Capacity of Binary Symmetric POST Channels

Conference Paper

Author(s):
Permuter, Haim H.; Asnani, Himanshu; Weissman, Tsachy

Publication date:
2014

Permanent link:
https://doi.org/10.3929/ethz-a-010094656

Capacity of Binary Symmetric POST Channels

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

Himanshu Asnani
Stanford University
asnani@stanford.edu

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 these as POST (Previous Output is the STate) channels. We focus on POST(a, b) channels. These channels have binary inputs and outputs, where the state determines if the channel behaves as a binary with parameters (a, b) or (b, a). We show that the non feedback capacity of the POST(a, b) 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 of the feedback. We show that this is a sufficient condition for the feedback capacity to equal the non feedback capacity for any finite state channel.**

Keywords*—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 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 in [10].

In this paper we introduce and consider 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: Consider the POST channel depicted in Fig. 1 with the following behavior. When $y_{i-1} = 0$, then the channel behaves as a binary channel with transition matrix

$$\begin{bmatrix} a & \bar{b} \\ \bar{a} & b \end{bmatrix} \tag{1}$$

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

$$\begin{bmatrix} b & \bar{a} \\ \bar{b} & a \end{bmatrix}. \tag{2}$$

We refer to this channel as the POST(a, b) channel. The

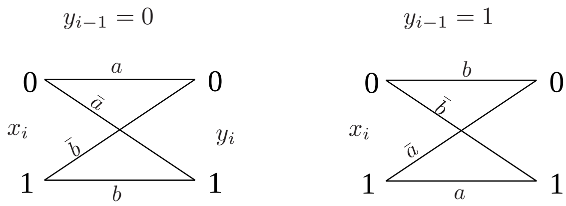

Fig. 1. POST(a, b) channel. If $y_{i-1} = 0$ then the channel behaves as DMC with parameters (a, b) and if $y_{i-1} = 1$ then the channel behaves as DMC with parameters (b, a).

POST(α) which was considered in [11] is a special case of POST(a, b), where $a = 1$ and $b = \bar{\alpha}$. The results in this paper extends our previouse results in [11]. An extended version of this conference paper that includes all the proofs may be found in [12].

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 $a' + b' - 1 > 0$.

This channel arose in the investigation of controlled feedback in the setting of "to feed or not to feed back" [13]. 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 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 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 $P^*(x^n||y^{n-1})$ are the same. This necessary and sufficient condition that we establish, in the generality of any finite state channel, follows from the KKT conditions [14, 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 consider a binary POST(a, b) channel and we show that feedback does not increase capacity for this considerably larger class of channels. In Section V, 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 [15]. We denote by $p(x^n||y^{n-d})$ the probability mass function of $X^n = (X_1, \ldots, X_n)$, *causally conditioned* on Y^{n-d} for some integer $d \geq 0$, which is defined as

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

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 $d = 0, 1$:

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

$$p(x^n||y^{n-1}) := \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-1}). \tag{5}$$

The directed information was defined by Massey [16], inspired by Marko's work [17] on bidirectional communication, as

$$I(X^n \to Y^n) := \sum_{i=1}^{n} I(X^i; Y_i|Y^{i-1}). \tag{6}$$

The directed information can also be rewritten as

$$I(X^n \to Y^n) = \sum_{x^n, y^n} p(x^n||y^{n-1})p(y^n||x^n) \log \frac{p(y^n||x^n)}{\sum_{x^n} p(x^n||y^{n-1})p(y^n||x^n)} \tag{7}$$

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

$$p(x^n, y^n) = p(x^n||y^{n-1})p(y^n||x^n). \tag{8}$$

We will make use the fact that directed information $I(X^n \to Y^n)$ is concave in $P(x^n||y^{n-1})$ for a fixed $P(y^n||x^n)$.

Directed information characterizes the capacity of point-to-point channels with feedback [10], [18]–[20]. 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

$$C_{fb} = \lim_{n\to\infty} \frac{1}{n} \max_{P(x^n||y^{n-1})} I(X^n \to Y^n). \tag{9}$$

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

$$C = \lim_{n\to\infty} \frac{1}{n} \max_{P(x^n)} I(X^n \to Y^n), \tag{10}$$

since the channel is indecomposable [21]. 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 [16].

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:

$$\max_{P(x^n||y^{n-1})} I(X^n \to Y^n) \tag{11}$$

and

$$\max_{P(x^n)} I(X^n \to Y^n). \tag{12}$$

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

A convex optimization problem, as defined in [14, Ch. 4], is a problem of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq b_i \quad i = 1, \cdots, k \\ & g_j(x) = 0 \quad j = 1, \cdots, l \end{array} \tag{13}$$

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 (11) into a convex optimization problem, as presented in (13), 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:

$$\begin{array}{ll} p(x^n||y^{n-1}) \geq 0, & \forall x^n, y^{n-1}, \\ \sum_{x_{i+1}^n} p(x^n||y^{n-1}) = \gamma_{x^i, y^{i-1}}, & \forall x^i, y^{n-1}, i \geq 1, \\ \sum_{x_1^n} p(x^n||y^{n-1}) = 1, & \forall y^{n-1}. \end{array} \tag{14}$$

Note that the two equalities in (14) 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 $\sum_x P(x) = 1$.

Note that the optimization problem given in (11) 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 $P(y^n||x^n)$ [22, Lemma 2]. Therefore, the KKT conditions [14, Ch 5.5.3] are necessary and sufficient. The next theorem states these conditions explicitly for our setting.

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 (10) is that for some numbers $\beta_{y^{n-1}}$

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

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

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

IV. 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 [23, Ex 3.7] by applying [23, Theorem 3.3.3] and is given by

$$C=\log\left[2^{\frac{\bar a H_b(b)-bH_b(a)}{a+b-1}}+2^{\frac{\bar b H_b(a)-aH_b(b)}{a+b-1}}\right]. \tag{17}$$

The capacity achieving input distribution is

$$\begin{aligned} P(x=0)&=c_0\left(b2^{\frac{H(b)}{a+b-1}}-\bar b 2^{\frac{H(a)}{a+b-1}}\right),\\ P(x=1)&=c_0\left(-\bar a 2^{\frac{H(b)}{a+b-1}}+a2^{\frac{H(a)}{a+b-1}}\right), \end{aligned} \tag{18}$$

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

$$P(y=0)=c_0(ab-\bar a\bar b)2^{\frac{H(b)}{a+b-1}} \tag{19}$$

$$P(y=1)=c_0(ab-\bar a\bar b)2^{\frac{H(a)}{a+b-1}}. \tag{20}$$

Lemma 3 (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 (17).

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:

$$\begin{aligned}
\mathcal{L}_1&=\Bigg\{\max(\frac{\bar a}{\bar b}\gamma,\frac{\gamma(\bar a+b)-\sqrt{\gamma^2(\bar a+b)^2-4a\bar b}}{2\bar b})\\
&\qquad\le\beta\le\frac{\gamma(\bar a+b)+\sqrt{\gamma^2(\bar a+b)^2-4a\bar b}}{2\bar b}.\Bigg\}\\
\mathcal{L}_2&=\left\{\frac{(a+\bar b)+\sqrt{(a+\bar b)^2-4\bar a b\gamma^2}}{2b\gamma}\le\beta\le\frac{\bar a}{\bar b}\gamma\right\}\\
\mathcal{L}_3&=\left\{\beta\le\min(\frac{\bar a}{\bar b}\gamma,\frac{(a+\bar b)-\sqrt{(a+\bar b)^2-4\bar a b\gamma^2}}{2b\gamma})\right\}\\
\mathcal{L}_4&=\left\{\beta\le\min(\frac{b\gamma}{a},\frac{\gamma(\bar a+b)-\sqrt{\gamma^2(\bar a+b)^2-4a\bar b}}{2a})\right\}\\
\mathcal{L}_5&=\left\{\frac{\gamma(\bar a+b)+\sqrt{\gamma^2(\bar a+b)^2-4a\bar b}}{2a}\le\beta\le\frac{b\gamma}{a}\right\}\\
\mathcal{L}_6&=\Bigg\{\max(\frac{b\gamma}{a},\frac{(a+\bar b)-\sqrt{(a+\bar b)^2-4\bar a b\gamma^2}}{2\bar a\gamma})\\
&\qquad\le\beta\le\frac{(a+\bar b)+\sqrt{(a+\bar b)^2-4\bar a b\gamma^2}}{2\bar a\gamma}\Bigg\},
\end{aligned} \tag{21}$$

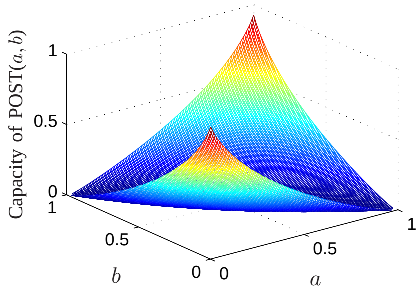

Fig. 2. The capacity of the POST(a,b) channel with and without feedback. This is also the capacity of the binary DMC with parameters (a,b)

where γ is defined as

$$\gamma=2^{\frac{H(b)-H(a)}{a+b-1}}. \tag{22}$$

In addition, let

$$\mathcal{L}_0=\left\{1\le\beta\le\min(\frac{a}{\bar a\gamma},\frac{b\gamma}{\bar b})\right\} \tag{23}$$

Lemma 4: 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 5: The condition in Lemma 4 holds for all POST channel parameters (a,b). Thus, feedback does not increase capacity of POST(a,b).

A. Deriving the sufficient conditions of Lemma 4

Proof of Lemma 4: Let $P_{n,0}$ and $P_{n,1}$ be defined as

$$P_{n,0}=\begin{bmatrix} a\cdot P_{n-1,0} & \bar b\cdot P_{n-1,0}\\ \bar a\cdot P_{n-1,1} & b\cdot P_{n-1,1}\end{bmatrix} \tag{24}$$

and

$$P_{n,1}=\begin{bmatrix} b\cdot P_{n-1,0} & \bar a\cdot P_{n-1,0}\\ \bar b\cdot P_{n-1,1} & a\cdot P_{n-1,1}\end{bmatrix} \tag{25}$$

where $P_{0,0} = P_{0,1} = 1$. Inverting the matrices, we obtain

$$P_{n,0}^{-1} = \begin{bmatrix} \frac{b}{ba-\bar{a}\bar{b}}P_0^{-1} & -\frac{\bar{b}}{ba-\bar{a}\bar{b}}P_1^{-1} \\ -\frac{\bar{a}}{ba-\bar{a}\bar{b}}P_0^{-1} & \frac{a}{ba-\bar{a}\bar{a}}P_1^{-1} \end{bmatrix} \tag{26}$$

$$P_{n,1}^{-1} = \begin{bmatrix} \frac{a}{ba-\bar{a}\bar{b}}P_0^{-1} & -\frac{\bar{a}}{ba-\bar{a}\bar{b}}P_1^{-1} \\ -\frac{\bar{b}}{ba-\bar{a}\bar{b}}P_0^{-1} & \frac{b}{ba-\bar{a}\bar{b}}P_1^{-1} \end{bmatrix} \tag{27}$$

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

$$P_0(x^n) = P_{n,0}^{-1}P_0(y^n) \tag{28}$$

$$P_1(x^n) = P_{n,1}^{-1}P_1(y^n) \tag{29}$$

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

$$P_0(x^n) = \frac{1}{(a+b-1)(\gamma+1)} \begin{bmatrix} b\gamma P_0(x^{n-1}) - \bar{b}P_1(x^{n-1}) \\ -\bar{a}\gamma P_0(x^{n-1}) + aP_1(x^{n-1}) \end{bmatrix} \tag{30}$$

$$P_1(x^n) = \frac{1}{(a+b-1)(\gamma+1)} \begin{bmatrix} aP_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1}) \\ -\bar{b}P_0(x^{n-1}) + b\gamma P_1(x^{n-1}) \end{bmatrix} \tag{31}$$

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},

$$\begin{aligned} \min\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\}P_0(x^{n-1}) &\geq P_1(x^{n-1}) \\ \min\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\}P_1(x^{n-1}) &\geq P_0(x^{n-1}). \end{aligned} \tag{32}$$

For $n = 1$ this follows from the fact that $\min\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\} \geq 1$. To prove for $n \geq 1$ we use the following lemma. ∎

Lemma 6: If the condition in Lemma 4 holds then there exists, $1 \leq \beta \leq \min\{\frac{a}{\bar{a}\gamma}, \frac{b\gamma}{\bar{b}}\}$ such that $\forall n$, the inequalities

$$\begin{aligned} \beta P_1(x^{n-1}) \geq P_0(x^{n-1}),\ \forall x^{n-1}, \\ \beta P_0(x^{n-1}) \geq P_1(x^{n-1}),\ \forall x^{n-1}, \end{aligned} \tag{33}$$

imply

$$\begin{aligned} \beta P_1(x^n) \geq P_0(x^n),\ \forall x^n, \\ \beta P_0(x^n) \geq P_1(x^n),\ \forall x^n. \end{aligned} \tag{34}$$

V. 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 which optimizes the directed information when feedback is allowed, and then proving that this output pmf can be 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 simple capacity-achieving codes.

Acknowledgement

The authors are grateful to Jiantao Jiao who suggested a proof of an inequity that we used in order to prove Lemma 6.

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] A.J. Goldsmith and P.P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inf. Theory*, 43(6):1986 –1992, 1997.

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

[11] H. Asnani, H. Permuter, and T. Weissman. Capacity of a post channel with and without feedback. In *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.

[12] H. Permuter, H. Asnani, and T. Weissman. Capacity of a post channel with and without feedback. submitted to *IEEE Trans. Inf. Theory.* Available at arxiv.org/abs/1309.5440, 2013.

[13] 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.

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

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

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

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

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

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

[20] 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.

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

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

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