(y^n || x^n) - \lambda_{x^n, y^{n-1}} + \sum_{i=0}^n v_{x^i, y^{i-1}} + v_{y^{n-1}} \\ &= -\sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} - \lambda_{x^n, y^{n-1}} \\ &\quad + \sum_{i=0}^n v_{x^i, y^{i-1}} + v_{y^{n-1}} \end{aligned} \quad (25)$$

$$\frac{\partial L}{\partial \gamma_{x^i, y^{i-1}}} = -\sum_{i=0}^n v_{x^i, y^{i-1}}. \quad (26)$$

Hence, the KKT conditions given in (24) become $\forall x^n, y^{n-1}$

$$\begin{aligned} \sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} &= v_{y^{n-1}} \quad \text{if } P(x^n || y^{n-1}) > 0, \\ \sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} &\leq v_{y^{n-1}} \quad \text{if } p(x^n || y^{n-1}) = 0, \end{aligned} \quad (27)$$

which is exactly (15). Now, to obtain (16) we use (27) and observe that

$$\begin{aligned} \sum_{x^n} \sum_{y^{n-1}} p(x^n || y^{n-1}) \sum_{y^n} p(y^n || x^n) \log \frac{p(y^n || x^n)}{ep(y^n)} \\ = \sum_{x^n} \sum_{y^{n-1}} p(x^n || y^{n-1}) v_{y^{n-1}}. \end{aligned} \quad (28)$$

Indeed, the LHS of (28) equals to $I(X^n \rightarrow Y^n) - 1$. The RHS equals to $\sum_{y^{n-1}} \nu_{y^{n-1}}$, since $\sum_{x^n} p(x^n|y^{n-1}) = 1$ for all y^{n-1} and, therefore, (28) implies that (16) holds. ■

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

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

$$p^*(y^n) = \sum_{x^n} p(y^n|x^n)p(x^n), \quad (29)$$

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

Proof: Note that the sufficient and necessary condition given in (15) depends only on the channel causal conditioning pmf $P(y^n|x^n)$ and the output pmf $P(y^n)$. Furthermore, note that if (15) is satisfied then

$$\sum_{y^n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{ep(y^n)} = \sum_{y^{n-1}} \beta_{y^{n-1}}, \quad \forall x^n, \quad (30)$$

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

$$\sum_{y^n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{ep(y^n)} = \sum_{y^{n-1}} \beta_{y^{n-1}}, \quad \forall x^n. \quad (31)$$

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 POST(α) CHANNEL WITH AND WITHOUT FEEDBACK

Lemma 4 (Feedback Capacity): The feedback capacity of the POST(α) channel is the same as of the memoryless Z channel with parameter α , which is $C = -\log_2 c$ where

$$c = (1 + \bar{\alpha}\alpha^{\frac{\alpha}{1-\alpha}})^{-1}, \quad (32)$$

and $\bar{\alpha} \triangleq 1 - \alpha$.

The behavior of the capacity as a function of α is depicted in Fig. 3. *Proof of the achievability and the capacity of Z channel with parameter α :* the achievability proof is trivial since both the encoder and the decoder know if the channel behaves as a Z or an S channel. The input probability that maximizes the mutual information, i.e. $\arg \max_{P_X} I(X; Y)$, for the memoryless Z channels with parameter α is

$$P(x=1) = c\alpha^{\frac{\alpha}{1-\alpha}} \quad P(x=0) = c(1 - \alpha^{\frac{\alpha}{1-\alpha}}), \quad (33)$$

where c is a normalization coefficient and is given in (32). The output probability for the Z channel with parameter α is

$$\begin{aligned} P(y=1) &= \bar{\alpha}P(x=1) = c\bar{\alpha}\alpha^{\frac{\alpha}{1-\alpha}} \\ P(y=0) &= 1 - c\bar{\alpha}\alpha^{\frac{\alpha}{1-\alpha}} = c. \end{aligned} \quad (34)$$

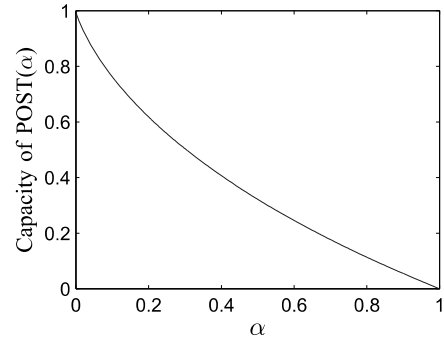


Fig. 3. The capacity of the POST(α) channel with and without feedback. This is also the capacity of the Z channel with parameter α .

The capacity of the Z channel with parameter α is

$$C = \max_{P_X} I(X; Y) = -\log_2 c, \quad (35)$$

which is also an achievable rate for the POST(α) channel with feedback. ■

Proof of Converse: The upper bound is given in the following set of equalities and inequalities:

C_{fb}

$$\begin{aligned} &\stackrel{(a)}{=} \lim_{n \rightarrow \infty} \max_{P(x^n|y^{n-1})} \frac{1}{n} I(X^n \rightarrow Y^n) \\ &\stackrel{(b)}{=} \lim_{n \rightarrow \infty} \max_{P(x^n|y^{n-1})} \frac{1}{n} I(X^n \rightarrow Y^n | Y_0) \\ &= \lim_{n \rightarrow \infty} \max_{P(x^n|y^{n-1})} \frac{1}{n} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X^i) \\ &\stackrel{(c)}{\leq} \lim_{n \rightarrow \infty} \max_{P(x^n|y^{n-1})} \frac{1}{n} \sum_{i=1}^n H(Y_i | Y_{i-1}) - H(Y_i | Y_{i-1}, X_i) \\ &= \lim_{n \rightarrow \infty} \max_{P(x^n|y^{n-1})} \frac{1}{n} \sum_{i=1}^n P(y_i=0) I(X_i; Y_i | y_{i-1}=0) \\ &\quad + P(y_i=1) I(X_i; Y_i | y_{i-1}=1) \\ &\stackrel{(d)}{=} \lim_{n \rightarrow \infty} \max_{\{P(x_i|y_{i-1})\}_{i \geq 1}} \frac{1}{n} \sum_{i=1}^n P(y_i=0) I(X_i; Y_i | y_{i-1}=0) \\ &\quad + P(y_i=1) I(X_i; Y_i | y_{i-1}=1) \\ &= \max_{P(x_i|y_{i-1}=0)} P(y_i=0) I(X_i; Y_i | y_{i-1}=0) \\ &\quad + \max_{P(x_i|y_{i-1}=1)} P(y_i=1) I(X_i; Y_i | y_{i-1}=1), \text{ for some } i \\ &\stackrel{(e)}{=} -P(y_i=0) \log_2 c - P(y_i=1) \log_2 c, \\ &= -\log_2 c, \end{aligned} \quad (36)$$

where (a) follows from the capacity formula given in [10] and [19], (b) from the inequality $|I(X^n \rightarrow Y^n) - I(X^n \rightarrow Y^n | S)| \leq H(S)$ [19, Lemma 4], (c) from the fact that conditioning reduces entropy and from the Markov chain $Y_i - (X_i, Y_{i-1}) - (X^{i-1}, Y^{i-2})$, (d) from the fact that the set of causal conditioning $P(x^n|y^{n-1})$ is equivalent to the set of $\{P(x_i|x^{i-1}, y^{i-1})\}$ and in this particular case it's enough to maximize only over $\{P(x_i|y_{i-1})\}$ because of the objective. Finally, Step (e) follows from the capacity of the memoryless Z-channel with parameter α . ■

TABLE I
CONDITIONAL PROBABILITIES $P(Y_1|X_1, s_0 = 0)$ (ON LEFT) AND
 $P(Y_1|X_1, s_0 = 1)$ (ON THE RIGHT) OF THE
POST(α) CHANNEL

	X_1	0	1
Y_1			
0		1	α
1		0	$\bar{\alpha}$
	X_1	0	1
Y_1			
0		$\bar{\alpha}$	0
1		α	1

TABLE II
CONDITIONAL PROBABILITIES $P(Y^2||X^2, s_0 = 0)$

	X_1X_2	00	01	10	11
Y_1Y_2					
00		1	α	α	α^2
01		0	$\bar{\alpha}$	0	$\alpha\bar{\alpha}$
10		0	0	$\bar{\alpha}^2$	0
11		0	0	$\bar{\alpha}\alpha$	$\bar{\alpha}$

Note that the induced Y_i is a Markov chain with transition probability $c\bar{\alpha}\alpha^{\frac{c}{\bar{\alpha}}}$ (as X_i depends on the past X^{i-1}, Y^{i-1} only through Y_{i-1}). Now, we are interested in expressing the conditional pmf of the POST(α) channel recursively. This recursive formula will be used later to find an input distribution that does not utilize the feedback for the case of a POST Channel without feedback and achieves the same output distribution, namely, a Markov chain with transition probability $c\bar{\alpha}\alpha^{\frac{c}{\bar{\alpha}}}$.

Table I presents the conditional pmf of a POST(α) 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 is 0 and 1, respectively. Namely,

$$\begin{aligned} P_{n,0} &\triangleq P(y^n||x^n, s_0 = 0) \\ P_{n,1} &\triangleq P(y^n||x^n, s_0 = 1). \end{aligned} \quad (37)$$

The columns of the matrices $P_{n,0}$ and $P_{n,1}$ are indexed by $x^n = (x_1, x_2, \dots, x_n)$ and the rows by $y^n = (y_1, y_2, \dots, y_n)$ arranged via lexicographical order, 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

$$P_{1,0} = \begin{bmatrix} 1 & \alpha \\ 0 & \bar{\alpha} \end{bmatrix} \quad P_{1,1} = \begin{bmatrix} \bar{\alpha} & 0 \\ \alpha & 1 \end{bmatrix}. \quad (38)$$

Table II presents $P(Y^2||X^2, s_0 = 0)$ and the corresponding matrix $P_{2,0}$ is given in eq. (39).

$$P_{2,0} = \begin{bmatrix} 1 & \alpha & \alpha & \alpha^2 \\ 0 & \bar{\alpha} & 0 & \alpha\bar{\alpha} \\ 0 & 0 & \bar{\alpha}^2 & 0 \\ 0 & 0 & \bar{\alpha}\alpha & \alpha \end{bmatrix} \quad (39)$$

From the channel definition, the following recursive relation holds

$$P_{n,0} = \begin{bmatrix} 1 \cdot P_{n-1,0} & \alpha \cdot P_{n-1,0} \\ 0 \cdot P_{n-1,1} & \bar{\alpha} \cdot P_{n-1,1} \end{bmatrix} \quad (40)$$

and

$$P_{n,1} = \begin{bmatrix} \bar{\alpha} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\ \alpha \cdot P_{n-1,1} & 1 \cdot P_{n-1,1} \end{bmatrix}, \quad (41)$$

where $P_{0,0} = P_{0,1} = 1$ (i.e., the one by one unit matrix).

Recall that 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 $c\bar{\alpha}\alpha^{\frac{c}{\bar{\alpha}}}$, see (34). Let $p_0(y^n)$ and $p_1(y^n)$ denote the probability of y^n given the initial state 0 and 1, respectively. Hence, for $n = 1$ we have

$$\begin{aligned} p_0(y_1 = 0) &= p_1(y_1 = 1) = c \\ p_0(y_1 = 1) &= p_1(y_1 = 0) = c\bar{\alpha}\alpha^{\frac{c}{\bar{\alpha}}}. \end{aligned} \quad (42)$$

For $n \geq 2$,

$$p_0(y_n|y^{n-1}) = p_1(y_n|y^{n-1}) = p_{y_{n-1}}(y_n). \quad (43)$$

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

Lemma 5 (pmf Vector Representation of Markov Processes): Let Y^n be binary symmetric Markov with transition probability δ . Let $P_0(y^n)$ and $P_1(y^n)$ be the vector pmf when the initial state is 0 and 1, respectively. One can describe the vector pmf using the following recursive relation

$$P_0(y^n) = \begin{bmatrix} \bar{\delta}P_0(y^{n-1}) \\ \delta P_1(y^{n-1}) \end{bmatrix} \quad \text{and} \quad P_1(y^n) = \begin{bmatrix} \delta P_0(y^{n-1}) \\ \bar{\delta}P_1(y^{n-1}) \end{bmatrix}, \quad (44)$$

where $P_0(y^0) = P_1(y^0) = 1$, and $\begin{bmatrix} u \\ v \end{bmatrix}$ denotes the column vector obtained by concatenating the two column vectors u and v .

Proof: We prove this claim by induction. Note that, for $n = 1$ (44) implies

$$\begin{aligned} p_0(y_1 = 0) &= p_1(y_1 = 1) = \bar{\delta} \\ p_0(y_1 = 1) &= p_1(y_1 = 0) = \delta. \end{aligned} \quad (45)$$

Now, we need to show that, regardless of the initial state, (43) holds. Assume $y_1 = 1$ and the initial state is 0, then for any y_2^n

$$\begin{aligned} p_0(y_n|y^{n-1}) &= \frac{p_0(y^n)}{p_0(y^{n-1})} \\ &= \frac{p_0(1, y_2^n)}{p_0(1, y_2^{n-1})} \\ &= \frac{\delta p_1(y_2^n)}{\delta p_1(y_2^{n-1})} \\ &= \frac{p_1(y_2^n)}{p_1(y_2^{n-1})} \\ &= p_1(y_n|y_2^{n-1}). \end{aligned} \quad (46)$$

In a similar way we obtain for any y_2^n and any initial state

$$p_0(y_n|y^{n-1}) = p_1(y_n|y^{n-1}) = p_{y_1}(y_n|y_2^{n-1}). \quad (47)$$

By repeating the procedure in (46) i times, we obtain that for $1 \leq i < n$ and $\forall y_2^n$

$$p_0(y_n|y^{n-1}) = p_1(y_n|y^{n-1}) = p_{y_i}(y_n|y_{i+1}^{n-1}). \quad (48)$$

Now, note that choosing $i = n - 1$ we obtain (43), which means that the process is indeed Markov with transition probability δ . ■

The following is our first main result:

Theorem 6: Feedback does not increase the capacity of the POST(α) channel.

Proof: According to Corollary 3, in order to show that the nonfeedback capacity equals the feedback capacity, it suffices to show that there exists an input pmf $P(x^n)$ that induces the optimal $P^*(y^n)$ of the feedback case, which is the binary symmetric Markov chain with transition probability $c\bar{\alpha}\alpha^{\frac{\alpha}{a}}$.

We will now find such a pmf by calculating $P_1(x^n) = P_{n,1}^{-1}P_1^*(y^n)$ and $P_0(x^n) = P_{n,0}^{-1}P_0^*(y^n)$ and verifying that $P_0(x^n)$ and $P_1(x^n)$ are indeed a valid pmf. Recall that when A and D are invertible,

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}. \quad (49)$$

Hence,

$$\begin{aligned} P_{n,0}^{-1} &= \begin{bmatrix} 1 \cdot P_{n-1,0} & \alpha \cdot P_{n-1,0} \\ 0 \cdot P_{n-1,1} & \bar{\alpha} \cdot P_{n-1,1} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} P_{n-1,0}^{-1} & -\frac{\alpha}{\bar{\alpha}} P_{n-1,1}^{-1} \\ 0 & \frac{1}{\bar{\alpha}} P_{n-1,1}^{-1} \end{bmatrix} \end{aligned} \quad (50)$$

$$\begin{aligned} P_{n,1}^{-1} &= \begin{bmatrix} \bar{\alpha} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\ \alpha \cdot P_{n-1,1} & 1 \cdot P_{n-1,1} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \frac{1}{\bar{\alpha}} P_{n-1,0}^{-1} & 0 \\ -\frac{\alpha}{\bar{\alpha}} P_{n-1,1}^{-1} & P_{n-1,1}^{-1} \end{bmatrix} \end{aligned} \quad (51)$$

Now we compute $P_1(x^n)$ and $P_0(x^n)$.

$$\begin{aligned} P_0(x^n) &= P_{n,0}^{-1}P_0(y^n) \\ &= \begin{bmatrix} P_{n-1,0}^{-1} & -\frac{\alpha}{\bar{\alpha}} P_{n-1,1}^{-1} \\ 0 & \frac{1}{\bar{\alpha}} P_{n-1,1}^{-1} \end{bmatrix} \begin{bmatrix} cP_0^*(y^{n-1}) \\ c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_1^*(y^{n-1}) \end{bmatrix} \\ &= c \begin{bmatrix} P_0(x^{n-1}) - \alpha^{\frac{1}{a}}P_1(x^{n-1}) \\ \alpha^{\frac{\alpha}{a}}P_1(x^{n-1}) \end{bmatrix}, \end{aligned} \quad (52)$$

$$\begin{aligned} P_1(x^n) &= P_{n,1}^{-1}P_1(y^n) \\ &= \begin{bmatrix} \frac{1}{\bar{\alpha}} P_{n-1,0}^{-1} & 0 \\ -\frac{\alpha}{\bar{\alpha}} P_{n-1,1}^{-1} & P_{n-1,1}^{-1} \end{bmatrix} \begin{bmatrix} c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_0^*(y^{n-1}) \\ cP_1(y^{n-1}) \end{bmatrix} \\ &= c \begin{bmatrix} \alpha^{\frac{\alpha}{a}}P_0(x^{n-1}) \\ P_1(x^{n-1}) - \alpha^{\frac{1}{a}}P_0(x^{n-1}) \end{bmatrix}, \end{aligned} \quad (53)$$

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

Now, we need to show that the probability expressions are valid, namely, nonnegative and sum to 1. The fact that they sum to 1 can be seen from the recursion immediately by verifying that $c(\alpha^{\frac{\alpha}{a}} + 1 - \alpha^{\frac{1}{a}}) = 1$.

In order to show the nonnegativity we need to show that

$$\begin{aligned} P_0(x^{n-1}) - \alpha^{\frac{1}{a}}P_1(x^{n-1}) &\geq 0 \\ P_1(x^{n-1}) - \alpha^{\frac{1}{a}}P_0(x^{n-1}) &\geq 0. \end{aligned} \quad (54)$$

For $n = 1$ this is true since $\alpha^{\frac{1}{a}} \leq 1$ for $0 \leq \alpha \leq 1$ (see Lemma 16 in the appendix). The following lemma (Lemma 7) states that if (54) holds for $n - 1$ it also holds for n and, by induction, we conclude that (54) holds for all n . ■

Lemma 7: There exists a $1 \leq \beta \leq \alpha^{-\frac{1}{a}}$, for which the condition

$$\begin{aligned} \beta P_1(x^{n-1}) &\geq P_0(x^{n-1}), \quad \forall x^{n-1}, \\ \beta P_0(x^{n-1}) &\geq P_1(x^{n-1}), \quad \forall x^{n-1}, \end{aligned} \quad (55)$$

implies

$$\begin{aligned} \beta P_1(x^n) &\geq P_0(x^n), \quad \forall x^n, \\ \beta P_0(x^n) &\geq P_1(x^n), \quad \forall x^n. \end{aligned} \quad (56)$$

Proof: Let's assume that (55) holds and we need to show that (56) holds, which is equivalent to showing the following four inequalities:

$$\beta \left(cP_0(x^{n-1}) - c\alpha^{\frac{1}{a}}P_1(x^{n-1}) \right) \geq c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_0(x^{n-1}), \quad (57)$$

$$\beta c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_1(x^{n-1}) \geq cP_1(x^{n-1}) - c\alpha^{\frac{1}{a}}P_0(x^{n-1}), \quad (58)$$

$$\beta c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_0(x^{n-1}) \geq cP_0(x^{n-1}) - c\alpha^{\frac{1}{a}}P_1(x^{n-1}), \quad (59)$$

$$\beta \left(cP_1(x^{n-1}) - c\alpha^{\frac{1}{a}}P_0(x^{n-1}) \right) \geq c\bar{\alpha}\alpha^{\frac{\alpha}{a}}P_1(x^{n-1}). \quad (60)$$

Because of the symmetry it suffices to show that (57) and (58) hold. We start by showing that (57) holds. The inequality (57) is equivalent to

$$\beta \left(P_0(x^{n-1}) - \alpha^{\frac{1}{a}}P_1(x^{n-1}) \right) \geq \alpha^{\frac{\alpha}{a}}P_0(x^{n-1}), \quad (61)$$

which can be further written as

$$P_0(x^{n-1})(\beta - \alpha^{\frac{\alpha}{a}}) \geq \beta\alpha^{\frac{1}{a}}P_1(x^{n-1}), \quad (62)$$

and as

$$P_0(x^{n-1}) \frac{\beta - \alpha^{\frac{\alpha}{a}}}{\beta\alpha^{\frac{1}{a}}} \geq P_1(x^{n-1}). \quad (63)$$

This is true using the induction assumption in (55) if

$$\frac{\beta - \alpha^{\frac{\alpha}{a}}}{\beta\alpha^{\frac{1}{a}}} \geq \beta. \quad (64)$$

Equivalently

$$\alpha^{\frac{1}{a}}\beta^2 - \beta + \alpha^{\frac{\alpha}{a}} \leq 0. \quad (65)$$

$$\frac{1 - \sqrt{1 - 4\alpha^{\frac{\alpha+1}{a}}}}{2\alpha^{\frac{1}{a}}} \leq \beta \leq \frac{1 + \sqrt{1 - 4\alpha^{\frac{\alpha+1}{a}}}}{2\alpha^{\frac{1}{a}}}, \quad (66)$$

where the condition $4\alpha^{\frac{\alpha+1}{a}} \leq 1$ can be verified to hold for all $0 \leq \alpha \leq 1$, as shown in Lemma 17 in the appendix.

Now let's consider inequality (58). We need to show that

$$\beta\alpha^{\frac{\alpha}{a}}P_1(x^{n-1}) \geq P_1(x^{n-1}) - \alpha^{\frac{1}{a}}P_0(x^{n-1}), \quad (67)$$

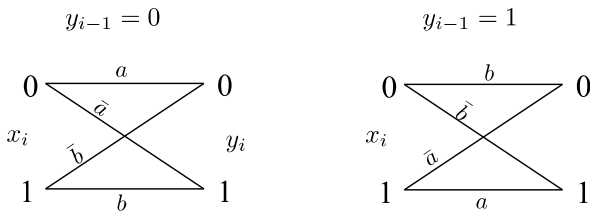


Fig. 4. POST(a, b) channel. If $y_{i-1} = 0$ then the channel behaves as DMC with parameters (a, b) and if $y_{i-1} = 1$ then the channel behaves as DMC with parameters (b, a) .

or, equivalently,

$$P_0(x^{n-1}) \geq \frac{1 - \beta \alpha^{\frac{a}{\alpha}}}{\alpha^{\frac{1}{\alpha}}} P_1(x^{n-1}). \quad (68)$$

This is true using the induction assumption in (55) if

$$\frac{1}{\beta} \geq \frac{1 - \beta \alpha^{\frac{a}{\alpha}}}{\alpha^{\frac{1}{\alpha}}}, \quad (69)$$

and equivalently

$$\alpha^{\frac{a}{\alpha}} \beta^2 - \beta + \alpha^{\frac{1}{\alpha}} \geq 0. \quad (70)$$

This holds if

$$\beta \geq \frac{1 + \sqrt{1 - 4\alpha^{\frac{a+1}{\alpha}}}}{2\alpha^{\frac{a}{\alpha}}}. \quad (71)$$

Combining (66) with (71) we obtain that there exists a β in the interval

$$\frac{1 + \sqrt{1 - 4\alpha^{\frac{a+1}{\alpha}}}}{2\alpha^{\frac{a}{\alpha}}} \leq \beta \leq \frac{1 + \sqrt{1 - 4\alpha^{\frac{a+1}{\alpha}}}}{2\alpha^{\frac{1}{\alpha}}} \quad (72)$$

that satisfies the lemma. Note that the interval is nonempty since for $0 \leq \alpha \leq 1$, $\alpha^{\frac{a}{\alpha}} \geq \alpha^{\frac{1}{\alpha}}$. Finally, note that since $\frac{1 + \sqrt{1 - 4\alpha^{\frac{a+1}{\alpha}}}}{2} \leq 1$ we obtain that $\beta \leq \alpha^{-\frac{1}{\alpha}}$. Furthermore, it is shown in Lemma 18 in the appendix that $\frac{1 + \sqrt{1 - 4\alpha^{\frac{a+1}{\alpha}}}}{2\alpha^{\frac{a}{\alpha}}} \geq 1$ for any $0 \leq \alpha \leq 1$, and, therefore, also $\beta \geq 1$. ■

V. BINARY POST(a, b) CHANNEL

In this section, we extend the scope and study the capacity of what we refer to as the POST(a, b), which is a generalization of POST(α) channel. The POST(a, b) channel has two states and in each state there is a binary channel with respective parameters (a, b) . We develop, using similar analyses as for the POST(α) channel, simple conditions on a, b for which, if satisfied, feedback does not increase capacity. We then show that the conditions are satisfied for all parameter values $(a, b) \in [0, 1] \times [0, 1]$.

A. Definition of the POST(a, b) Channel

Consider the POST channel depicted in Fig. 4 with the following behavior. When $y_{i-1} = 0$, then the channel behaves as a binary channel with transition matrix

$$\begin{bmatrix} a & \bar{b} \\ \bar{a} & b \end{bmatrix} \quad (73)$$

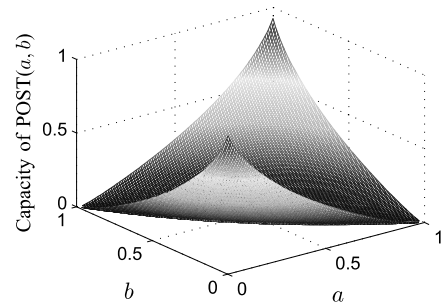


Fig. 5. The capacity of the POST(a, b) channel with and without feedback. This is also the capacity of the binary DMC with parameters (a, b) .

and when $y_{i-1} = 1$ then it behaves as a binary channel with the transition matrix

$$\begin{bmatrix} b & \bar{a} \\ \bar{b} & a \end{bmatrix}. \quad (74)$$

We refer to this channel as the POST(a, b) channel. POST(α) is a special case of POST(a, b), where $a = 1$ and $b = \bar{\alpha}$.

Without loss of generality, we assume throughout that $a + b - 1 > 0$. It is easy to see that in the case where $a + b - 1 = 0$ or, equivalently, where $a = \bar{b}$, the capacity is simply 0. Additionally, if $a + b - 1 < 0$ then $\bar{a} + \bar{b} > 1$; hence by relabeling the inputs ($0 \leftrightarrow 1$) we obtain a new channel (with parameter a', b' rather than a, b) where $a' = \bar{a}$ and $b' = \bar{b}$ and we have $a' + b' - 1 > 0$.

B. Capacity of the POST(a, b) Channel With and Without Feedback

Before considering the POST(a, b) let us first consider the binary DMC with parameters (a, b) . The capacity of the binary DMC with parameters (a, b) was derived by Ash in [22, Example 3.7] by applying [22, Th. 3.3.3] and is given by

$$C = \log \left[2^{\frac{\bar{a}H_b(b) - bH_b(a)}{a+b-1}} + 2^{\frac{\bar{b}H_b(a) - aH_b(b)}{a+b-1}} \right]. \quad (75)$$

The capacity achieving input distribution is

$$\begin{aligned} P(x = 0) &= c_0 \left(b 2^{\frac{H(b)}{a+b-1}} - \bar{b} 2^{\frac{H(a)}{a+b-1}} \right), \\ P(x = 1) &= c_0 \left(-\bar{a} 2^{\frac{H(b)}{a+b-1}} + a 2^{\frac{H(a)}{a+b-1}} \right), \end{aligned} \quad (76)$$

where c_0 is a normalizing coefficient so that the sum $P(x = 0) + P(x = 1)$ is equal to 1. The induced output distribution is

$$P(y = 0) = c_0(ab - \bar{a}\bar{b}) 2^{\frac{H(b)}{a+b-1}} \quad (77)$$

$$P(y = 1) = c_0(ab - \bar{a}\bar{b}) 2^{\frac{H(a)}{a+b-1}}. \quad (78)$$

Lemma 8 (Feedback Capacity of POST(a, b)): The feedback capacity of the POST(a, b) channel is the same as of the memoryless DMC with parameters (a, b) , which is given in (75).

The proof follows the same arguments as the proof of the feedback capacity of POST(α) in Lemma 4 and is, therefore, omitted. The behavior of the capacity as a function of (a, b) is depicted in Fig. 5.

We now present sufficient conditions on a, b implying that feedback does not increase the capacity of the POST(a, b) channel. That these conditions are indeed sufficient we establish in the next subsection. Define the following intervals:

$$\begin{aligned} \mathcal{L}_1 &= \left\{ \max\left(\frac{\bar{a}}{b}\gamma, \frac{\gamma(\bar{a}+b) - \sqrt{\gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b}}}{2\bar{b}}\right) \leq \beta \right. \\ &\quad \left. \leq \frac{\gamma(\bar{a}+b) + \sqrt{\gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b}}}{2\bar{b}} \right\} \\ \mathcal{L}_2 &= \left\{ \frac{(a+\bar{b}) + \sqrt{(a+\bar{b})^2 - 4\bar{a}b\gamma^2}}{2b\gamma} \leq \beta \leq \frac{\bar{a}}{b}\gamma \right\} \\ \mathcal{L}_3 &= \left\{ \beta \leq \min\left(\frac{\bar{a}}{b}\gamma, \frac{(a+\bar{b}) - \sqrt{(a+\bar{b})^2 - 4\bar{a}b\gamma^2}}{2b\gamma}\right) \right\} \\ \mathcal{L}_4 &= \left\{ \beta \leq \min\left(\frac{b\gamma}{a}, \frac{\gamma(\bar{a}+b) - \sqrt{\gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b}}}{2a}\right) \right\} \\ \mathcal{L}_5 &= \left\{ \frac{\gamma(\bar{a}+b) + \sqrt{\gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b}}}{2a} \leq \beta \leq \frac{b\gamma}{a} \right\} \\ \mathcal{L}_6 &= \left\{ \max\left(\frac{b\gamma}{a}, \frac{(a+\bar{b}) - \sqrt{(a+\bar{b})^2 - 4\bar{a}b\gamma^2}}{2\bar{a}\gamma}\right) \leq \beta \right. \\ &\quad \left. \leq \frac{(a+\bar{b}) + \sqrt{(a+\bar{b})^2 - 4\bar{a}b\gamma^2}}{2\bar{a}\gamma} \right\}, \end{aligned} \quad (79)$$

where γ is defined as

$$\gamma = 2^{\frac{H(b)-H(a)}{a+b-1}}. \quad (80)$$

In addition, let

$$\mathcal{L}_0 = \left\{ 1 \leq \beta \leq \min\left(\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\right) \right\} \quad (81)$$

Lemma 9 (When Doesn't Feedback Increase Capacity): If the intersections of the intervals $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \mathcal{L}_3$ with $\mathcal{L}_4 \cup \mathcal{L}_5 \cup \mathcal{L}_6$ and \mathcal{L}_0 is nonempty then feedback does not increase the capacity of the POST(a, b) channel.

Lemma 10: The condition in Lemma 9 holds for all POST channel parameters (a, b) . Thus, feedback does not increase capacity of POST(a, b).

Proof: It suffices to show that either the interval $(\mathcal{L}_1 \cap \mathcal{L}_5 \cap \mathcal{L}_0)$ or the interval $(\mathcal{L}_2 \cap \mathcal{L}_6 \cap \mathcal{L}_0)$ is non empty. First we claim that the expression in the square roots of (79) are nonnegative, i.e.,

$$\begin{aligned} \gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b} &\geq 0 \\ (a+\bar{b})^2 - 4\bar{a}b\gamma^2 &\geq 0 \end{aligned} \quad (82)$$

as shown in Appendix D-B (Note that the second inequality follows from the first by switching between a and b).

Recall that $a+b-1 \geq 0$ which implies $a \geq \bar{b}$ and $\bar{a} \leq b$. In addition, assume that $\bar{a}a \leq b\bar{b}$, and we show in Appendix D-A that together with $a \geq \bar{b}$ it implies that $a \geq b$. We now prove that if $\bar{a}a \leq b\bar{b}$, then $(\mathcal{L}_1 \cap \mathcal{L}_5 \cap \mathcal{L}_0)$ is nonempty. (Similarly, one can show that if $\bar{a}a \geq b\bar{b}$, $(\mathcal{L}_2 \cap \mathcal{L}_6 \cap \mathcal{L}_0)$ is non empty.) We first want to claim that the

intersection $(\mathcal{L}_1 \cap \mathcal{L}_5)$ is nonempty. Indeed, the lower bound of \mathcal{L}_5 is smaller than the upper bound of \mathcal{L}_1 since $\frac{1}{2a} \leq \frac{1}{2\bar{b}}$. The upper bound of \mathcal{L}_5 is larger than the lower bound of \mathcal{L}_1 since $\frac{b\gamma}{a} \geq \frac{\bar{a}\gamma}{\bar{b}}$ because $\bar{a}a \leq b\bar{b}$, and because,

$$\frac{\gamma(\bar{a}+b) - \sqrt{\gamma^2(\bar{a}+b)^2 - 4\bar{a}\bar{b}}}{2\bar{b}} \leq \frac{b\gamma}{a}, \quad (83)$$

where the inequality follows the fact that the LHS is less or equal to 1, as shown in Appendix D-C, and the RHS is great or equal to 1, as show in Appendix D-D.

Now we need to show that the intersection $(\mathcal{L}_1 \cap \mathcal{L}_5)$ with \mathcal{L}_0 is still non empty. We will first show that $\min\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\} \geq \frac{b\gamma}{a}$. Recall that $a \geq \bar{b}$ hence $\frac{b\gamma}{\bar{b}} \geq \frac{b\gamma}{a}$. In addition $\gamma^2 \leq \frac{a^2}{\bar{a}a}$ as shown in Appendix D-E. Now we want to show that the lower bound of \mathcal{L}_0 which is 1 is not larger than the upper bound of the intersection $(\mathcal{L}_1 \cap \mathcal{L}_5)$. First we claim that the upper bound of \mathcal{L}_5 is larger than 1, i.e., $\frac{b\gamma}{a} \geq 1$ for $a \geq \bar{b}$, as shown in Appendix D-D. Finally, we need to show that the upper bound of \mathcal{L}_1 is larger than 1, i.e.,

$$\frac{\gamma(\bar{a}+b)}{2\bar{b}} \geq 1, \quad (84)$$

as shown in Appendix D-F. ■

C. Deriving the Sufficient Conditions of Lemma 9

Proof of Lemma 9: Let $P_{n,0}$ and $P_{n,1}$ be defined as in (37). Following the channel definition we have

$$P_{n,0} = \begin{bmatrix} a \cdot P_{n-1,0} & \bar{b} \cdot P_{n-1,0} \\ \bar{a} \cdot P_{n-1,1} & b \cdot P_{n-1,1} \end{bmatrix} \quad (85)$$

and

$$P_{n,1} = \begin{bmatrix} b \cdot P_{n-1,0} & \bar{a} \cdot P_{n-1,0} \\ \bar{b} \cdot P_{n-1,1} & a \cdot P_{n-1,1} \end{bmatrix} \quad (86)$$

where $P_{0,0} = P_{0,1} = 1$. Using the identity

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{Z}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{Z}^{-1} \\ -\mathbf{Z}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{Z}^{-1} \end{bmatrix} \quad (87)$$

where $\mathbf{Z} \triangleq (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})$, we obtain

$$P_{n,0}^{-1} = \begin{bmatrix} \frac{b}{ba - \bar{a}\bar{b}} P_0^{-1} & -\frac{\bar{b}}{ba - \bar{a}\bar{b}} P_1^{-1} \\ -\frac{\bar{a}}{ba - \bar{a}\bar{b}} P_0^{-1} & \frac{a}{ba - \bar{a}\bar{b}} P_1^{-1} \end{bmatrix} \quad (88)$$

$$P_{n,1}^{-1} = \begin{bmatrix} \frac{a}{ba - \bar{a}\bar{b}} P_0^{-1} & -\frac{\bar{a}}{ba - \bar{a}\bar{b}} P_1^{-1} \\ -\frac{\bar{b}}{ba - \bar{a}\bar{b}} P_0^{-1} & \frac{b}{ba - \bar{a}\bar{b}} P_1^{-1} \end{bmatrix} \quad (89)$$

Now we compute $P_1(x^n)$ and $P_0(x^n)$

$$\begin{aligned} P_0(x^n) &= P_{n,0}^{-1} P_0(y^n) \\ &= \frac{1}{(a+b-1)(2^{\frac{H(b)}{a+b-1}} + 2^{\frac{H(a)}{a+b-1}})} \\ &\quad \times \begin{bmatrix} b P_0^{-1} & -\bar{b} P_1^{-1} \\ -\bar{a} P_0^{-1} & a P_1^{-1} \end{bmatrix} \begin{bmatrix} 2^{\frac{H(b)}{a+b-1}} P_0(y^{n-1}) \\ 2^{\frac{H(a)}{a+b-1}} P_1(y^{n-1}) \end{bmatrix} \end{aligned}$$

$$= \frac{1}{(a+b-1)(2^{\frac{H(b)}{a+b-1}} + 2^{\frac{H(a)}{a+b-1}})} \times \begin{bmatrix} b2^{\frac{H(b)}{a+b-1}} P_0(x^{n-1}) - \bar{b}2^{\frac{H(a)}{a+b-1}} P_1(x^{n-1}) \\ -\bar{a}2^{\frac{H(b)}{a+b-1}} P_0(x^{n-1}) + a2^{\frac{H(a)}{a+b-1}} P_1(x^{n-1}) \end{bmatrix}, \quad (90)$$

$$P_1(x^n) = P_{n,1}^{-1} P_1(y^n) = \frac{1}{(a+b-1)(2^{\frac{H(b)}{a+b-1}} + 2^{\frac{H(a)}{a+b-1}})} \times \begin{bmatrix} a2^{\frac{H(a)}{a+b-1}} P_0(x^{n-1}) - \bar{a}2^{\frac{H(b)}{a+b-1}} P_1(x^{n-1}) \\ -\bar{b}2^{\frac{H(a)}{a+b-1}} P_0(x^{n-1}) + b2^{\frac{H(b)}{a+b-1}} P_1(x^{n-1}) \end{bmatrix}, \quad (91)$$

where $P_0(x^0) = P_1(x^0) = 1$. We can rewrite $P_0(x^n)$ and $P_1(x^n)$ as follows:

$$P_0(x^n) = \frac{1}{(a+b-1)(\gamma+1)} \begin{bmatrix} b\gamma P_0(x^{n-1}) - \bar{b}P_1(x^{n-1}) \\ -\bar{a}\gamma P_0(x^{n-1}) + aP_1(x^{n-1}) \end{bmatrix}, \quad (92)$$

$$P_1(x^n) = \frac{1}{(a+b-1)(\gamma+1)} \begin{bmatrix} aP_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1}) \\ -\bar{b}P_0(x^{n-1}) + b\gamma P_1(x^{n-1}) \end{bmatrix}. \quad (93)$$

We need to show that indeed the probability expressions are valid, namely nonnegative and sum to 1. Showing the non-negativity of each of the terms in the above expression is equivalent to showing $\forall n \geq 1$ and for all x^{n-1} ,

$$\min\left\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\right\} P_0(x^{n-1}) \geq P_1(x^{n-1})$$

$$\min\left\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\right\} P_1(x^{n-1}) \geq P_0(x^{n-1}). \quad (94)$$

For $n = 1$ this follows from the fact that $\min\left\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\right\} \geq 1$ which is proved in Appendix D-G. To prove for $n \geq 1$ we use the following lemma, whose proof appears in Appendix C. ■

Lemma 11: If the condition in Lemma 9 holds then there exists, $1 \leq \beta \leq \min\left\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\right\}$ such that $\forall n$, the inequalities

$$\beta P_1(x^{n-1}) \geq P_0(x^{n-1}), \quad \forall x^{n-1},$$

$$\beta P_0(x^{n-1}) \geq P_1(x^{n-1}), \quad \forall x^{n-1}, \quad (95)$$

imply

$$\beta P_1(x^n) \geq P_0(x^n), \quad \forall x^n,$$

$$\beta P_0(x^n) \geq P_1(x^n), \quad \forall x^n. \quad (96)$$

VI. DOES THERE EXIST A POST CHANNEL FOR WHICH FEEDBACK INCREASES THE CAPACITY?

For the nonfeedback capacity we do not have an analytical expression, but only the infinite letter expression given in (8). In general, for any finite state channel we have the following upper bounds [20, Th. 4.6.1] [19, Th. 15]

$$C \leq \frac{1}{n} \max_{s_0} \max_{P(x^n)} I(X^n; Y^n | s_0) + \frac{\log |\mathcal{S}|}{n}, \quad (97)$$

for any integer $n \geq 1$. The notation, $I(X^n; Y^n | s_0)$ is conditional mutual information, which is defined in the same way as regular mutual information, but as a function of the joint $P(x^n, y^n | s_0)$ rather than as a function of $P(x^n, y^n)$.

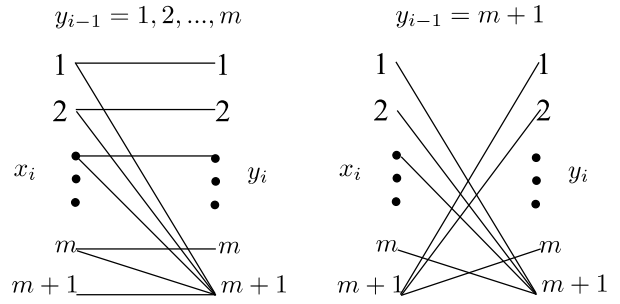


Fig. 6. A POST channel where feedback increases capacity when m is large. The probability associated with each edge is either $\frac{1}{2}$ or 1 when $y_{i-1} = 1, 2, \dots, m$ and either 1 or $\frac{1}{m}$ when $y_{i-1} = m+1$.

The notation $|\mathcal{S}|$ refers to the number of channel states. However, for the POST channel we have a tighter upper bound given in the following corollary which is obtained directly from [23, Th. 3.11].

Corollary 12 (Upper Bound on the Capacity): For any POST channel the following upper bound holds

$$C \leq \frac{1}{n} \max_{s_0} \max_{P(x^n)} I(X^n; Y^n | s_0) \quad (98)$$

for any integer $n \geq 1$.

Given the results that feedback does not increase the capacity of a $\text{POST}(a, b)$, the question naturally arises: does there exist a specific POST channel where feedback strictly increases the capacity? The answer is affirmative. The idea is to find a POST channel that consists of two states such that when there is feedback, the optimal input distributions given the states differ significantly among the different states. For the binary DMC it was shown in [24] and independently in [25] that the input probability that achieves the capacities is in $[\frac{1}{e}, 1 - \frac{1}{e}]$ for each of the two alphabet symbols. Hence, in the case of a binary post channel with feedback the optimal input probability as a function of the state will not vary too much. But as we increase the alphabet size we can construct a POST channel where the optimal input probabilities as a function of the state would vary significantly. Such a channel is presented in Fig. 6.

We can determine the feedback capacity analytically using the following lemma, which is proved in Appendix E.

Lemma 13: The feedback capacity of the channel in Fig. 6, is given by,

$$C_{fb} = \max_{\gamma, \delta \in [0, 1]} \left\{ \frac{2\delta}{2\delta + 1 + \gamma} \left(\frac{\bar{\gamma}}{2} \log_2(m) + h_2\left(\frac{1+\gamma}{2}\right) - (1-\gamma) \right) + \frac{1+\gamma}{2\delta + 1 + \gamma} \left(h_2(\delta) \right) \right\} \quad (99)$$

Corollary 14: Note that as m approaches infinity, $\frac{\bar{\gamma}}{2} \log_2(m) + h_2\left(\frac{1+\gamma}{2}\right) - (1-\gamma) \simeq \frac{\bar{\gamma}}{2} \log_2(m) \gg h_2(\delta)$, thus $C_{fb} \simeq \max_{\gamma \in [0, 1]} \left\{ \frac{1-\gamma}{3+\gamma} \log_2(m) \right\} = \frac{\log_2(m)}{3}$.

This gives us the intuition to suggest the following simple scheme approximately capacity-achieving for large m . If $y_{i-1} \leq m$, then transmit $\log_2(m)$ bits via input $X_i = 1, 2, \dots, m$. The probability that these bits would be received at the decoder is $\frac{1}{2}$. If $y_{i-1} = m+1$, then $X_i = m+1$.

TABLE III
CAPACITY WITH AND WITHOUT FEEDBACK OF THE POST CHANNEL IN FIG. 6 AS THE ALPHABET GROWS.
EVIDENTLY, FEEDBACK INCREASES CAPACITY FOR SOME VALUES OF m

m	upper bound on capacity $\frac{1}{6} \max_{s_0} \max_{P(x^6)} I(X^6; Y^6 s_0)$	lower bound on feedback capacity $R = \frac{\log_2 m}{3}$	feedback capacity (99)
2^0	0.7918	0	0.7595
2^1	0.8568	0.3333	0.8325
2^2	0.9803	0.6667	1.0000
2^3	1.1711	1.0000	1.2599
2^4	1.3865	1.3333	1.5366
2^5	1.6098	1.6667	1.8260
2^6	1.8374	2.0000	2.1252
2^7	2.0683	2.3333	2.4319
2^8	2.3019	2.6667	2.7444
2^9	2.5376	3.0000	3.06140
2^{10}	2.7751	3.3333	3.3818

Thus, the rate transmitted error free for this scheme is

$$R = \frac{\text{average bit transmitted}}{\text{average usage of channels}} = \frac{\frac{1}{2} \log_2 m + \frac{1}{2} 0}{\frac{1}{2} 1 + \frac{1}{2} 2} = \frac{\log_2 m}{3}. \quad (100)$$

It is possible to calculate the term $\max_{P(x^n)} I(X^n; Y^n | s_0)$ using the Blahut-Arimoto (BA) Algorithm [1], [2] and hence to obtain an upper.. The BA algorithm consists of iterations where each iteration has two steps in which $Q(x^n | y^n)$ and $P(x^n)$ are computed as follows

$$Q(x^n | y^n) = \frac{1}{c_1} P(x^n) P(y^n | x^n), \quad (101)$$

$$P(x^n) = \frac{1}{c_2} \prod_{y^n} Q(x^n | y^n) P(y^n | x^n), \quad (102)$$

where c_1 and c_2 are normalization factors that yields $\sum_{x^n} Q(x^n | y^n) = 1 \forall y^n$, and $\sum_{x^n} P(x^n) = 1$. The initialization of the algorithm is done by choosing $P(x^n)$ uniform over all x^n . In general, (101) and (102) need to be calculated for any x^n and y^n . However, we note that for the channel presented in Fig. 6 there is symmetry and hence $P(x^n)$ depends only on the condition if $x_i \in \{1, \dots, m\}$ (and not on the value itself) or if $x_i = m + 1$. Therefore, it suffices to calculate only 2^n values of $P(x^n)$. Furthermore, $Q(x^n | y^n) = 0$ if for some $i \in \{1, 2, \dots, n\}$, $x_i \neq y_i$ and $y_i \neq m + 1$. Therefore, it suffices to calculate only values where $x_i = y_i$ or $y_i = m + 1$. Finally, as in the case of $P(x^n)$, $Q(x^n | y^n)$ depends only on the condition if $x_i \in \{1, \dots, m\}$ or if $x_i = m + 1$. Hence, for $Q(x^n | y^n)$ it suffices to calculate only $2^n \times 2^n$ values.

The simple observations above make it feasible to apply the BA algorithm in order to calculate $\max_{P(x^n)} I(X^6; Y^6 | s_0)$ for the channel in Fig. 6 for large m , as presented in Table III. We clearly see from Table III that for $m \geq 2^2$ the feedback capacity is strictly larger than the non feedback capacity. The difference increases with m .

VII. CONCLUSION AND FURTHER RESEARCH

We have introduced and studied the family of POST channels and showed, somewhat surprisingly, that feedback does not increase the capacity of the general $POST(a, b)$ channel. The proof 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 also be induced by an input distribution without feedback. 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 implementable capacity-achieving codes.

APPENDIX A

CONCAVITY OF DIRECTED INFORMATION IN $P(x^n || y^{n-1})$

Lemma 15 (Concavity of Directed Information in $P(x^n || y^{n-1})$): Directed information $I(X^n \rightarrow Y^n)$ is concave in $P(x^n || y^{n-1})$ for a fixed $P(y^n || x^n)$.

Proof: We need to show that for $0 \leq \theta \leq 1$

$$\begin{aligned} & \sum_{x^n, y^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n) \\ & \times \log \frac{p(y^n || x^n)}{\sum_{x^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n)} \\ & \geq \sum_{x^n, y^n} \theta p_1(x^n || y^{n-1}) p(y^n || x^n) \\ & \times \log \frac{p(y^n || x^n)}{\sum_{x^n} p_1(x^n || y^{n-1}) p(y^n || x^n)} \\ & + \bar{\theta} p_2(x^n || y^{n-1}) p(y^n || x^n) \log \frac{p(y^n || x^n)}{\sum_{x^n} p_2(x^n || y^{n-1}) p(y^n || x^n)} \end{aligned} \quad (103)$$

This inequality may be written as

$$\begin{aligned} & \sum_{x^n, y^n} \theta p_1(x^n || y^{n-1}) p(y^n || x^n) \\ & \times \log \frac{\sum_{x^n} p_1(x^n || y^{n-1}) p(y^n || x^n)}{\sum_{x^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n)} \\ & + \bar{\theta} p_2(x^n || y^{n-1}) p(y^n || x^n) \\ & \times \log \frac{\sum_{x^n} p_2(x^n || y^{n-1}) p(y^n || x^n)}{\sum_{x^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n)} \geq 0 \end{aligned} \quad (104)$$

Furthermore,

$$\begin{aligned} & \sum_{y^n} \theta \left\{ \sum_{x^n} p_1(x^n || y^{n-1}) p(y^n || x^n) \right\} \\ & \times \log \frac{\sum_{x^n} p_1(x^n || y^{n-1}) p(y^n || x^n)}{\sum_{x^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n)} \\ & + \sum_{y^n} \bar{\theta} \left\{ \sum_{x^n} p_2(x^n || y^{n-1}) p(y^n || x^n) \right\} \\ & \times \log \frac{\sum_{x^n} p_2(x^n || y^{n-1}) p(y^n || x^n)}{\sum_{x^n} (\theta p_1(x^n || y^{n-1}) + \bar{\theta} p_2(x^n || y^{n-1})) p(y^n || x^n)} \geq 0 \end{aligned} \quad (105)$$

Finally, note that the RHS is a sum of two divergences between pmf's of y^n and therefore it is nonnegative. ■

APPENDIX B SUPPORTING INEQUALITIES

Lemma 16: The inequality

$$\alpha^{\frac{1}{\alpha}} \leq 1 \quad (106)$$

holds for $0 \leq \alpha \leq 1$.

Proof: For $0 \leq \alpha \leq 1$

$$\log \alpha \leq 0, \quad (107)$$

which implies

$$\frac{1}{\bar{\alpha}} \log \alpha \leq 0, \quad (108)$$

and equivalently

$$2^{\frac{1}{\bar{\alpha}} \log \alpha} \leq 2^0. \quad (109)$$

Note that the last inequality is actually (106) for nonnegative α . ■

Lemma 17: The inequality

$$4\alpha^{\frac{\alpha+1}{\alpha}} \leq 1 \quad (110)$$

holds for $0 \leq \alpha \leq 1$.

Proof: By taking \ln of both sides, we need to show

$$\frac{1+\alpha}{\bar{\alpha}} \ln \alpha + \ln 4 \leq 0, \quad (111)$$

which is equivalent to

$$(1+\alpha) \ln \alpha + \bar{\alpha} \ln 4 \leq 0. \quad (112)$$

In order to prove (112) we claim that the LHS increases in α for $0 \leq \alpha \leq 1$ and, therefore, the maximum value is obtained at $\alpha = 1$ and is 0. In order to show that the LHS of (112) is increasing we need to show that its derivative is nonnegative, i.e.,

$$\frac{1}{\alpha} + 1 + \ln \alpha - \ln 4 \geq 0, \quad (113)$$

or equivalently

$$1 + \alpha + \alpha \ln \alpha - \alpha \ln 4 \geq 0. \quad (114)$$

The LHS of (114) is a convex function and the minimum is obtained when the derivative is zero, i.e.,

$$1 + \ln \alpha + 1 - \ln 4 = 0, \quad (115)$$

which implies that $\ln \alpha = \ln \frac{4}{e^2}$. Hence, the minimum value of the LHS of (114) is $1 + \frac{4}{e^2} + \frac{4}{e^2} \ln \frac{4}{e^2} - \frac{4}{e^2} \ln 4 = 1 - \frac{4}{e^2}$, which is positive. Therefore (114) holds, which implies that (112) holds, which implies that (111) holds. ■

Lemma 18: The following inequality holds

$$\frac{1 + \sqrt{1 - 4\alpha^{\frac{\alpha+1}{\alpha}}}}{2\alpha^{\frac{\alpha}{\alpha}}} \geq 1 \quad (116)$$

for any $0 \leq \alpha \leq 1$.

Proof: Since α is nonnegative we need to show

$$\sqrt{1 - 4\alpha^{\frac{\alpha+1}{\alpha}}} \geq 2\alpha^{\frac{\alpha}{\alpha}} - 1. \quad (117)$$

This would be true if

$$1 - 4\alpha^{\frac{\alpha+1}{\alpha}} \geq 4\alpha^{\frac{2\alpha}{\alpha}} - 4\alpha^{\frac{\alpha}{\alpha}} + 1, \quad (118)$$

which can be simplified to

$$1 \geq \alpha^{\frac{\alpha}{\alpha}} + \alpha^{\frac{1}{\alpha}}, \quad (119)$$

which can be written as

$$1 \geq \alpha^{\frac{\alpha}{\alpha}} (1 + \alpha). \quad (120)$$

When $\alpha \rightarrow 0^+$ we have equality, hence it suffices to show that the equality holds after taking the derivative with respect to α . We use the equality $f'(\alpha) = f(\alpha)(\ln f(\alpha))'$ to find

$$(\alpha^{\frac{\alpha}{\alpha}})' = \alpha^{\frac{\alpha}{\alpha}} \left(\frac{\ln \alpha}{\bar{\alpha}^2} + \frac{1}{\bar{\alpha}} \right). \quad (121)$$

Hence, applying the derivative on (120) it remains to show that

$$\alpha^{\frac{\alpha}{\alpha}} \left(\frac{\ln \alpha}{\bar{\alpha}^2} + \frac{1}{\bar{\alpha}} \right) (1 + \alpha) + \alpha^{\frac{\alpha}{\alpha}} \leq 0. \quad (122)$$

or more simply

$$\ln \alpha + \bar{\alpha} + \frac{\bar{\alpha}^2}{1 + \alpha} \leq 0. \quad (123)$$

Note that if $\alpha = 1$ there is equality. Hence it suffices to show that the derivative of the LHS is non-negative for $0 \leq \alpha \leq 1$. I.e.,

$$\frac{1}{\alpha} - 1 - \frac{2\bar{\alpha}}{1 + \alpha} - \frac{\bar{\alpha}^2}{(1 + \alpha)^2} \geq 0 \quad (124)$$

which is equivalent to

$$\frac{\bar{\alpha}}{\alpha} - \frac{\bar{\alpha}(3+\alpha)}{(1+\alpha)^2} \geq 0 \quad (125)$$

and this is true if

$$(1+\alpha)^2 - \alpha(3+\alpha) \geq 0 \quad (126)$$

which is equivalent to

$$1 - \alpha \geq 0, \quad (127)$$

which is true. \blacksquare

APPENDIX C PROOF OF LEMMA 11

Proof: Suppose Eq. (95) holds; for Eq. (96) to hold, we need to have $\forall x^{n-1}$,

$$\beta(b\gamma P_0(x^{n-1}) - \bar{b}P_1(x^{n-1})) \geq aP_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1}) \quad (128)$$

$$\beta(-\bar{a}\gamma P_0(x^{n-1}) + aP_1(x^{n-1})) \geq -\bar{b}P_0(x^{n-1}) + b\gamma P_1(x^{n-1}) \quad (129)$$

$$\beta(aP_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1})) \geq b\gamma P_0(x^{n-1}) - \bar{b}P_1(x^{n-1}) \quad (130)$$

$$\beta(-\bar{b}P_0(x^{n-1}) + b\gamma P_1(x^{n-1})) \geq -\bar{a}\gamma P_0(x^{n-1}) + aP_1(x^{n-1}), \quad (131)$$

or equivalently, (128) and (129) become

$$(b\gamma\beta - a)P_0(x^{n-1}) \geq (\bar{b}\beta - \bar{a}\gamma)P_1(x^{n-1}), \quad (132)$$

$$(\bar{b} - \bar{a}\gamma\beta)P_0(x^{n-1}) \geq (b\gamma - a\beta)P_1(x^{n-1}), \quad (133)$$

and (130) and (131) become

$$(\bar{b} - \bar{a}\gamma\beta)P_1(x^{n-1}) \geq (b\gamma - a\beta)P_0(x^{n-1}) \quad (134)$$

$$(b\gamma\beta - a)P_1(x^{n-1}) \geq (\bar{b}\beta - \bar{a}\gamma)P_0(x^{n-1}). \quad (135)$$

Because of the similarity of the equations its enough to consider only (132) and (133). Now we will consider a few cases.

The Region of β That Satisfies (132) and (135): we will divide the treatment of (132) (or equivalently (135)) into two cases.

Case 1: $\beta\bar{b} - \bar{a}\gamma > 0$ or equivalently $\beta > \frac{\bar{a}\gamma}{\bar{b}}\gamma$; Eq. (132) becomes

$$P_0(x^{n-1}) \frac{b\gamma\beta - a}{\bar{b}\beta - \bar{a}\gamma} \geq P_1(x^{n-1}). \quad (136)$$

By the assumption of the induction this would be true for all x^{n-1} if

$$\frac{b\gamma\beta - a}{\bar{b}\beta - \bar{a}\gamma} \geq \beta, \quad (137)$$

or equivalently

$$\bar{b}\beta^2 - \gamma(\bar{a} + b)\beta + a \leq 0. \quad (138)$$

This implies

$$\begin{aligned} & \frac{\gamma(\bar{a} + b) - \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}}}{2\bar{b}} \\ & \leq \beta \leq \frac{\gamma(\bar{a} + b) + \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}}}{2\bar{b}}, \end{aligned} \quad (139)$$

which is the interval \mathcal{L}_1 defined in (79).

Case 2: $\beta\bar{b} - \bar{a}\gamma < 0$ or equivalently $\beta < \frac{\bar{a}\gamma}{\bar{b}}\gamma$; Eq. (132) becomes

$$P_0(x^{n-1}) \frac{b\gamma\beta - a}{\bar{b}\beta - \bar{a}\gamma} \leq P_1(x^{n-1}), \quad (140)$$

which is true based on the induction assumption if

$$\frac{b\gamma\beta - a}{\bar{b}\beta - \bar{a}\gamma} \leq \frac{1}{\beta}. \quad (141)$$

This is equivalent to

$$b\gamma\beta^2 - (a + \bar{b})\beta + \bar{a}\gamma \geq 0, \quad (142)$$

and this is true if

$$\beta \geq \frac{(a + \bar{b}) + \sqrt{(a + \bar{b})^2 - 4\bar{a}b\gamma^2}}{2b\gamma}, \quad (143)$$

which is the interval \mathcal{L}_2 , or

$$\beta \leq \frac{(a + \bar{b}) - \sqrt{(a + \bar{b})^2 - 4\bar{a}b\gamma^2}}{2b\gamma}, \quad (144)$$

which is the interval \mathcal{L}_3 .

The Region of β That Satisfies (133) and (134): we will divide the treatment of (133) (or equivalently (134)) into two cases.

Case 1: $b\gamma - \beta a > 0$ or equivalently $\beta < \frac{b\gamma}{a}$; Eq. (133) becomes

$$P_0(x^{n-1}) \frac{\bar{b} - \bar{a}\gamma\beta}{b\gamma - \beta a} \geq P_1(x^{n-1}). \quad (145)$$

By the assumption of the induction this would be true for all x^{n-1} if

$$\frac{\bar{b} - \bar{a}\gamma\beta}{b\gamma - \beta a} \geq \beta, \quad (146)$$

or equivalently

$$a\beta^2 - \gamma(b + \bar{a})\beta + \bar{b} \geq 0. \quad (147)$$

This implies

$$\beta \leq \frac{\gamma(\bar{a} + b) - \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}}}{2a}, \quad (148)$$

which is the interval \mathcal{L}_4 , or

$$\beta \geq \frac{\gamma(\bar{a} + b) + \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}}}{2a}, \quad (149)$$

which is the interval \mathcal{L}_5 .

Case 2: $b\gamma - \beta a < 0$ or equivalently $\beta > \frac{b\gamma}{a}$; Eq. (133) becomes

$$P_0(x^{n-1}) \frac{\bar{b} - \bar{a}\gamma\beta}{b\gamma - \beta a} \leq P_1(x^{n-1}). \quad (150)$$

By the assumption of the induction this would be true for all x^{n-1} if

$$\frac{\bar{b} - \bar{a}\gamma\beta}{b\gamma - \beta a} \leq \frac{1}{\beta} \quad (151)$$

$$\beta\bar{b} - \bar{a}\gamma\beta^2 \geq b\gamma - \beta a \quad (152)$$

$$\bar{a}\gamma\beta^2 - \beta(a + \bar{b}) + b\gamma \leq 0. \quad (153)$$

This implies

$$\frac{(a + \bar{b}) - \sqrt{(a + \bar{b})^2 - 4\bar{a}b\gamma^2}}{2\bar{a}\gamma} \leq \beta \leq \frac{(a + \bar{b}) + \sqrt{(a + \bar{b})^2 - 4\bar{a}b\gamma^2}}{2\bar{a}\gamma} \quad (154)$$

which is the interval \mathcal{L}_6 . ■

APPENDIX D

INEQUALITIES NEEDED FOR THE $POST(a, b)$ CHANNEL

In this appendix we prove inequalities that are needed for proving that feedback does not increase the capacity of $POST(a, b)$ channel. All inequalities contains γ which is defined in (80) and obviously $0 \leq a \leq 1$ and $0 \leq b \leq 1$.

A. $a \geq \bar{b}$ and $a\bar{a} \leq b\bar{b}$ Implies That $a \geq b$

Proof: Let $\rho_a = |\frac{1}{2} - a|$ and $\rho_b = |\frac{1}{2} - b|$. Hence,

$$\begin{aligned} a\bar{a} &= \left(\frac{1}{2} - \rho_a\right)\left(\frac{1}{2} + \rho_a\right) \\ b\bar{b} &= \left(\frac{1}{2} - \rho_b\right)\left(\frac{1}{2} + \rho_b\right) \end{aligned} \quad (155)$$

Since $a\bar{a} \leq b\bar{b}$ it follows that $\rho_a \geq \rho_b$. And since $a \geq \bar{b}$ it follows that $a = \frac{1}{2} + \rho_a$ and this implies $a \geq b$. ■

B. $\gamma^2(\bar{a} + b)^2 - 4a\bar{b} \geq 0$

Proof: Assume first that $a \geq \bar{b}$

$$2^{2\frac{H(\bar{b}) - H(a)}{a - \bar{b}}} \geq \frac{4a\bar{b}}{(2 - a - \bar{b})^2} \quad (156)$$

Taking log on both sides we obtain

$$2H(\bar{b}) - 2H(a) \geq (a - \bar{b})(\log 4\bar{b} + \log a - \log(2 - a - \bar{b})^2) \quad (157)$$

Note that if $a = \bar{b}$ we have equality, hence it suffices to prove that the inequality holds after applying the derivative with respect to a .

$$-2 \ln \frac{\bar{a}}{a} \geq \ln 4\bar{b} + \ln a - 2 \ln(2 - a - \bar{b}) + (a - \bar{b})\left(\frac{1}{a} + \frac{2}{2 - a - \bar{b}}\right) \quad (158)$$

If $a = \bar{b}$ we note that there is equality hence it suffices to prove that the inequality holds after taking the derivative with respect to a .

$$\begin{aligned} \frac{2}{a} + \frac{2}{1 - a} &\geq \frac{1}{a} + \frac{2}{2 - a - \bar{b}} + \frac{1}{a} + \frac{2}{2 - a - \bar{b}} \\ &+ (a - \bar{b})\left(-\frac{1}{a^2} + \frac{2}{(2 - a - \bar{b})^2}\right) \end{aligned} \quad (159)$$

After simplifying we obtain

$$\frac{a - \bar{b}}{a^2} + \frac{2}{\bar{a}} + \frac{2a + 6\bar{b} - 8}{(2 - a - \bar{b})^2} \geq 0 \quad (160)$$

which after simple algebra yields

$$\frac{a - \bar{b}}{a^2} + 2\frac{(a - \bar{b})(1 - \bar{b})}{\bar{a}(2 - a - \bar{b})^2} \geq 0, \quad (161)$$

and obviously the inequality holds since $a \geq \bar{b}$ and $b \leq 1$.

Now, if we consider $a \leq b$, then the sign \geq in (157) should be replaced by the sign \leq . In (158) the sign should be the same, but in (161) the sign should be the opposite again, which is true. ■

C. $\gamma(\bar{a} + b) - \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}} \leq 2\bar{b}$, for $a \geq \bar{b}$ and $a\bar{a} \leq b\bar{b}$

Proof: We need to show

$$\gamma(\bar{a} + b) - 2\bar{b} \leq \sqrt{\gamma^2(\bar{a} + b)^2 - 4a\bar{b}}, \quad (162)$$

since the RHS is nonnegative it suffices to show that

$$\gamma^2(\bar{a} + b)^2 - 4\bar{b}(\bar{a} + b)\gamma + 4\bar{b}^2 \leq \gamma^2(\bar{a} + b)^2 - 4a\bar{b}, \quad (163)$$

which simplifies to

$$\gamma(\bar{a} + b) \geq a + \bar{b}. \quad (164)$$

After applying log on both sides, we need to show that

$$H(b) - H(a) \geq (a - \bar{b})(\log(a + \bar{b}) - \log(\bar{a} + b)). \quad (165)$$

If $a = \bar{b}$ we have equality, hence it suffices to show that after we take derivative with respect to a the inequality holds, i.e.,

$$\ln \frac{a}{\bar{a}} \geq \log \frac{a + \bar{b}}{\bar{a} + b} + \frac{a - \bar{b}}{a + \bar{b}} + \frac{a - \bar{b}}{\bar{a} + b} \quad (166)$$

Again, if $b = \bar{a}$ we have equality, and we take the derivative with respect to b , i.e.,

$$0 \leq -\frac{1}{a + \bar{b}} - \frac{1}{\bar{a} + b} + \frac{2a}{(a + \bar{b})^2} + \frac{2\bar{a}}{(\bar{a} + b)^2} \quad (167)$$

which after simplification equals to

$$0 \leq \frac{a - \bar{b}}{(a + \bar{b})^2} + \frac{\bar{a} - b}{(\bar{a} + b)^2} \quad (168)$$

and since $a - \bar{b} = b - \bar{a}$, we only need to show that

$$\frac{1}{(a + \bar{b})^2} \leq \frac{1}{(\bar{a} + b)^2} \quad (169)$$

which is true since $a + 1 - b \geq 1 - a + b$ when $a \geq b$ and this follows as shown in Appendix D-A. ■

D. $\frac{by}{a} \geq 1$ for $a \geq \bar{b}$

Proof: We need to show that

$$-b \log b - \bar{b} \log \bar{b} + a \log a + \bar{a} \log \bar{a} \geq (a - \bar{b})(\log a - \log b) \quad (170)$$

If $a = \bar{b}$, then both sides equal to zero. Hence, it suffices to show that the inequality holds after applying the derivative with respect to a .

$$\ln \frac{a}{\bar{a}} \geq (\ln a - \log b) + 1 - \frac{\bar{b}}{a}. \quad (171)$$

We need to show that for $b \geq \bar{a}$

$$\log \frac{b}{\bar{a}} + \frac{\bar{b}}{a} - 1 \geq 0 \quad (172)$$

If $b = \bar{a}$, then equality holds, hence its enough to show that the derivative with respect to b is nonnegative. Namely,

$$\frac{1}{b} - \frac{1}{a} \geq 0. \quad (173)$$

This is true since we also have $a\bar{a} \leq b\bar{b}$ and it implies $a \geq b$ as shown in Appendix D-A. ■

E. $\gamma^2 \leq \frac{a^2}{b\bar{a}}$ for $a \geq \bar{b}$

Proof: Taking log on both sides we need to show

$$\frac{2h(b) - 2h(a)}{a - \bar{b}} \leq 2 \log a - \log b - \log \bar{a}. \quad (174)$$

After simple algebra we obtain

$$(-b - \bar{a}) \log b - 2\bar{b} \log \bar{b} + 2\bar{b} \log a + (\bar{a} + b) \log \bar{a} \leq 0 \quad (175)$$

We need to show that the inequality holds for $a \geq \bar{b}$. Note that when $a = \bar{b}$. we obtain equality. Therefore it's enough to show that the derivative of LHS with respect to a is nonpositive and therefore decreasing from 0. At this point we transform all the log to be natural base.

$$\ln b + 2\frac{\bar{b}}{a} - \ln \bar{a} - 1 - \frac{b}{\bar{a}} \leq 0 \quad (176)$$

Again if $b = \bar{a}$ we obtain equality. Now we take derivative with respect to b and need to show that it is nonpositive.

$$\frac{1}{b} - \frac{2}{a} - \frac{1}{\bar{a}} \leq 0$$

This is true since $b \geq \bar{a}$. ■

F. $\frac{\gamma(\bar{a}+b)}{2b} \geq 1$ for $a \geq \bar{b}$ and $a\bar{a} \leq b\bar{b}$

Proof: We would like to show

$$\frac{\gamma(\bar{a} + b)}{2\bar{b}} \geq 1. \quad (177)$$

Equivalently after taking log on both sides,

$$\frac{h(b) - h(a)}{a - \bar{b}} \geq \log \frac{2\bar{b}}{\bar{a} + b}, \quad (178)$$

$$\begin{aligned} & -b \log b - \bar{b} \log \bar{b} + a \log a + \bar{a} \log \bar{a} \\ & \geq (a - \bar{b})(\log 2 + \log \bar{b} - \log(\bar{a} + b)), \end{aligned} \quad (179)$$

Note that if $a = \bar{b}$ we have equality, hence it suffices to show that for $b \geq \bar{a}$ the inequality holds after taking the derivative with respect to b , i.e.,

$$\ln \frac{\bar{b}}{b} \geq \ln 2 + \ln \bar{b} - \ln(\bar{a} + b) + (a - \bar{b}) \left(-\frac{1}{\bar{b}} - \frac{1}{\bar{a} + b} \right) \quad (180)$$

If $a = \bar{b}$ we get 0 on both sides. Hence, it suffices to show that in we take derivative with respect to a (180) holds.

$$0 \geq \frac{1}{\bar{a} + b} - \frac{1}{\bar{b}} - \frac{1}{\bar{a} + b} - \frac{a - \bar{b}}{(\bar{a} + b)^2} \quad (181)$$

and this trivially holds since the RHS is nonpositive while the LHS is 0. ■

G. $\gamma \geq \frac{\bar{b}}{b}$ and $\gamma \leq \frac{a}{\bar{a}}$

Proof: Recall $\gamma = 2^{\frac{H(b)-H(a)}{a+b-1}}$. We prove only $\gamma \geq \frac{\bar{b}}{b}$ and $\gamma \leq \frac{a}{\bar{a}}$ follows from identical steps just replacing a with b . We need to show that

$$b2^{\frac{H(b)}{a+b-1}} \geq \bar{b}2^{\frac{H(a)}{a+b-1}}. \quad (182)$$

Equivalently,

$$\log b + \frac{H(b)}{a+b-1} \geq \log \bar{b} + \frac{H(a)}{a+b-1}. \quad (183)$$

$$\begin{aligned} (a+b-1) \log b - b \log b - \bar{b} \log \bar{b} \\ \geq (a+b-1) \log \bar{b} - a \log a - \bar{a} \log \bar{a}. \\ -\bar{a} \log b \geq a \log \bar{b} - a \log a - \bar{a} \log \bar{a}. \\ \bar{a} \log \frac{\bar{a}}{b} + a \log \frac{a}{\bar{b}} \geq 0. \end{aligned} \quad (184)$$

Note that the last equality is a divergence expression between two Bernoulli distributions with parameters a and \bar{b} and hence it's nonnegative. ■

APPENDIX E PROOF OF LEMMA 13

Proof: By symmetry, for any optimal distribution we will have, $P(X_i = m + 1 | Y_{i-1} = k)$ equal for all $k \in \{1, \dots, m\}$. Hence define

$$\gamma \triangleq P(X_i = m + 1 | Y_{i-1} = k) \quad \forall k \in \{1, \dots, m\} \quad (185)$$

$$\delta \triangleq P(X_i = m + 1 | Y_{i-1} = m + 1). \quad (186)$$

Also, by symmetry, $P(X_i = l | Y_{i-1} = k) = \frac{\bar{\gamma}}{m} \quad \forall k, l \in \{1, \dots, m\}$ and $P(X_i = l | Y_{i-1} = m + 1) = \frac{\bar{\delta}}{m} \quad \forall l \in \{1, \dots, m\}$. Thus we have the following transition kernel for the induced output Markov Chain,

$$\begin{aligned} P(Y_i = m + 1 | Y_{i-1} = m + 1) &= 1 - \delta \\ P(Y_i = m + 1 | Y_{i-1} = k) &= \text{frac}1 + \gamma 2 \quad \forall k \in \{1, 2, \dots, m\} \\ P(Y_i = k | Y_{i-1} = m + 1) &= \frac{\delta}{m} \quad \forall k \in \{1, \dots, m\} \\ P(Y_i = k | Y_{i-1} = l) &= \frac{\bar{\gamma}}{2m} \quad \forall k, l \in \{1, 2, \dots, m\}. \end{aligned} \quad (187)$$

Define the stationary distribution by $\pi_k \triangleq P(Y_i = k)$, for all $k \in \{1, \dots, m + 1\}$. To obtain the stationary distribution, we have the following balance equations,

$$\sum_{k=1}^{m+1} \pi_k = 1$$

$$\pi_{m+1} \delta = \left(\sum_{k=1}^m \pi_k \right) \frac{1 + \gamma}{2}$$

$$\pi_k \left(1 - \frac{\bar{\gamma}}{2m} \right) = \pi_{m+1} \delta + \left(\sum_{l=1, l \neq k}^m \pi_l \right) \frac{\bar{\gamma}}{2m} \quad \forall k \in \{1, \dots, m\}.$$

We solve for the **Case: m = 1**, first, for which the equations are,

$$\pi_1 + \pi_2 = 1 \quad (188)$$

$$\pi_2 \delta = \pi_1 \frac{1 + \gamma}{2}, \quad (189)$$

which yields, $\pi_1 = \frac{\delta}{\delta + \frac{1+\gamma}{2}}$ and $\pi_2 = \frac{\frac{1+\gamma}{2}}{\delta + \frac{1+\gamma}{2}}$. Now note,

$$\begin{aligned} I(X_i; Y_i | Y_{i-1}) &= \pi_1 \left(H(Y_i | Y_{i-1} = 1) - H(Y_i | X_i, Y_{i-1} = 1) \right) \\ &\quad + \pi_2 \left(H(Y_i | Y_{i-1} = 2) - H(Y_i | X_i, Y_{i-1} = 2) \right) \\ &= \pi_1 \left(h_2 \left(\frac{1+\gamma}{2} \right) - (1-\gamma) \right) + \pi_2 h_2(\delta). \end{aligned} \quad (190)$$

Now, as $C_{fb} = \max_{\gamma, \delta \in [0,1]} I(X_i; Y_i | Y_{i-1})$, we obtain

$$C_{fb} = \max_{\delta, \gamma \in [0,1]} \left\{ \frac{\delta}{\delta + \frac{1+\gamma}{2}} \left(h_2 \left(\frac{1+\gamma}{2} \right) - (1-\gamma) \right) + \frac{\frac{1+\gamma}{2}}{\delta + \frac{1+\gamma}{2}} \left(h_2(\delta) \right) \right\} \quad (191)$$

$$C_{fb} = \max_{\delta, \gamma \in [0,1]} \left\{ \frac{2\delta}{2\delta + 1 + \gamma} \left(h_2 \left(\frac{1+\gamma}{2} \right) - (1-\gamma) \right) + \frac{1+\gamma}{2\delta + 1 + \gamma} \left(h_2(\delta) \right) \right\}. \quad (192)$$

Now, we deal with general **Case**: $m \geq 2$, where from the symmetry in balance equations, π_k are equal $\forall k \in \{1, \dots, m\}$, and hence we obtain the following on solving the balance equations,

$$\pi_1 = \frac{\delta}{m\delta + m\frac{1+\gamma}{2}} \quad (193)$$

$$\pi_k = \frac{m\frac{1+\gamma}{2}}{m\delta + m\frac{1+\gamma}{2}} \quad \forall k \in \{1, \dots, m\}, \quad (194)$$

where constant $A = \frac{\delta(1-\gamma)}{2m\delta + m(1+\gamma)}$. Note that

$$\begin{aligned} I(X_i; Y_i | Y_{i-1} = k) &= h_2 \left(\underbrace{\frac{\bar{\gamma}}{2m}, \dots, \frac{\bar{\gamma}}{2m}}_{m \text{ times}}, \frac{1+\gamma}{2} \right) - (1-\gamma) \quad \forall k \in \{1, \dots, m\}. \end{aligned}$$

Also note that

$$\begin{aligned} I(X_i; Y_i | Y_{i-1} = m+1) &= h_2 \left(\underbrace{\frac{\delta}{m}, \dots, \frac{\delta}{m}}_{m \text{ times}}, \bar{\delta} \right) - \delta \log_2(m) = h_2(\delta). \end{aligned}$$

Thus we obtain

$$\begin{aligned} I(X_i; Y_i | Y_{i-1}) &= \left(\sum_{k=1}^m \pi_m \right) \left(h_2 \left(\underbrace{\frac{\bar{\gamma}}{2m}, \dots, \frac{\bar{\gamma}}{2m}}_{m \text{ times}}, \frac{1+\gamma}{2} \right) - (1-\gamma) \right) + \pi_{m+1} h_2(\delta) \end{aligned} \quad (195)$$

$$\begin{aligned} &= \frac{2\delta}{2\delta + 1 + \gamma} \left(h_2 \left(\underbrace{\frac{\bar{\gamma}}{2m}, \dots, \frac{\bar{\gamma}}{2m}}_{m \text{ times}}, \frac{1+\gamma}{2} \right) - (1-\gamma) \right) \\ &\quad + \frac{1+\gamma}{2\delta + 1 + \gamma} \left(h_2(\delta) \right). \end{aligned} \quad (196)$$

Now, as $C_{fb} = \max_{\gamma, \delta \in [0,1]} I(X_i; Y_i | Y_{i-1})$, after some basic algebraic manipulation we obtain the following expression:

$$C_{fb} = \max_{\delta, \gamma \in [0,1]} \left\{ \frac{2\delta}{2\delta + 1 + \gamma} \left(\frac{\bar{\gamma}}{2} \log_2(m) + h_2 \left(\frac{1+\gamma}{2} \right) - (1-\gamma) \right) + \frac{1+\gamma}{2\delta + 1 + \gamma} \left(h_2(\delta) \right) \right\}. \quad (197)$$

ACKNOWLEDGEMENT

The authors are grateful to Jiantao Jiao who suggested the proof of (82). The authors would like to thank the associate editor and the anonymous reviewers for very helpful comments.

REFERENCES

- [1] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.
- [2] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.
- [3] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949.
- [4] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Hawkesbury Showground, Australia: Holden-Day, 1964.
- [5] F. Alajaji and T. Fuja, "Effect of feedback on the capacity of discrete additive channels with memory," in *Proc. IEEE ISIT*, Jun./Jul. 1994.
- [6] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 57–85, Jan. 2010.
- [7] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2009.
- [8] O. Elishco and H. Permuter, "Capacity and coding for the Ising channel with feedback," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5138–5149, Sep. 2014.
- [9] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [10] 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.
- [11] T. Berger and F. Bonomi, "Capacity and zero-error capacity of Ising channels," *IEEE Trans. Inf. Theory*, vol. 36, no. 1, pp. 173–180, Jan. 1990.
- [12] H. Asnani, H. H. Permuter, and T. Weissman, "To feed or not to feed back," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5150–5172, Sep. 2014.
- [13] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.
- [14] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, ETH Zurich Univ., Zürich, Switzerland, 1998.
- [15] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Nov. 1990, pp. 303–305.
- [16] H. Marko, "The bidirectional communication theory—A generalization of information theory," *IEEE Trans. Commun.*, vol. COM-21, no. 12, pp. 1335–1351, Dec. 1973.
- [17] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1488–1499, Apr. 2008.
- [18] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.
- [19] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.
- [20] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

- [21] I. Naiss and H. H. Permuter, "Extension of the Blahut–Arimoto algorithm for maximizing directed information," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 204–222, Jan. 2013.
- [22] R. Ash, *Information Theory*. New York, NY, USA: Wiley, 1965.
- [23] H. Permuter J. Chen and T. Weissman, "Tighter bounds on the capacity of finite-state channels via Markov set-chains," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3660–3691, Aug. 2010.
- [24] N. Shulman and M. Feder, "The uniform distribution as a universal prior," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1356–1362, Jun. 2004.
- [25] G. Kumar and A. Manolacos, "No input symbol should occur more frequently than $1-1/e$," Jan. 2012. [Online]. Available: <http://arxiv.org/abs/1201.6425>

Haim Henri Permuter (M'08–SM'14) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008. Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor. Prof. Permuter is a recipient of several awards, among them the Fulbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY.

Himanshu Asnani is currently System Architect in Ericsson R&D, San Jose, CA (USA). He received his Ph.D. in Electrical Engineering Department in 2014 from Stanford University, working under Professor Tsachy Weissman, where he was a Stanford Graduate Fellow. His research interests include information theory, statistical learning and computer networks. He was the recipient of 2014 Marconi Society Paul Baran Young Scholar Award, Best Paper Award at MobiHoc 2009 and was also the finalist for Student Paper Award in ISIT 2011, Saint Petersburg, Russia. Prior to that, he received his B.Tech. from IIT Bombay in 2009 and M.S. from Stanford University in 2011, both in Electrical Engineering.

Tsachy Weissman (S'99–M'02–SM'07–F'13) is on the faculty of the department of Electrical Engineering at Stanford University, where he holds the STMicroelectronics Chair in the School of Engineering. He received his BSc and PhD from Technion in 1997 and 2001. He has published extensively on Information Theory, Statistical Signal Processing, the interplay between them, and their applications. He is inventor of several patents and involved in a number of hi-tech companies as member of the technical board. Much of his recent research has been dedicated to the theory and practice of genomic data compression. His research has been recognized with numerous awards, including best paper awards, a fellowships for Leaders in Science and Technology, and prizes for excellence in research. He serves on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY and *Foundations and Trends in Communications and Information Theory*.