Feedback capacity and coding for the BIBO channel with a no-repeated-ones input constraint

Oron Sabag, *Student Member, IEEE,* Haim H. Permuter, *Senior Member, IEEE,* Navin Kashyap, *Senior Member, IEEE*

Abstract—In this paper, a general binary-input binary-output (BIBO) channel is investigated in the presence of feedback and input constraints. The feedback capacity and the optimal input distribution of this setting are calculated for the case of an $(1, \infty)$-RLL input constraint, that is, the input sequence contains no consecutive ones. These results are obtained via explicit solution of an equivalent dynamic programming optimization problem. A simple coding scheme is designed based on the principle of posterior matching, which was introduced by Shayevitz and Feder for memoryless channels. The posterior matching scheme for our input-constrained setting is shown to achieve capacity using two new ideas: *history bits*, which captures the memory embedded in our setting, and *message-interval splitting*, which eases the analysis of the scheme. Additionally, in the special case of an S-channel, we give a very simple zero-error coding scheme that is shown to achieve capacity. For the input-constrained BSC, we show using our capacity formula that feedback increases capacity when the cross-over probability is small.

Index Terms—Binary channels, dynamic programming, feedback capacity, posterior matching scheme, runlength-limited (RLL) constraints.

I. INTRODUCTION

Consider the binary symmetric channel (BSC), described in Fig. 1 with $\alpha = \beta$, in the presence of output feedback. This setting is well understood in terms of capacity, $C = 1 - H_2(\alpha)$, but also in terms of efficient and capacity-achieving coding schemes such as the Horstein scheme [1] and the posterior matching scheme (PMS) [2]. However, imposing constraints on the input sequence, even in the simplest cases, makes the capacity calculation challenging, since this setting is equivalent to a finite-state channel. A special case of the setting studied here is the BSC with feedback and a no-consecutive-ones input constraint (Fig. 2), that is, the channel input sequence cannot contain adjacent ones. We will show for instance, that its feedback capacity still has a simple expression:

$$C = \max_p \frac{H_2(p) + pH_2\left(\frac{\alpha(1-\alpha)}{p}\right)}{1+p} - H_2(\alpha), \quad (1)$$

and that there exists an efficient coding scheme that achieves this feedback capacity. It is also interesting to understand

Fig. 1. BIBO channel with transition probabilities (α, β). Special cases are the Z and S channels, which correspond to $\alpha = 0$ and $\beta = 0$, respectively, and the BSC when $\alpha = \beta$.

the role of feedback on capacity when input constraints are present, and it will be proven that in contrast to the unconstrained BSC, *feedback does increase capacity* for the input-constrained BSC.

The capacity of input-constrained memoryless channels has been extensively investigated in the literature, but still there are no computable expressions for the capacity without feedback [3]–[7]. On the other hand, in [8], it was shown that if there is a noiseless feedback link to the encoder (Fig. 2), then the feedback capacity can be formulated as a dynamic programming (DP) problem, for which there exist efficient numerical algorithms for capacity computation [9]–[16]. However, as indicated by the authors of [8], analytic expressions for the feedback capacity and the optimal input distributions are still hard to obtain and remain an open problem. In this paper, both feedback capacity and the optimal input distribution of the binary-input binary-output (BIBO) channel (Fig. 1) with a no-consecutive-ones input constraint are derived by solving the corresponding DP problem. The BIBO channel includes as special cases the BSC ($\alpha = \beta$), which was studied in [8], the Z-channel ($\alpha = 0$) and the S-channel ($\beta = 0$).

Shannon proved that feedback does not increase the capacity of a memoryless channel [17]; following the proof of his theorem, he also claimed that "feedback does not increase the capacity for channels with memory if the internal channel state can be calculated at the encoder". The input-constrained setting studied here can be cast as a state-dependent channel, so it fits Shannon's description of such a channel. Therefore, we investigate the role of feedback for the special case of input-constrained BSC. In the regime $\alpha \to 0$, the feedback capacity from (1) is compared with a corresponding expression obtained for the capacity without feedback [18]. This comparison reveals that feedback increases capacity, at least for small enough values of α for the input-constrained BSC in contrast to Shannon's claim. However, this is not the first counterexample to Shannon's claim; two other such counterexamples can be found in [19], [20].
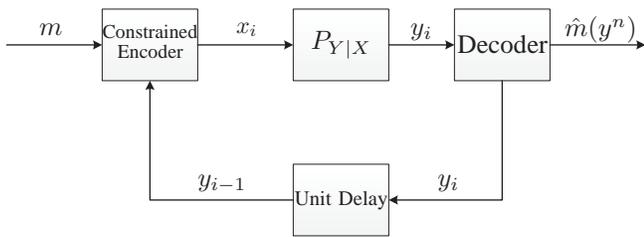
Fig. 2.   System model for an input-constrained memoryless channel with noiseless feedback.

In past works on channels with memory, such as [11]–[14], the optimal input distribution provided insights into the construction of simple coding schemes with zero error probability. This methodology also works for the S-channel, for which we are able to give a simple zero-error coding scheme. The coding scheme is similar to the "repeat each bit until it gets through" policy that is optimal for a binary erasure channel with feedback. In our case, each bit is repeated with its complement until $Y = 0$ is received, so the formed sequence is of alternating bits and satisfies the input constraint. However, a coding scheme for the general BIBO channel is challenging since $p(y|x) > 0$, for all (x,y), and therefore, there is no particular pattern of outputs for which a bit can be decoded with certainty. Nonetheless, we are able to use the structure of the optimal input distribution to give a simple coding scheme, based on the principle of posterior matching as is elaborated below.

Two fundamental schemes on sequential coding for memoryless channels with feedback date back to the work of Horstein [1] for the BSC, and that of Schalkwijk and Kailath [21] for the additive white Gaussian noise (AWGN) channel. In [2], Shayevitz and Feder established the strong connection between these coding schemes by introducing a generic coding scheme, termed the *posterior matching scheme* (PMS), for all memoryless channels. This work provided a rigorous proof for the optimality of such sequential schemes, a fact that may be intuitively correct but difficult to prove. Subsequent works proved the coding optimality using different approaches [22], [23], including an original idea by Li and El Gamal in [24] to introduce a randomizer that is available both to the encoder and the decoder. This assumption markedly simplifies the coding analysis, and it was adopted thereafter by [25] to simplify their original analysis in [2]. In our coding scheme, it is also assumed that there is a common randomizer available to all parties as a key step to the derivations of an optimal PMS for the BIBO channel.

The encoder principle in the PMS is to determine the channel inputs such that the optimal input distribution is simulated. For a memoryless channel, the optimal input distribution is i.i.d. so the encoder simulates the same experiment at all times. In the input-constrained setting, the input distribution is given by $p(x_i|x_{i-1}, y^{i-1})$ (inputs are constrained with probability 1), so the conditioning injects new information on which the encoder should depend. The first element in the conditioning, y^{i-1}, can be viewed as a time-sharing (not i.i.d.) since both the encoder and the decoder know this tuple. Indeed, it is

shown that they do not need to track the entire tuple y^{i-1}, but a recursive quantization of it on a directed graph. The second element, x_{i-1}, is a new element in the PMS since it is only available to the encoder, and it is handled by introducing a new idea called the *history bit* for each message. The analysis of the scheme is simplified using *message-interval splitting*, which results in a homogenous Markov chain instead of a time-dependent random process. These two ideas constitute the core of the PMS for the input-constrained setting, and it is shown that the coding scheme achieves the capacity of the general input-constrained BIBO channel.

The remainder of the paper is organized as follows. Section II presents our notation and a description of the problem we consider. Sections III and IV contain statements of the main technical results of the paper. In Section V, we provide the PMS for our input-constrained setting, while the optimality of this scheme is proved in Section VI. The DP formulation of feedback capacity together with its solution is presented in Section VII. Section VIII contains some concluding remarks. Some of the more technically involved proofs are given in appendices to preserve the flow of the presentation.

II. NOTATION AND PROBLEM DEFINITION

Random variables will be denoted by upper-case letters, such as X, while realizations or specific values will be denoted by lower-case letters, e.g., x. Calligraphic letters, e.g., $\mathcal{X}$, will denote the alphabets of the random variables. Let X^n denote the n-tuple $(X_1, \ldots, X_n)$ and let x^n denote the realization vectors of n elements, i.e., $x^n = (x_1, x_2, \ldots, x_n)$. For any scalar $\alpha \in [0,1]$, $\bar{\alpha}$ stands for $\bar{\alpha} = 1 - \alpha$. Let $H_2(\alpha)$ denote the binary entropy for the scalar $\alpha \in [0,1]$, i.e., $H_2(\alpha) = -\alpha \log_2 \alpha - \bar{\alpha} \log_2 \bar{\alpha}$.

The probability mass function (pmf) of a random variable X is denoted by $p_X(x)$, and conditional and joint pmfs are denoted by $p_{Y|X}(y|x)$ and $p_{X,Y}(x,y)$, respectively; when the random variables are clear from the context we use the shorthand notation $p(x)$, $p(y|x)$ and $p(x,y)$. The conditional distribution $p_{Y|X}$ is specified by a stochastic matrix $P_{Y|X}$, the rows of which are indexed by $\mathcal{X}$, the columns by $\mathcal{Y}$, and the (x,y)th entry is the conditional probability $p_{Y|X}(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

The communication setting (Fig. 2) consists of a message M that is drawn uniformly from the set $\{1, \ldots, 2^{nR}\}$ and made available to the encoder. At time i, the encoder produces a binary output, $x_i \in \{0, 1\}$, as a function of m, and the output samples y^{i-1}. The sequence of encoder outputs, $x_1 x_2 x_3 \ldots$, must satisfy the $(1,\infty)$-RLL input constraint, i.e., no consecutive ones are allowed. The transmission is over the BIBO channel (Fig. 1) that is characterized by two transition probabilities, $p_{Y|X}(1|0) = \alpha$ and $p_{Y|X}(0|1) = \beta$, where α and β are scalars from $[0,1]$. The channel is memoryless, i.e., $p(y_i|x^i, y^{i-1}) = p_{Y|X}(y_i|x_i)$ for all i.

Definition 1. A $(n, 2^{nR}, (1,\infty))$ *code* for an input-constrained channel with feedback is defined by a set of encoding functions:

$$f_i : \{1, \ldots, 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}, \ i = 1, \ldots, n,$$

satisfying $f_i(m, y^{i-1}) = 0$ if $f_{i-1}(m, y^{i-2}) = 1$ (the mapping $f_1(\cdot)$ is not constrained), for all (m, y^{i-1}), and by a decoding function $\Psi : \mathcal{Y}^n \to \{1, \ldots, 2^{nR}\}$.

The *average probability of error* for a code is defined as $P_e^{(n)} = \Pr[M \neq \Psi(Y^n)]$. A rate R is said to be $(1, \infty)$-*achievable* if there exists a sequence of $(n, 2^{nR}, (1, \infty))$ codes such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The *capacity*, $C^{\mathrm{fb}}(\alpha, \beta)$ is defined as the supremum over all $(1, \infty)$-achievable rates.

The transition probabilities can be restricted to $\alpha + \beta \leq 1$, a fact that is justified by:

Lemma II.1. *The capacity of a BIBO channel satisfies* $C(\alpha, \beta) = C(1 - \alpha, 1 - \beta)$, *for all* α, β.

Proof: For a channel with parameters (α, β), apply an invertible mapping $\tilde{Y} = Y \oplus 1$ on channel outputs so that the capacity remains the same but the parameters are changed to $(1 - \alpha, 1 - \beta)$. ∎

The proof of the lemma is valid even when the inputs are constrained and there is feedback to the encoder.

III. MAIN RESULTS

In this section, we present our main results concerning the feedback capacity of the BIBO channel, and thereafter, we show that feedback increases capacity for the BSC. The optimal PMS for the BIBO channel is not included in this section and appears in Section V.

A. Feedback capacity

The general expression for the feedback capacity is given by the following theorem

Theorem III.1 (BIBO capacity). *The feedback capacity of the input-constrained BIBO channel is*

$$C^{\mathrm{fb}}(\alpha, \beta) = \max_{z_L \leq z \leq z_U} R_{\alpha,\beta}(z), \qquad (3)$$

where $\alpha + \beta \leq 1$, $z_L = \frac{\sqrt{\alpha}}{\sqrt{\alpha}+\sqrt{\beta}}$, $z_U = \frac{\sqrt{\bar{\alpha}}}{\sqrt{\bar{\alpha}}+\sqrt{\beta}}$ *and* $R_{\alpha,\beta}(z)$ *is defined in* (2).
The feedback capacity can also be expressed by:

$$C^{\mathrm{fb}}(\alpha, \beta) = \log\left(\frac{1 - p_{\alpha,\beta}}{p_{\alpha,\beta} - \alpha\bar{\beta}}\right) + \beta\frac{H_2(\alpha)}{1 - \alpha - \beta} - \bar{\alpha}\frac{H_2(\beta)}{1 - \alpha - \beta}, \qquad (4)$$

where $p_{\alpha,\beta}$ *is the unique solution of*

$$(1 - \alpha\bar{\beta})[H_2(\beta) - H_2(\alpha)]$$
$$= (\bar{\beta} - \alpha)[2\log(1 - p) - \log(p - \alpha\bar{\beta})(1 + \alpha\bar{\beta}) + \alpha\bar{\beta}\log\alpha\bar{\beta}]. \qquad (5)$$

The proof of (3) in Theorem III.1 appears in Section VII and relies on the formulation of feedback capacity as a DP problem. From the solution of the DP, we only obtain that

the maximization in (3) is over $z \in [0, 1]$, but this can be strengthened using the following result:

Lemma III.1. *Let* $R_{\alpha,\beta}(z)$ *be as defined in* (2) *with* $0 \leq z \leq 1$. *The argument that achieves the maximum of* $R_{\alpha,\beta}(z)$ *is unique and lies within* $[z_L, z_U]$, *for all* $\alpha + \beta \leq 1$. *Additionally, for the BSC* ($\alpha = \beta$), *the maximum is attained when the argument is within* $[z_L, 0.5]$.

The proof of Lemma III.1 appears in Appendix A. The alternative capacity expression (4) is obtained by taking the derivative of (3) and substituting the resulting relation into the capacity expression (3). Note that the LHS of (5) is a decreasing function of p, and hence, efficient methods can be applied to calculate (4).

Remark 1. The feedback capacity can also be calculated using upper and lower bounds from [26], which turn out to meet for this channel, instead of the DP approach that is taken in this paper.

Theorem III.1 provides the capacity of three special cases: the BSC, the S-channel and the Z-channel. Their feedback capacities are calculated by substituting their corresponding parameters in Theorem III.1.

Corollary 1 (BSC capacity). *The feedback capacity of the input-constrained BSC* ($\alpha = \beta$) *is*

$$C^{\mathrm{BSC}}(\alpha) = \max_{\sqrt{\alpha\bar{\alpha}} \leq p \leq 0.5} \frac{H_2(p) + pH_2\left(\frac{\alpha\bar{\alpha}}{p}\right)}{1 + p} - H_2(\alpha), \qquad (6)$$

where $\alpha \leq 0.5$. *An alternative capacity expression is*

$$C^{\mathrm{BSC}}(\alpha) = \log\left(\frac{1 - p_\alpha}{p_\alpha - \alpha\bar{\alpha}}\right) - H_2(\alpha), \qquad (7)$$

where p_α *is the unique solution of* $(\alpha\bar{\alpha})^{\alpha\bar{\alpha}}(1 - p)^2 = (p - \alpha\bar{\alpha})^{1+\alpha\bar{\alpha}}$.

By operational considerations, the feedback capacity in Theorem III.1 serves as an upper bound for the non-feedback setting, which is still an open problem. For the BSC, it will be shown further in Theorem III.2 that feedback increases capacity, at least for small values of α, so this upper bound is not tight.

Corollary 2 (S-channel capacity). *The feedback capacity of the input-constrained S-channel* ($\beta = 0$) *is*

$$C^{\mathrm{S}}(\alpha) = \max_{\sqrt{\alpha} \leq p \leq 1} \frac{H_2(p) + pH_2\left(\frac{\alpha}{p}\right) - H_2(\alpha)}{1 + p}$$
$$= \max_{\sqrt{\alpha} \leq p \leq 1} \bar{\alpha}\frac{H_2\left(\frac{1-p}{1-\alpha}\right)}{1 + p} \qquad (8)$$

The capacity can also be expressed by:

$$C^{\mathrm{S}}(\alpha) = \log\left(\frac{1 - p_\alpha}{p_\alpha - \alpha}\right), \qquad (9)$$

$$R_{\alpha,\beta}(z) = \frac{H_2(\alpha z + \bar{\beta}z) + (\alpha z + \bar{\beta}z)H_2\left(\frac{\alpha\bar{\beta}}{\alpha\bar{z}+\bar{\beta}z}\right) - (\bar{z} + \bar{\beta}z)H_2(\alpha) - (z + \alpha\bar{z})H_2(\beta)}{1 + \alpha\bar{z} + \bar{\beta}z}. \qquad (2)$$

where p_α is the unique solution of $(1-p)^2 = (p-\alpha)^{1+\alpha}\bar{\alpha}^{\bar{\alpha}}$.

The second capacity expression in (8) reveals a simple zero-error coding scheme for the S-channel. To describe this, we first fix a $z \in (0,1)$, and consider a set, $\mathcal{M}$, consisting of $|\mathcal{M}| = 2^{NH_2(z)}$ messages, where N is a large integer.[1] The coding scheme operates in two stages:

1) Message shaping: The set of messages, $\mathcal{M}$, is mapped in a one-to-one fashion into the set of length-N binary sequences containing Nz 1s. Thus, each message is identified with a binary sequence of length N with fraction of 1s equal to z. This "message shaping" can be implemented, for instance, using the enumerative source coding technique [27, Example 2].

2) Message transmission: Each of the N bits, $b_1, b_2, \ldots, b_N$, in the shaped message sequence is transmitted by the encoder using the following procedure:

   To send the message bit b, the encoder transmits the sequence $b\bar{b}b\bar{b}\ldots$, where $\bar{b}$ denotes the complement (NOT) of b, until a 0 is received at the channel output, at which point the transmission is stopped.

   Note that if $y_1 y_2 \ldots y_{\ell-1} 0$ is the sequence received at the S-channel output in response to the transmission of the message bit b, then the decoder can determine whether $b = 0$ or $b = 1$ from the parity of ℓ: if ℓ is odd, then $b = 0$; if ℓ is even, then $b = 1$.

By the law of large numbers, the number of S-channel uses needed for the transmission of an N-bit shaped message sequence is close to $N \times \mathbb{E}[L_z]$, where $\mathbb{E}[L_z]$ denotes the expected number of transmissions needed for sending a single Bernoulli(z) bit b using the procedure described above. It is easy to check that $\mathbb{E}[L_z]$ equals

$$\bar{z}\bar{\alpha}\sum_{k=1}^{\infty}(2k-1)\alpha^{k-1} + z\bar{\alpha}\sum_{k=1}^{\infty}2k\alpha^{k-1} = -\bar{z} + \bar{\alpha}\sum_{k=1}^{\infty}2k\alpha^{k-1}$$

$$= \frac{1+p}{1-\alpha}, \qquad (10)$$

where $p = z + \alpha\bar{z}$. Thus, the rate achieved by this scheme is (arbitrarily close to) $\frac{\log|\mathcal{M}|}{N\,\mathbb{E}[L_z]} = \frac{H_2(z)}{\mathbb{E}[L_z]} = \frac{1-\alpha}{1+p}H_2\left(\frac{1-p}{1-\alpha}\right)$. Maximizing over $z \in (0,1)$, we conclude that the scheme achieves the S-channel capacity given by (8).

Corollary 3 (Z-channel capacity). *The feedback capacity of the input-constrained Z-channel ($\alpha = 0$) is*

$$C^Z(\beta) = \max_{0 \le p \le \bar{\beta}} \frac{H_2(p) - \frac{p}{1-\beta}H_2(\beta)}{1+p}$$

$$= -\log(1 - p_\beta), \qquad (11)$$

where p_β is the unique solution of the quadratic equation $(1-p)^2 = p \cdot 2^{\frac{H_2(\beta)}{1-\beta}}$.

The feedback capacities of the input-constrained S and Z channels are different because of the asymmetry imposed by the input constraint (Fig. 3). Note that for most values of the channel parameters, the capacity of the S-channel exceeds that of the Z-channel; intuitively, the decoder can gain more information when observing two consecutive ones in the channel output because it knows that there is one error in this transmission pair.
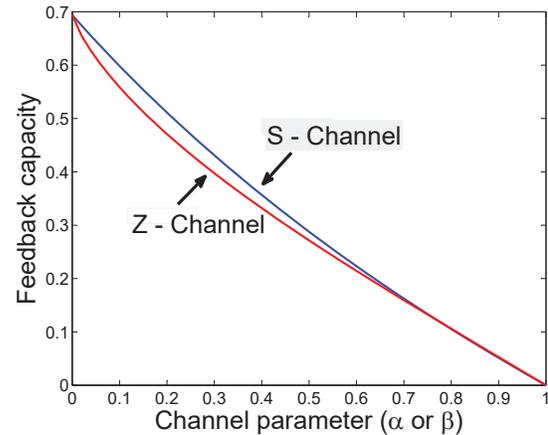
Fig. 3. Comparison between the capacities of the constrained Z- and S-channels.

B. Feedback increases capacity

In this section, we show that feedback increases capacity for the input-constrained BSC.

Theorem III.2. *Feedback increases capacity for the $(1,\infty)$-RLL input-constrained BSC, for all values of α in some neighborhood around 0.*

As discussed in Section I, this gives a counterexample to a claim of Shannon's from [17]. A subsequent work [28] related to the conference version of our paper [29] used a novel technique to calculate upper bounds on the non-feedback capacity of the input-constrained BSC. The upper bound in [28] is a tighter upper bound than our feedback capacity, which shows that feedback increases capacity not only for small values of α, but actually for all α.

In order to show Theorem III.2, we provide the asymptotic expressions of the input-constrained BSC with and without feedback.

Theorem III.3. *The feedback capacity of the input-constrained BSC is:*

$$C^{\mathrm{BSC}}(\alpha) = \log\lambda + \frac{2-\lambda}{3-\lambda}\alpha\log\alpha$$

$$+ \left(\frac{\log(2-\lambda) - (2-\lambda)}{3-\lambda}\right)\alpha + O(\alpha^2\log^2\alpha), \qquad (12)$$

where λ is the golden ratio $\left(\lambda = \frac{1+\sqrt{5}}{2}\right)$.

The derivation of Theorem III.3 is more involved than standard Taylor series expansion about $\alpha = 0$, since the second-order term of (12) is $O(\alpha\log\alpha)$. The proof of Theorem III.3 appears in Appendix B.

The asymptotic behaviour of the capacity of the input-constrained BSC without feedback is captured by the following result.

Theorem III.4. *[18, Example 4.1] The non-feedback capacity of the $(1, \infty)$-RLL input-constrained BSC is:*

$$C^{\mathrm{NF}}(\alpha) = \log \lambda + \frac{2\lambda + 2}{4\lambda + 3} \alpha \log \alpha + O(\alpha). \qquad (13)$$

It is now easy to prove Theorem III.2.

Proof of Theorem III.2: The coefficients of the term $\alpha \log \alpha$ in (12) and (13) satisfy $\frac{2\lambda+2}{4\lambda+3} > \frac{2-\lambda}{3-\lambda}$. Therefore, there exists $\alpha^* > 0$ such that $C^{\mathrm{BSC}}(\alpha) - C^{\mathrm{NF}}(\alpha) > 0$, for all $\alpha < \alpha^*$. ∎

IV. The Optimal Input Distribution

In this section, we present the optimal input distribution for the input-constrained BIBO, based on which the capacity-achieving coding scheme of the next section is derived. The optimization problem that needs to be solved when calculating the feedback capacity of our setting is given in the following theorem.

Theorem IV.1 ([13], Theorem 3)**.** *The capacity of an $(1, \infty)$-RLL input-constrained memoryless channel with feedback can be written as:*

$$C^{\mathrm{fb}} = \sup \liminf_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} I(X_t; Y_t | Y^{t-1}), \qquad (14)$$

where the supremum is taken with respect to $\{p_{X_t|X_{t-1}, Y^{t-1}} : p_{X_t|X_{t-1}, Y^{t-1}}(1|1, y^{t-1}) = 0\}_{t \geq 1}$.

The input at time t depends on the previous channel input, x_{t-1}, and the output samples y^{t-1}. The description of such an input distribution is difficult since the conditioning contains a time-increasing domain, $\mathcal{Y}^{t-1}$. The essence of the DP formulation is to replace the conditioning on y^{t-1} with $p_{X_{t-1}|Y^{t-1}}(0|y^{t-1})$, which is a sufficient statistic of the outputs tuple. Furthermore, the DP solution in Section VII reveals that the *optimal input distribution* can be described with an even simpler notion called a Q-graph, which is suitable for scenarios where the DP state, $p_{X_{t-1}|Y^{t-1}}(0|y^{t-1})$, takes a finite number of values.

Definition 2. For an output alphabet, $\mathcal{Y}$, a Q-graph is a directed, connected and labeled graph. Additionally, each node should have $|\mathcal{Y}|$ outgoing edges, with distinct labels.

The Q-graph depicted in Fig. 4 will be used to describe the optimal input distribution. Let $\mathcal{Q} = \{1, 2, 3, 4\}$ denote the set of nodes of this Q-graph. We will use a function $g : \mathcal{Q} \times \{0, 1\} \to \mathcal{Q}$ to record the transitions along the edges of the graph. Specifically, $g(1, 0) = 4$, $g(1, 1) = 2$, $g(2, 0) = 3$, $g(2, 1) = 1$, $g(3, 0) = 3$, $g(3, 1) = 1$, $g(4, 0) = 3$, and $g(4, 1) = 1$. Given some initial node $q_0 \in \mathcal{Q}$ and an output sequence $y^t \in \mathcal{Y}^t$ of arbitrary length, a unique node $q_t \in \mathcal{Q}$ is determined by a walk on the Q-graph starting at q_0 and following the edges labeled by $y_1, y_2, \ldots, y_t$, in that order. We will write this as $q_t = \Phi(q_0, y^t)$, where $\Phi : \mathcal{Q} \times \bigcup_{t \geq 1} \mathcal{Y}^t \to \mathcal{Q}$ is the mapping recursively described by $\Phi(q_0, y_1) = g(q_0, y_1)$,
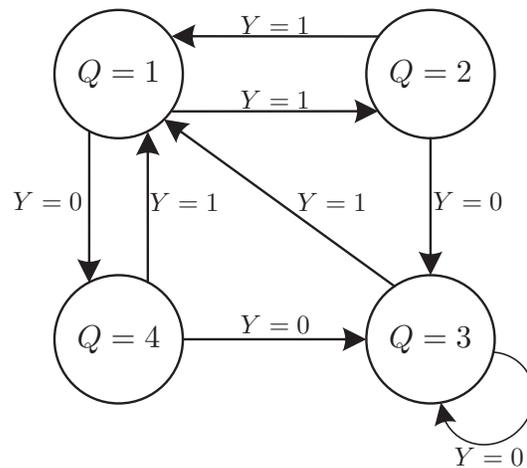
Fig. 4. The Q-graph that characterizes the optimal input distribution.

and $\Phi(q_0, y^t) = g(\Phi(q_0, y^{t-1}), y_t)$ for $t \geq 2$. The importance of the Q-graph for our scheme is that the encoder and decoder need only track the value $\Phi(q_0, y^{t-1})$, instead of the entire output sequence y^{t-1}.

For the description of the optimal input distribution, define $z_2^{\alpha, \beta} \triangleq \arg\max_{0 \leq z \leq 1} R_{\alpha, \beta}(z)$ as the unique maximizer from Lemma III.1, and the following subsequent quantities:

$$z_1^{\alpha, \beta} \triangleq \frac{\alpha \bar{z}_2^{\alpha, \beta}}{\alpha \bar{z}_2^{\alpha, \beta} + \bar{\beta} z_2^{\alpha, \beta}}$$

$$z_3^{\alpha, \beta} \triangleq \frac{\bar{\alpha} \bar{z}_2^{\alpha, \beta}}{\bar{\alpha} \bar{z}_2^{\alpha, \beta} + \beta z_2^{\alpha, \beta}}$$

$$z_4^{\alpha, \beta} \triangleq \frac{\bar{\alpha} \bar{\beta} z_2^{\alpha, \beta}}{\bar{\alpha} \bar{\beta} z_2^{\alpha, \beta} + \alpha \beta \bar{z}_2^{\alpha, \beta}}. \qquad (15)$$

It can be shown that $z_1^{\alpha, \beta} \leq z_2^{\alpha, \beta} \leq z_3^{\alpha, \beta} \leq z_4^{\alpha, \beta}$ for all $\alpha + \beta \leq 1$. For instance, the relation $z_1 \leq z_2$ (superscripts (α, β) are omitted) can be simplified to $(\bar{\beta} - \alpha)z_2^2 + 2\alpha z_2 - \alpha \geq 0$. Now, the polynomial $(\bar{\beta} - \alpha)x^2 + 2\alpha x - \alpha$ has two roots, one is negative and the other is at $x = z_L$. Since the polynomial is convex, $(\bar{\beta} - \alpha)z_2^2 + 2\alpha z_2 - \alpha \geq 0$ is equivalent to $z_2 \geq z_L$. Using the same methodology, it can be shown that $z_1 \leq z_2 \leq z_3 \leq z_4$ is equivalent to $z_L \leq z_2 \leq z_U$, which is proved in Lemma III.1.

Define the conditional distributions $p^*_{X|X^-, Q}$ via the conditional probability matrices

$$p^*_{X|X^-, Q=1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$p^*_{X|X^-, Q=2} = p^*_{X|X^-, Q=1}$$

$$p^*_{X|X^-, Q=3} = \begin{bmatrix} 1 - \frac{z_2^{\alpha, \beta}}{z_3^{\alpha, \beta}} & \frac{z_2^{\alpha, \beta}}{z_3^{\alpha, \beta}} \\ 1 & 0 \end{bmatrix} \qquad (16)$$

$$p^*_{X|X^-, Q=4} = \begin{bmatrix} 1 - \frac{z_2^{\alpha, \beta}}{z_4^{\alpha, \beta}} & \frac{z_2^{\alpha, \beta}}{z_4^{\alpha, \beta}} \\ 1 & 0 \end{bmatrix},$$

in which X^- indexes the rows and X indexes the columns. To be precise, the first (resp. second) row of each matrix is a conditional pmf of X given $X^- = 0$ (resp. $X^- = 1$).

The optimal input distribution and alternative capacity expression are given in the following theorem:

Theorem IV.2 (Optimal input distribution). *For any $q_0 \in \mathcal{Q}$, the input distribution $p_{X_i|X_{i-1},Y^{i-1}}(x \mid x^-, y^{i-1}) = p^*_{X|X^-,Q}(x \mid x^-, \Phi(q_0, y^{i-1}))$, defined via (16) and Fig. 4, is capacity-achieving. Moreover, the random process $\{(X_i, Q_i)\}_{i \geq 1}$ induced by $p^*_{X|X^-,Q}$ is an irreducible and aperiodic Markov chain on $\{0,1\} \times \mathcal{Q}$. The stationary distribution of this Markov chain is given by $\pi_{X^-,Q} = \pi_Q \pi_{X^-|Q}$, where π_Q is the pmf on $\mathcal{Q}$ defined by $[\pi_Q(1), \pi_Q(2), \pi_Q(3), \pi_Q(4)] = \left[\frac{p}{1+p}, \frac{pq}{1+p}, \frac{1-p}{1+p}, \frac{p(1-q)}{1+p}\right]$, with $p = \alpha(1 - z_2^{\alpha,\beta}) + (1 - \beta)z_2^{\alpha,\beta}$ and $q = \frac{\alpha(1-\beta)}{p}$, and $\pi_{X^-|Q}(0|i) = 1 - \pi_{X^-|Q}(1|i) = z_i^{\alpha,\beta}$ for $i = 1, \ldots, 4$. The feedback capacity $C^{\mathrm{fb}}(\alpha, \beta)$ can be expressed as $I(X; Y|Q)$, where the joint distribution is $\pi_{Q,X,Y}(q, x, y) = \sum_{x^-} p_{Y|X}(y|x) p^*_{X|X^-,Q}(x|x^-, q) \pi_{X^-,Q}(x^-, q).$*

The scheme uses the joint probability distribution $p_{Y|X} p^*_{X|X^-,Q} \pi_{X^-,Q}$ induced by the optimal input distribution $p^*_{X|X^-,Q}$. Here, X and X^- should be viewed as the channel inputs during the current and previous time instances, respectively, and Q is the value of the node on the Q-graph prior to the transmission of X. In the analysis of the coding scheme, we will use the Markov property of $\{(X_i, Q_i)\}_{i \geq 1}$ to show that $I(X; Y|Q)$ is achievable. The proof of Theorem IV.2 is presented at the end of Section VII.

V. The Coding Scheme

The coding scheme we describe here consists of two phases: Phase I is based on a posterior matching scheme (PMS), and Phase II is a clean-up phase based on block codes.

The main element of any PMS is the posterior distribution of the message given the channel outputs. The posterior distribution is represented by the lengths of sub-intervals that form a partition of the unit interval. Each sub-interval is associated with a particular message, and henceforth, it will be referred to as a "message interval". The initial lengths are equal for all message intervals, since the decoder is assumed to have no prior information about the messages. The lengths of the message intervals are updated throughout the transmission based on the outputs that are made available to the decoder (and to the encoder from the feedback). The encoder's job is to refine the decoder's knowledge about the correct message by simulating samples from a desired input distribution. When this is done properly, as time progresses, the length of the true message interval will increase towards 1, and the decoder can then successfully declare its estimate of the correct message.

The above description of PMS is generic and applies to any setting of channel coding with feedback. What is specific to our PMS in the input-constrained setting is the input distribution that the encoder attempts to simulate during its operation. Most of the adaptations needed for our PMS that are not present in the baseline PMS for memoryless channels in [2] and [24] are a natural consequence of the input constraints and the structure of the input distribution in Theorem IV.2. However, these adaptations complicate the analysis of the scheme as the evolution in time of the involved random variables results in a random process that is difficult to analyze.

The analysis becomes easier upon introducing a certain message-interval splitting operation (described in Section V-D) that induces a Markov chain structure on the time-evolution of the random variables in the scheme. However, the splitting operation prevents the length of the correct message interval from increasing to 1, but we will show that, with high probability, the length of this interval will eventually go above some positive constant. The PMS output at the end of Phase I will be a list of messages whose interval lengths are above this constant.

In Phase II, a fixed-length block coding scheme, asymptotically of zero rate, is used to determine which message in the list produced at the end of Phase I is the correct message.

The remainder of this section is organized as follows. In Section V-A, the key elements of the PMS of Phase I are described. This is followed by a description of Phase II of our coding scheme in Section V-B. The overall coding scheme that combines the two phases is shown to be capacity-achieving in Section V-C. At the heart of the PMS of Phase I is a recursive construction of message intervals, which is described in detail in Section V-D. This technical description has been left to the end so as not to distract the reader from the main ideas of the coding scheme.

A. Phase I: PMS

The PMS is based on the joint probability distribution $\pi_{Y,X,X^-,Q}$ on $\mathcal{Y} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Q}$ defined by $\pi_{Y,X,X^-,Q} := \pi_{X^-,Q} p^*_{X|X^-,Q} p_{Y|X}$, where $\pi_{X^-,Q}$ and $p^*_{X|X^-,Q}$ constitute the optimal input distribution described in Section IV, and $p_{Y|X}$ is the channel law. In what is to follow, we routinely use notation such as π_Q, $\pi_{X^-|Q}$, $\pi_{X,X^-|Q}$ etc. to denote certain marginal and conditional probability distributions specified by $\pi_{X,X^-,Q}$. Thus, for example, π_Q denotes the marginal distribution on Q of $\pi_{X,X^-,Q}$. The joint distribution $\pi_{Y,X,X^-,Q}$ is known to the encoder and the decoder.

It is further assumed that the encoder and the decoder share some common randomness:

- a random initial Q-state[2] Q_0, distributed according to π_Q;
- a sequence $(U_i)_{i=1}^n$ of i.i.d. random variables, each U_i being Unif$[0, 1]$-distributed.

At the ith time instant, just prior to the transmission of the ith symbol by the encoder, both the decoder and the encoder have the output sequence y^{i-1} available to them. From this, and having shared knowledge of a realization q_0 of the initial Q-state Q_0, each of them can compute $q_{i-1} = \Phi(q_0, y^{i-1})$.

The assumption of shared randomness simplifies much of our analysis. It should be noted that by standard averaging arguments, the shared knowledge can be "de-randomized", in the sense that there exists a deterministic instantiation of Q_0 and (U_i) for which our probability of error analysis will remain valid.

---

[2]For the purposes of this description, we use the term "Q-state" to denote a node on the Q-graph.

Aside from the shared randomness, the encoder alone has access to an i.i.d. sequence $(V_i)_{i=1}^n$, with $V_i \sim \text{Unif}[0,1]$ for all i. We next describe the main elements of our PMS.

1) Messages and message intervals: At the outset, there is a set of messages $\mathcal{M} = \{1, 2, \ldots, 2^{nR}\}$, where n is a sufficiently large positive integer and $R < I(X; Y \mid Q)$.[3] A message $M \in \mathcal{M}$ is selected uniformly at random, and transmitted using the PMS scheme. Our aim is to show that the probability of error $P_e^{(n)}$ at the decoder goes to 0 exponentially in n.

Message intervals are central to the operation of the PMS scheme. At each time instant $i \geq 1$, again just prior to the transmission of the ith symbol by the encoder, the decoder and the encoder can compute a common set, $\mathcal{J}_i$, of message intervals, which form a partition of $[0, 1)$ into disjoint sub-intervals of varying lengths. The construction is recursive, with $\mathcal{J}_1$ being computed from the initial knowledge of q_0, and for $i \geq 1$, $\mathcal{J}_{i+1}$ being computed from $\mathcal{J}_i$ based on the additional knowledge of y_i and u_i. Here, u_i denotes a realization of the shared randomness U_i. We will describe the recursive construction in Section V-D, for now just noting that each message interval in $\mathcal{J}_{i+1}$ is obtained from a unique "parent" in $\mathcal{J}_i$.

Each message interval in $\mathcal{J}_i$ is indexed by an integer between 1 and $|\mathcal{J}_i|$, with $\iota(\mathsf{J})$ denoting the index of the message interval J. Additionally, each message interval is associated with a message from $\mathcal{M}$, the association being specified by a surjective mapping $\mu_i : \mathcal{J}_i \to \mathcal{M}$. Thus, for each $\mathsf{J} \in \mathcal{J}_i$, $\mu_i(\mathsf{J})$ is the message associated with J. One of the message intervals $\mathsf{J} \in \mathcal{J}_i$ with $\mu_i(\mathsf{J}) = m$, where $m \in \mathcal{M}$ is the actual message to be transmitted, is designated as the *true message interval*, and is denoted by J_i. The identity of J_i is *a priori* known only to the encoder. This is the message interval that determines the symbol to be transmitted at time i.

Each $\mathsf{J} \in \mathcal{J}_i$ has a "history bit", denoted by $x_i^-(\mathsf{J})$. This is the bit that the encoder would have transmitted at time $i-1$ if it were the case that the parent of J was the true message interval J_{i-1}.

A message interval $\mathsf{J} \in \mathcal{J}_i$ can be uniquely identified by its left end-point $t_i(\mathsf{J})$ and its length $s_i(\mathsf{J})$, so that $\mathsf{J} = [t_i(\mathsf{J}), t_i(\mathsf{J}) + s_i(\mathsf{J}))$. The length of J equals the posterior probability that J is the true message interval, given y^{i-1} and u^{i-1}, i.e.,

$$s_i(\mathsf{J}) = \Pr[J_i = \mathsf{J} \mid Y^{i-1} = y^{i-1}, U^{i-1} = u^{i-1}]. \quad (17)$$

In section V-D, it will be shown that the lengths $s_i(\mathsf{J})$, $\mathsf{J} \in \mathcal{J}_i$, can be computed recursively as a simple function of the lengths of the parent intervals in $\mathcal{J}_{i-1}$.

The left end-points are then computed as

$$t_i(\mathsf{J}) = \pi_{X^-|Q}(0|q_{i-1}) \mathbb{1}_{\{x_i^-(\mathsf{J})=1\}} + \sum_{\substack{\mathsf{J}':\ \iota(\mathsf{J}')<\iota(\mathsf{J}),\\ x_i^-(\mathsf{J}')=x_i^-(\mathsf{J})}} s_i(\mathsf{J}'). \quad (18)$$

[3]For ease of description, 2^{nR} is assumed to be an integer; we may otherwise take the number of messages to be $\lceil 2^{nR} \rceil$.
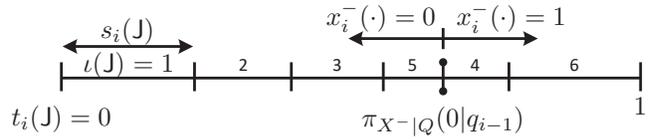
Fig. 5.   Illustration of the message intervals. Note that the lengths of all messages intervals with $x_i^- = 0$ sum up to $\pi_{X^-|Q}(0|q)$.

We will show later (see Lemma V.4) that

$$\sum_{\mathsf{J} \in \mathcal{J}_i : x_i^-(\mathsf{J})=0} s_i(\mathsf{J}) = \pi_{X^-|Q}(0|q_{i-1}),$$

so that the positioning of the left end-points in (18) implies that the message intervals with history bit 0 (respectively, 1) form a partition of $[0, \pi_{X^-|Q}(0|q_{i-1}))$ (respectively, $[\pi_{X^-|Q}(0|q_{i-1}), 1))$. The positioning of the message intervals is illustrated in Fig. 5.

In summary, a message interval $\mathsf{J} \in \mathcal{J}_i$ stores five pieces of data: $\iota(\mathsf{J})$, $\mu_i(\mathsf{J})$, $s_i(\mathsf{J})$, $t_i(\mathsf{J})$ and $x_i^-(\mathsf{J})$. As we will see in Section V-D, these are computable at both the encoder and the decoder from their common knowledge at time i.

2) Encoder operation at time i: To describe the symbol transmitted by the encoder at time i, we need to introduce the labeling functions $\mathcal{L}_q : [0, 1) \to \{0, 1\}$, defined for each $q \in \mathcal{Q}$. The labeling $\mathcal{L}_q$ assigns the label '1' to the interval $[0, \pi_{X,X^-|Q}(1, 0|q))$, and '0' to the interval $[\pi_{X,X^-|Q}(1, 0|q), 1)$. In other words,

$$\mathcal{L}_q(x) = \begin{cases} 1 & \text{if } 0 \leq x < \pi_{X,X^-|Q}(1, 0|q) \\ 0 & \text{if } \pi_{X,X^-|Q}(1, 0|q) \leq x < 1. \end{cases} \quad (19)$$

The labeling $\mathcal{L}_q$ is depicted in Fig. 6.

At time i, the encoder knows the true message interval J_i in $\mathcal{J}_i$. Let t_i and s_i denote its left end-point and length, respectively, and let x_i^- be its history bit. If $x_i^- = 1$, then J_i is contained in $[\pi_{X^-|Q}(0|q_{i-1}), 1)$, which is a subset of $\mathcal{L}_{q_{i-1}}^{-1}(0) = [\pi_{X,X^-|Q}(1, 0|q_{i-1}), 1)$. In this case, the encoder transmits $x_i = \mathcal{L}_{q_{i-1}}(J_i) = 0$, in keeping with the $(1, \infty)$-RLL constraint.

On the other hand, if $x_i^- = 0$, the encoder transmits the bit $x_i = \mathcal{L}_{q_{i-1}}(w_i)$ with

$$w_i = t_i + u_i(q_{i-1}) + s_i \cdot v_i \mod \pi_{X^-|Q}(0|q_{i-1}),$$

where $u_i(q_{i-1}) := u_i \cdot \pi_{X^-|Q}(0|q_{i-1})$, and v_i denotes a realization of the encoder's private randomness $V_i \sim \text{Unif}[0,1]$. In other words, the encoder picks w_i uniformly at random from the interval $J_i + u_i(q_{i-1}) \mod \pi_{X^-|Q}(0|q_{i-1})$, which is obtained by cyclically shifting the message interval J_i by the amount $u_i(q_{i-1})$, within $[0, \pi_{X^-|Q}(0|q_{i-1}))$.

Remark 2. The rationale behind the cyclic shifting is the following: the random variable $W_i = t_i + U_i(q_{i-1}) + s_i \cdot V_i \mod \pi_{X^-|Q}(0|q_{i-1})$ is uniformly distributed over $[0, \pi_{X^-|Q}(0|q_{i-1}))$. Hence, $X_i = \mathcal{L}_{q_{i-1}}(W_i)$ is equal to 1 with probability $\frac{\pi_{X,X^-|Q}(1, 0|q_{i-1})}{\pi_{X^-|Q}(0|q_{i-1})} = p_{X|X^-,Q}^*(1|0, q_{i-1})$, and is equal to 0 with probability $p_{X|X^-,Q}^*(0|0, q_{i-1})$. Thus, the cyclic shifting ensures that the conditional distribution of X_i

(given that the previously transmitted bit was x_i^- and the Q-state just prior to the transmission of X_i is q_{i-1}) matches the optimal input distribution $p_{X|X^-,Q}^*(\cdot|x_i^-, q_{i-1})$.

In the analysis of the PMS schemes of [2] and [24], most of the effort goes in showing that the length of the true message interval gets arbitrarily close to 1, with high probability, as the number of transmissions n goes to ∞. In our case, however, the recursive construction of $\mathcal{J}_{i+1}$ from $\mathcal{J}_i$ involves a key message-interval splitting operation, which prevents the lengths of message intervals from getting too large. Nonetheless, we can show that the length of the true message interval $s_i(J_i)$ either exceeds

$$S_{\min} := \min_{x,q:\ \pi_{X,X^-|Q}(x,0|q)>0} \pi_{X,X^-|Q}(x,0|q) \quad (20)$$

at some time i, or as the time index i gets close to n, $s_i(J_i)$ exceeds a certain threshold $\xi > 0$ with high probability. The following theorem, proved in Section VI, gives a rigorous statement of this fact.

Theorem V.1. *Given $R < I(X;Y|Q)$, there exists $\xi > 0$ (which may depend on R) for which the following holds: for any sequence of non-negative integers $(\zeta_n)_{n\geq 1}$ growing as $o(n)$,[4] a PMS initiated with 2^{nR} messages has*

$$\Pr\left[s_{n_\zeta}(J_{n_\zeta}) < \xi \mid s_i(J_i) \leq S_{\min}, \text{ for } i \in [1:n_\zeta-1]\right] \longrightarrow 0,$$

as $n \to \infty$ and $n_\zeta := n - \zeta_n$.

Recall that the decoder does not know which of the intervals in $\mathcal{J}_i$ is the true message interval, but it is able to compute the lengths $s_i(\mathsf{J})$ for all $\mathsf{J} \in \mathcal{J}_i$. The theorem above allows the decoder to create a relatively short list of potential true message intervals.

3) Decoder decision: Let $\xi^* = \min\{\xi, S_{\min}\}$, where ξ is as in Theorem V.1 above, and $S_{\min}$ is as defined in (20). The decoder halts operations at time $T = n - \lceil\sqrt{n}\rceil$, and outputs the list of messages

$$\widehat{\mathcal{M}} = \{\mu_i(\mathsf{J}) : \mathsf{J} \in \mathcal{J}_i,\ s_i(\mathsf{J}) \geq \xi^*, 1 \leq i \leq T\}. \quad (21)$$

In words, this is the set of messages associated with message intervals whose lengths exceed either $S_{\min}$ or ξ at any point during the operation of the PMS. This signals the end of Phase I of our coding scheme.

Note that since the lengths of all message intervals at time i must sum to 1, there can be at most $\lfloor 1/\xi^* \rfloor$ messages contributed by $\mathcal{J}_i$ to $\widehat{\mathcal{M}}$, for each $i \in \{1, 2, \ldots, T\}$. Thus, we have $|\widehat{\mathcal{M}}| \leq \frac{1}{\xi^*}T \leq \frac{1}{\xi^*}n$.

B. Phase II: Clean-Up

The message list generated at the end of Phase I serves as the input for Phase II, a complementary coding scheme to determine the correct message within $\widehat{\mathcal{M}}$. The rate of the coding scheme in this phase can be made to go to zero, since it only has to distinguish between $O(n)$ many messages in $\mathcal{M}$. Each message in $\widehat{\mathcal{M}}$ is represented using $k_n = \lceil \log_2(n/\xi^*) \rceil$ bits, $b_1 b_2 \ldots b_{k_n}$, which are to be transmitted successively.

[4]This means that $\zeta_n/n \to 0$ as $n \to \infty$.

1) Encoding: Given a string of k bits, $b_1, \ldots b_{k_n}$, the encoder transmits each bit b_i using a length-L_n codeword $b_i 0 b_i 0 \ldots b_i 0$, where L_n is a suitably chosen even number. Thus, $k_n L_n$ channel uses are required to transmit these k_n bits, but as we will see, L_n will be chosen so that $k_n L_n = o(n)$, so that this is an asymptotically vanishing fraction of the overall number of channel uses, n. Note that the encoder does not make use of feedback.

2) Decoding: Based on the sequence of $L_n/2$ channel outputs received in response to the $L_n/2$ repetitions of the bit b_i, the decoder declares $\hat{b}_i = 0$ if the output sequence lies in the typical set $\mathcal{T}_\epsilon^{(L_n/2)}(p_{Y|X=0})$ for a well-chosen $\epsilon > 0$, and declares $\hat{b}_i = 1$ otherwise. The decoder's estimate of the transmitted message M is the message $\hat{M}$ represented by $\hat{b}_1 \hat{b}_2 \ldots \hat{b}_{k_n}$.

The next lemma shows that the probability of decoding error for one message bit b can be made arbitrarily small, by choosing L_n as a suitably slowly growing function of n. As usual, $\hat{b}$ in the statement of the lemma refers to the decoder's estimate of b.

Lemma V.1. *If the capacity of the BIBO channel (without feedback or input constraints) is nonzero, then there exists a constant $C_0 > 0$ such that for any $L_n \geq C_0 \log k_n$, we have $\Pr[\hat{b} \neq b] \leq \frac{1}{k_n^2}$.*

Proof: We use a standard typicality argument based on the fact that if capacity is non-zero, then

$$D(p_{Y|X=0} || p_{Y|X=1}) \neq 0,$$

and so there exists a sequence $\epsilon(\ell) > 0$ such that, for all sufficiently large ℓ, $\mathcal{T}_{\epsilon(\ell)}^{(\ell)}(p_{Y|X=0}) \bigcap \mathcal{T}_{\epsilon(\ell)}^{(\ell)}(p_{Y|X=1}) = \emptyset$ and both typical sets are nonempty.

Without loss of generality, assume that the transmitted bit is $b = 1$, so that the sequence of channel outputs in response to the sequence of transmitted 1s is i.i.d. $\sim p_{Y|X=1}$. By standard arguments, it can be shown that the probability that a length-$L/2$ output sequence is in $\mathcal{T}_{\epsilon'}^{(L/2)}(p_{Y|X=1})$ (for some $0 < \epsilon' < \epsilon(L/2)$) goes to 1 exponentially quickly in L, while the probability that the output sequence is in $\mathcal{T}_{\epsilon'}^{(L/2)}(p_{Y|X=0})$ decays to 0 exponentially quickly in L. Therefore, $\Pr[\hat{b} \neq b]$ can be made smaller than $\frac{1}{k_n^2}$ by choosing $L_n = C_0 \log k_n$ for a sufficiently large positive constant C_0. ∎

C. Combining Phases I and II

We now describe how Phases I and II are combined to obtain a capacity-achieving coding scheme. Fix an $R < I(X;Y|Q)$ and let ξ be as in Theorem V.1, which in turn determines $\xi^* = \min\{\xi, S_{\min}\}$. We will apply Theorem V.1 with $\zeta_n = \lceil\sqrt{n}\rceil$. Let $k_n = \lceil\log_2(n/\xi^*)\rceil$, and $L_n = \zeta_n/k_n$ be the parameters of the coding scheme in Phase II. Note that $L_n \geq C_0 \log k_n$ for all sufficiently large n, where C_0 is the constant in the statement of Lemma V.1.

We will run the PMS of Phase I on a message set $\mathcal{M}$ of size 2^{nR}. A message $M \sim \text{Unif}(\mathcal{M})$ is transmitted using n uses of the channel as follows: The PMS of Phase I is executed until time $T = n - \zeta_n$ (i.e., $T = n - \lceil\sqrt{n}\rceil$), at which time a list $\widehat{M}$ as in (21) is produced. We then execute Phase II for

the remaining $k_n L_n = \zeta_n$ channel uses, at the end of which the decoder produces an estimate $\hat{M}$ of M. The coding rate of the overall scheme is $\frac{1}{n} \log_2 |\mathcal{M}| = R$.

To assess the probability of error $P_e^{(n)}$, we observe that

$$
\begin{aligned}
P_e^{(n)} &= \Pr[M \neq \hat{M}] \\
&\leq \Pr[M \notin \widehat{\mathcal{M}}] + \Pr[M \neq \hat{M} \mid M \in \widehat{\mathcal{M}}] \\
&\leq \Pr[M \notin \widehat{\mathcal{M}}] + \sum_{i=1}^{k_n} \Pr[\hat{b}_i \neq b_i] \\
&\leq \Pr[M \notin \widehat{\mathcal{M}}] + \frac{1}{k_n}, \quad (22)
\end{aligned}
$$

the last inequality above being valid for all n large enough that $L_n \geq C_0 \log k_n$, so that the conclusion of Lemma V.1 holds. Let $n_\zeta := n - \zeta_n$ and then, the probability that $M \notin \widehat{\mathcal{M}}$ can be bounded as follows:

$$
\begin{aligned}
&\Pr[M \notin \widehat{\mathcal{M}}] \\
&\leq \Pr[s_i(J_i) \leq S_{\min} \text{ for } i \in [1 : n_\zeta - 1], \text{ and } s_{n_\zeta}(J_{n_\zeta}) < \xi] \\
&\leq \Pr[s_{n_\zeta}(J_{n_\zeta}) < \xi \mid s_i(J_i) \leq S_{\min} \text{ for } i \in [1 : n_\zeta - 1]],
\end{aligned}
$$

which, by Theorem V.1, goes to 0 as $n \to \infty$. Hence, by (22), we also have $P_e^{(n)} \to 0$ as $n \to \infty$.

We have thus shown that any rate $R < I(X; Y|Q)$ is $(1, \infty)$-achievable. Recall from Theorem IV.2 that $I(X; Y|Q) = C^{\mathrm{fb}}(\alpha, \beta)$. Thus, we have proved the main result of this section, stated below.

Theorem V.2. *For an input-constrained BIBO channel, any rate $R < C^{\mathrm{fb}}(\alpha, \beta)$ is $(1, \infty)$-achievable using a coding scheme that combines Phases I and II.*

It remains to tie a couple of loose ends in the description of our coding scheme, namely, the recursive construction of message intervals, and a proof of Theorem V.1. The former is presented in the subsection below, while the latter is given in Section VI.

D. Recursive Construction of Message Intervals

1) Initialization — Construction of $\mathcal{J}_1$: $\mathcal{J}_1$ consists of the 2^{nR} intervals $\mathsf{J}^{(j)} = [(j-1)2^{-nR}, j2^{-nR})$ of equal length, indexed by $j \in \{1, 2, \ldots, 2^{nR}\}$, in 1-1 correspondence with the 2^{nR} messages in $\mathcal{M}$. The length of each message interval is $s_1 := 2^{-nR}$. For the jth message interval $\mathsf{J}^{(j)}$, the index $\iota(\mathsf{J}^{(j)})$ and associated message $\mu(\mathsf{J}^{(j)})$ are both set to be j. The history bit of the $\mathsf{J}^{(j)}$ is set to be

$$
x_1^-(\mathsf{J}^{(j)}) = \begin{cases} 0 & \text{if } (j-1)2^{-nR} \leq \pi_{X^-|Q}(0|q_0) \\ 1 & \text{otherwise}, \end{cases}
$$

where q_0 is the initial Q-state known to both encoder and decoder.

We will, for simplicity of description, assume that n is chosen so that the message intervals in $\mathcal{J}_1$ with history bit equal to 0 form a partition of $[0, \pi_{X^-|Q}(0|q_0))$.[5] For a uniformly random message $M \in \mathcal{M}$, the true message interval

[5]If this is not the case, we can get this to happen by splitting into two the message interval that straddles the boundary point $\pi_{X^-|Q}(0|q_0)$, as described in the next subsection.

is $J_1 = \mathsf{J}^{(M)}$, and is *a priori* known only to the encoder. Note that with $X_0 = x_1^-(J_1)$, the pair (X_0, Q_0) has probability distribution $\pi_{X^-|Q} \pi_Q = \pi_{X^-, Q}$.

2) Recursion — Construction of $\mathcal{J}_{i+1}$ from $\mathcal{J}_i$: Recall that, for $i \geq 1$, the encoder and decoder can compute the set $\mathcal{J}_i$ based upon their common knowledge of q_0, y^{i-1} and u^{i-1}. After determining $\mathcal{J}_i$, they make use of their shared knowledge of u_i to compute a new partition, $\widetilde{\mathcal{J}}_i$, of $[0, 1)$ into message intervals. This is an intermediate step in the construction of $\mathcal{J}_{i+1}$ from $\mathcal{J}_i$. Message intervals $\mathsf{J} \in \widetilde{\mathcal{J}}_i$ also store five pieces of data: an index $\tilde{\iota}(\mathsf{J})$, a message $\tilde{\mu}_i(\mathsf{J}) \in \mathcal{M}$, a length $\tilde{s}_i(\mathsf{J})$, a left end-point $\tilde{t}_i(\mathsf{J})$, and a bit $x_i(\mathsf{J})$. These are explained as part of the description to follow.

Each message interval $\mathsf{J} \in \mathcal{J}_i$ is either retained as is in $\widetilde{\mathcal{J}}_i$, or is split into two intervals in $\widetilde{\mathcal{J}}_i$ using a procedure to be described shortly. Any message interval $\mathsf{J} \in \mathcal{J}_i$ with history bit $x_i^-(\mathsf{J}) = 1$ is retained as is in $\widetilde{\mathcal{J}}_i$. It retains its index, message, length and left end-point: $\tilde{\iota}(\mathsf{J}) = \iota(\mathsf{J})$, $\tilde{\mu}_i(\mathsf{J}) = \mu_i(\mathsf{J})$, $\tilde{s}_i(\mathsf{J}) = s_i(\mathsf{J})$, and $\tilde{t}_i(\mathsf{J}) = t_i(\mathsf{J})$. The bit $x_i(\mathsf{J})$ is set to be equal to 0, in keeping with the input constraint.

a) Cyclic shifting and message-interval splitting: To describe what happens to those $\mathsf{J} \in \mathcal{J}_i$ with history bit $x_i^-(\mathsf{J}) = 0$, we first recall that such message intervals form a partition of $[0, \pi_{X^-|Q}(0|q_{i-1}))$ — see Lemma V.4.

The fate of message intervals $\mathsf{J} \in \mathcal{J}_i$ with history bit $x_i^-(\mathsf{J}) = 0$ is determined by the labeling $\mathcal{L}_{q_{i-1}}$, as defined in (19). (Recall that q_{i-1} can be determined at both the encoder and the decoder from their common knowledge of q_0 and y^{i-1}.) Recall from the description of the encoder operation in Section V-A that when the history bit x_i^- for the true message interval J_i is equal to 0, then the encoder determines the next bit to be transmitted as $\mathcal{L}_{q_{i-1}}(w_i)$, where w_i is chosen uniformly at random from the interval obtained by cyclically shifting J_i within the interval $[0, \pi_{X^-|Q}(0|q_{i-1}))$ by $u_i(q_{i-1}) := u_i \cdot \pi_{X^-|Q}(0|q_{i-1})$. Since the decoder does not know the true J_i, it attempts to keep up with the encoder by applying the cyclic shifting operation to each $\mathsf{J} \in \mathcal{J}_i$. This results in a cyclically-shifted interval J^{u_i}, for each $\mathsf{J} \in \mathcal{J}_i$, with left end-point $t_i(\mathsf{J}^{u_i}) := t_i(\mathsf{J}) + u_i(q_{i-1}) \mod \pi_{X^-|Q}(0|q_{i-1})$, and right end-point $r_i(\mathsf{J}^{u_i}) := t_i(\mathsf{J}) + s_i(\mathsf{J}) + u_i(q_{i-1}) \mod \pi_{X^-|Q}(0|q_{i-1})$. If $\mathsf{J}^{u_i} \subseteq \mathcal{L}_{q_{i-1}}^{-1}(\mathsf{b})$ for some $\mathsf{b} \in \{0, 1\}$, then J^{u_i} is included in $\widetilde{\mathcal{J}}_i$, with $\tilde{\iota}(\mathsf{J}^{u_i}) = \iota(\mathsf{J})$, $\tilde{\mu}_i(\mathsf{J}^{u_i}) = \mu_i(\mathsf{J})$, $\tilde{s}_i(\mathsf{J}^{u_i}) = s_i(\mathsf{J})$, $\tilde{t}_i(\mathsf{J}^{u_i}) = t_i(\mathsf{J})$, and

$$
x_i(\mathsf{J}^{u_i}) = \begin{cases} 0 & \text{if } \mathsf{J}^{u_i} \subseteq \mathcal{L}_{q_{i-1}}^{-1}(0) \\ 1 & \text{if } \mathsf{J}^{u_i} \subseteq \mathcal{L}_{q_{i-1}}^{-1}(1). \end{cases}
$$

Note that $x_i(\mathsf{J}^{u_i})$ would have been the bit transmitted by the encoder at time i, had J been the true message interval.

A problem arises when $\mathsf{J}^{u_i} \not\subseteq \mathcal{L}_{q_{i-1}}^{-1}(\mathsf{b})$ for any $\mathsf{b} \in \{0, 1\}$, as it then straddles at least one of the boundary points, 0 and $\pi_{X, X^-|Q}(1, 0|q_{i-1})$, of the labeling $\mathcal{L}_{q_{i-1}}$. This means that the value of the bit x_i transmitted by the encoder at time i, had this J been the true message interval, is determined by the precise location of the random point w_i within J^{u_i}. While this is not a problem for the encoder, it creates an issue for the decoder as it will no longer know what to assign as the bit $x_i(\mathsf{J}^{u_i})$. We deal with this by *splitting* J^{u_i} into two or three
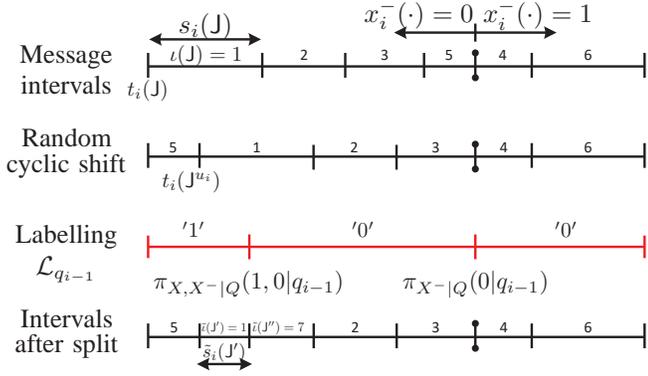
Fig. 6. Illustration of the cyclic shifting and splitting of message intervals. The lengths of message intervals are determined by their posterior probabilities, and the message intervals are positioned in ascending order of their indices, based on their history bits. Each message interval J with $x_i^-(J) = 0$ is cyclically shifted within the interval $[0, \pi_{X^-|Q}(0|q_{i-1}))$ by adding $u_i(q_{i-1})$, and its new left end-point is $t_i(J^{u_i})$. In this example, the labeling $\mathcal{L}_{q_{i-1}}$ implies that the message interval indexed with $\iota(J) = 1$ crosses the boundary point $\pi_{X,X^-|Q}(1,0|q_{i-1})$. Thus, this message interval is split into two intervals J' and J'', belonging to $\widetilde{\mathcal{J}}_i$, and indexed with $\tilde{\iota}(J') = 1$ and $\tilde{\iota}(J'') = 7$.

parts. To describe this, we first observe that if the length of J^{u_i} (which is the same as $s_i(J)$) is at most $S_{\min}$ (Eq. (20)), then any $J^{u_i} \not\subseteq \mathcal{L}_{q_{i-1}}^{-1}(b)$, $b = 0, 1$, can straddle exactly one of the boundary points of $\mathcal{L}_{q_{i-1}}$.

If J^{u_i} straddles only the boundary point 0, then we split J^{u_i} into two intervals $J' = [t_i(J^{u_i}), \pi_{X^-|Q}(0|q_{i-1}))$ and $J'' = [0, r_i(J^{u_i}))$ to be included in $\widetilde{\mathcal{J}}_i$, with $\tilde{\mu}_i(J') = \tilde{\mu}_i(J'') = \mu_i(J)$. The left end-points $\tilde{t}_i$ and lengths $\tilde{s}_i$ of J' and J'' are self-evident. Note that J' sits entirely within $\mathcal{L}_{q_{i-1}}^{-1}(0)$, so that we set $x_i(J') = 0$. By similar reasoning, we set $x_i(J'') = 1$. Finally, we set $\tilde{\iota}(J') = \iota(J)$, while J'' is assigned a brand new index: $\tilde{\iota}(J'')$ is set to be equal to the least positive integer that has not yet been assigned as an index to any message interval in $\widetilde{\mathcal{J}}_i$.

If J^{u_i} straddles only the boundary point $\pi_{X,X^-|Q}(1,0|q_{i-1})$, then we split J^{u_i} into the two intervals $J' = [t_i(J^{u_i}), \pi_{X,X^-|Q}(1,0|q_{i-1}))$ and $J'' = (\pi_{X,X^-|Q}(1,0|q_{i-1}), r_i(J^{u_i}))$ to be included in $\widetilde{\mathcal{J}}_i$. For the new message intervals J' and J'', we set $\tilde{\mu}_i$ and $\tilde{\iota}$ as above, while $x_i(J') = 1$ and $x_i(J'') = 0$. Fig. 6 illustrates the cyclic shifting and splitting operations on message intervals.

Finally, if $s_i(J) > S_{\min}$,[6] it is possible for J^{u_i} to stretch across both 0 and $\pi_{X,X^-|Q}(1,0|q_{i-1})$. In this case, we split J^{u_i} into three intervals J', J'', J''' such that each of these intervals lies entirely within one of $\mathcal{L}_{q_{i-1}}^{-1}(0)$ and $\mathcal{L}_{q_{i-1}}^{-1}(1)$. Then, for each of these new intervals, we set the bit x_i to be the $b \in \{0, 1\}$ for which the interval lies in $\mathcal{L}_{q_{i-1}}^{-1}(b)$. The left end-points $\tilde{t}_i$ and lengths $\tilde{s}_i$ are determined in the obvious manner. Finally, $\tilde{\mu}_i(J') = \tilde{\mu}_i(J'') = \tilde{\mu}_i(J''') = \mu_i(J)$, and $\tilde{\iota}(J') = \iota(J)$, while J'' and J''' get brand new indices.

Thus, each $J \in \mathcal{J}_i$ gives rise to either one interval (no

---

[6]While this case needs to be included for a complete description of the recursive construction of $\mathcal{J}_{i+1}$ from $\mathcal{J}_i$, it will play no role in the analysis of the PMS in Section VI. The analysis there is carried out under the simplifying assumption that the length of the true message interval never exceeds $S_{\min}$.

splitting) or two to three message intervals (after splitting) in $\widetilde{\mathcal{J}}_i$; we will refer to the interval(s) in $\widetilde{\mathcal{J}}_i$ as the *child(ren)* of $J \in \mathcal{J}_i$, and to J as their *parent*. Note that splitting affects only those message intervals in $\mathcal{J}_i$ that, when cyclically shifted by $u_i(q_{i-1})$, straddle the boundary points 0 and $\pi_{X,X^-|Q}(1,0|q_{i-1})$. Thus, splitting causes the number of intervals to increase by at most two: $|\mathcal{J}_i| \leq |\widetilde{\mathcal{J}}_i| \leq |\mathcal{J}_i| + 2$.

The splitting of message intervals lends itself to a simple alternative description of the encoding operation described earlier in Section V-A. Suppose that $J_i = J \in \mathcal{J}_i$. If J^{u_i} has exactly one child $\tilde{J} \in \widetilde{\mathcal{J}}_i$, then $x_i = x_i(\tilde{J})$. If J has two children J', J'' in $\widetilde{\mathcal{J}}_i$, then the encoder uses v_i to determine the transmitted bit x_i:

$$x_i = \begin{cases} x_i(J') & \text{if } 0 \leq v_i < \frac{\tilde{s}_i(J')}{s_i(J)} \\ x_i(J'') & \text{if } \frac{\tilde{s}_i(J')}{s_i(J)} \leq v_i < 1. \end{cases} \tag{23}$$

The case when J has three children is handled analogously:

$$x_i = \begin{cases} x_i(J') & \text{if } 0 \leq v_i < \frac{\tilde{s}_i(J')}{s_i(J)} \\ x_i(J'') & \text{if } \frac{\tilde{s}_i(J')}{s_i(J)} \leq v_i < \frac{\tilde{s}_i(J')+\tilde{s}_i(J'')}{s_i(J)} \\ x_i(J''') & \text{if } \frac{\tilde{s}_i(J')+\tilde{s}_i(J'')}{s_i(J)} \leq v_i < 1. \end{cases} \tag{24}$$

b) Construction of $\mathcal{J}_{i+1}$: The message intervals in $\mathcal{J}_{i+1}$ are in 1-1 correspondence with those in the set $\widetilde{\mathcal{J}}_i$. Specifically, for each $\tilde{J} \in \widetilde{\mathcal{J}}_i$, we introduce a message interval $J \in \mathcal{J}_{i+1}$, with $\iota(J) = \tilde{\iota}(\tilde{J})$, $\mu_{i+1}(J) = \tilde{\mu}_i(\tilde{J})$ and $x_{i+1}^-(J) = x_i(\tilde{J})$. We will refer to the message interval J as the *image* of $\tilde{J}$. The new interval lengths

$$s_{i+1}(J) = \Pr[J_{i+1} = J \mid Y^i = y^i, U^i = u^i], \quad \forall J \in \mathcal{J}_{i+1}, \tag{25}$$

can be determined from the set of interval lengths in $\widetilde{\mathcal{J}}_i$, as we will describe shortly. Once the interval lengths $s_{i+1}(J)$ have been determined, the left end-points can be computed as

$$t_{i+1}(J) = \pi_{X^-|Q}(0|q_i)\mathbb{1}_{\{x_{i+1}^-(J)=1\}} + \sum_{\substack{J':\ \iota(J')<\iota(J), \\ x_{i+1}^-(J')=x_{i+1}^-(J)}} s_{i+1}(J'). \tag{26}$$

To be able to describe how the lengths of intervals in $\mathcal{J}_{i+1}$ are computed from the lengths of intervals in $\widetilde{\mathcal{J}}_i$, we need to understand how the encoder decides the symbol x_i to be transmitted at time i, and how it picks the true message interval $J_{i+1} \in \mathcal{J}_{i+1}$.

c) Choice of the true message interval $J_{i+1} \in \mathcal{J}_{i+1}$: Suppose that $J_i = J \in \mathcal{J}_i$. If J has exactly one child $\tilde{J} \in \widetilde{\mathcal{J}}_i$, then J_{i+1} is taken to be the image of $\tilde{J}$ in $\mathcal{J}_{i+1}$, i.e., the message interval in $\mathcal{J}_{i+1}$ that has the same index as J (and $\tilde{J}$). On the other hand, if J has two children J', J'' in $\widetilde{\mathcal{J}}_i$, then J_{i+1} is set to be the image of J' (respectively, J'') if the transmitted bit x_i in (23) equals $x_i(J')$ (respectively, $x_i(J'')$). In any case, note that we always have $x_{i+1}^-(J_{i+1}) = x_i$. The case when J has three children is similarly handled, based on (24).

3) Computing $s_{i+1}(\mathsf{J})$, $\mathsf{J} \in \mathcal{J}_{i+1}$, *from* $\tilde{s}_i(\mathsf{J})$, $\mathsf{J} \in \widetilde{\mathcal{J}}_i$: In what follows, we use J to denote a message interval in $\widetilde{\mathcal{J}}_i$ as well as its image in $\mathcal{J}_{i+1}$. We state a preliminary lemma first.

Lemma V.2. *For any* $\mathsf{J} \in \mathcal{J}_{i+1}$, *we have* $\Pr[J_{i+1} = \mathsf{J} \mid Y^{i-1} = y^{i-1}, U^i = u^i] = \tilde{s}_i(\mathsf{J})$.

Proof: Let $\mathsf{J}^\# \in \mathcal{J}_i$ be the parent of J, so that $J_{i+1} = \mathsf{J}$ implies that $J_i = \mathsf{J}^\#$. Then,

$$\Pr[J_{i+1} = \mathsf{J} \mid Y^{i-1} = y^{i-1}, U^i = u^i]$$
$$= \Pr[J_{i+1} = \mathsf{J}, J_i = \mathsf{J}^\# \mid Y^{i-1} = y^{i-1}, U^i = u^i]$$
$$= \Pr[J_i = \mathsf{J}^\# \mid Y^{i-1} = y^{i-1}, U^{i-1} = u^{i-1}]$$
$$\cdot \Pr[J_{i+1} = \mathsf{J} \mid J_i = \mathsf{J}^\#, U_i = u_i]$$
$$= s_i(\mathsf{J}^\#) \frac{\tilde{s}_i(\mathsf{J})}{s_i(\mathsf{J}^\#)},$$

where $\Pr[J_{i+1} = \mathsf{J} \mid J_i = \mathsf{J}^\#, U_i = u_i] = \frac{\tilde{s}_i(\mathsf{J})}{s_i(\mathsf{J}^\#)}$ follows from the way that J_{i+1} is chosen, given J_i and u_i. ∎

We are now in a position to derive the means of computing $s_{i+1}(\mathsf{J})$ from $\tilde{s}_i(\mathsf{J})$.

Lemma V.3. *For each* $\mathsf{J} \in \mathcal{J}_{i+1}$, *we have*

$$s_{i+1}(\mathsf{J}) := \Pr[J_{i+1} = \mathsf{J} \mid Y^i = y^i, U^i = u^i]$$
$$= \tilde{s}_i(\mathsf{J}) \frac{p_{Y|X}(y_i \mid x_i(\mathsf{J}))}{\pi_{Y|Q}(y_i \mid q_{i-1})},$$

where $q_{i-1} = \Phi(q_0, y^{i-1})$.

Proof: We start with

$$\Pr[J_{i+1} = \mathsf{J} \mid Y^i = y^i, U^i = u^i]$$
$$= \frac{\Pr[J_{i+1} = \mathsf{J}, Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]}{\Pr[Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]}$$
$$\overset{(a)}{=} \tilde{s}_i(\mathsf{J}) \frac{\Pr[Y_i = y_i \mid J_{i+1} = \mathsf{J}, Y^{i-1} = y^{i-1}, U^i = u^i]}{\Pr[Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]}$$
$$\overset{(b)}{=} \tilde{s}_i(\mathsf{J}) \frac{\Pr[Y_i = y_i \mid X_i = x_i(\mathsf{J})]}{\Pr[Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]},$$

where (a) is by Lemma V.2, and (b) is due to the fact that if $J_{i+1} = \mathsf{J}$, then the bit transmitted at time instant i must have been $x_{i+1}^-(\mathsf{J}) = x_i(\mathsf{J})$.

The denominator on the right-hand-side above can be expressed as

$$\Pr[Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i] \quad (27)$$
$$= \sum_{\tilde{\mathsf{J}} \in \mathcal{J}_{i+1}} \Pr[J_{i+1} = \tilde{\mathsf{J}}, Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]$$
$$= \sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i} \tilde{s}_i(\tilde{\mathsf{J}}) \Pr[Y_i = y_i \mid X_i = x_i(\tilde{\mathsf{J}})] \quad (28)$$

Now, the last expression above can be written as

$$\left(\sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i: \, x_i(\tilde{\mathsf{J}})=1} \tilde{s}_i(\tilde{\mathsf{J}}) \right) \Pr[Y_i = y_i \mid X_i = 1]$$
$$+ \left(\sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i: \, x_i(\tilde{\mathsf{J}})=0} \tilde{s}_i(\mathsf{J}) \right) \Pr[Y_i = y_i \mid X_i = 0].$$

By construction of $\widetilde{\mathcal{J}}_i$ from $\mathcal{J}_i$, the message intervals $\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i$ with $x_i(\tilde{\mathsf{J}}) = 1$ form a partition of $[0, \pi_{X,X^-|Q}(1|0, q_{i-1}))$, so that

$$\sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i: \, x_i(\tilde{\mathsf{J}})=1} \tilde{s}_i(\tilde{\mathsf{J}}) = \pi_{X,X^-|Q}(1, 0|q_{i-1})$$

and

$$\sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i: \, x_i(\tilde{\mathsf{J}})=0} \tilde{s}_i(\tilde{\mathsf{J}}) = 1 - \sum_{\tilde{\mathsf{J}} \in \widetilde{\mathcal{J}}_i: \, x_i(\tilde{\mathsf{J}})=1} \tilde{s}_i(\tilde{\mathsf{J}})$$
$$= 1 - \pi_{X,X^-|Q}(1, 0|q_{i-1}).$$

Thus, (28) simplifies to

$$\Pr[Y_i = y_i \mid Y^{i-1} = y^{i-1}, U^i = u^i]$$
$$= \pi_{X,X^-|Q}(1, 0|q_{i-1}) \, p_{Y|X}(y_i|1)$$
$$+ \left(1 - \pi_{X,X^-|Q}(1, 0|q_{i-1}) \right) p_{Y|X}(y_i|0)$$
$$= \sum_{x,x^-} \pi_{X,X^-|Q}(x, x^-|q_{i-1}) \, p_{Y|X}(y_i|x),$$

recalling that $\pi_{X,X^-|Q}(x, x^-|q_{i-1}) = 0$ for $(x, x^-) = (1, 1)$. ∎

One simple consequence of Lemma V.3 is that for each $\mathsf{J} \in \mathcal{J}_{i+1}$, we have

$$s_{i+1}(\mathsf{J}) \leq s_i(\mathsf{J}^\#) \max_{x,y,q} \frac{p_{Y|X}(y|x)}{\pi_{Y|Q}(y|q)},$$

where $\mathsf{J}^\# \in \mathcal{J}_i$ is the parent of J. Recursively applying this inequality, we obtain

$$s_{i+1}(\mathsf{J}) \leq 2^{-nR} \left(\max_{x,y,q} \frac{p_{Y|X}(y|x)}{\pi_{Y|Q}(y|q)} \right)^i, \quad (29)$$

which is a crude, but useful, upper bound on interval lengths in $\mathcal{J}_{i+1}$.

A second consequence of Lemma V.3 is the fact, crucial for our description of cyclic shifting and message-interval splitting, that the message intervals $\mathsf{J} \in \mathcal{J}_i$ with history bit $x_i^-(\mathsf{J}) = 0$ form a partition of $[0, \pi_{X^-|Q}(0|q_{i-1}))$. This follows from the next lemma.

Lemma V.4. *For any* $i \geq 1$, *we have*

$$\sum_{\mathsf{J} \in \mathcal{J}_i: \, x_i^-(\mathsf{J})=0} s_i(\mathsf{J}) = \pi_{X^-|Q}(0 \mid q_{i-1}).$$

Proof: The proof is by induction on i. By construction, the statement is true for $i = 1$. So, suppose that it holds for some $i \geq 1$. We then consider $\sum_{\mathsf{J} \in \mathcal{J}_{i+1}: \, x_{i+1}^-(\mathsf{J})=0} s_{i+1}(\mathsf{J})$. By Lemma V.3, we have

$$\sum_{\mathsf{J} \in \mathcal{J}_{i+1}: \, x_{i+1}^-(\mathsf{J})=0} s_{i+1}(\mathsf{J})$$
$$= \sum_{\mathsf{J} \in \widetilde{\mathcal{J}}_i: \, x_i(\mathsf{J})=0} \tilde{s}_i(\mathsf{J}) \frac{p_{Y|X}(y_i \mid x_i(\mathsf{J}))}{\pi_{Y|Q}(y_i \mid q_{i-1})}$$
$$= \frac{p_{Y|X}(y_i \mid 0)}{\pi_{Y|Q}(y_i \mid q_{i-1})} \left(\sum_{\mathsf{J} \in \widetilde{\mathcal{J}}_i: \, x_i(\mathsf{J})=0} \tilde{s}_i(\mathsf{J}) \right)$$

$$\overset{(a)}{=} \frac{p_{Y|X}(y_i \mid 0)}{\pi_{Y|Q}(y_i \mid q_{i-1})} \left(1 - \pi_{X,X^-|Q}(1,0 \mid q_{i-1})\right)$$

$$= \frac{\sum_{x^-} p_{Y|X}(y_i \mid 0)\, \pi_{X,X^-|Q}(0, x^- \mid q_{i-1})}{\sum_{x,x^-} p_{Y|X}(y_i|x)\, \pi_{X,X^-|Q}(x, x^- \mid q_{i-1})}$$

$$= \frac{\sum_{x^-} p_{Y|X}(y_i \mid 0)\, p^*_{X|X^-,Q}(0 \mid x^-, q_{i-1})\, \pi_{X^-|Q}(x^-|q_{i-1})}{\sum_{x,x^-} p_{Y|X}(y_i|x)\, p^*_{X|X^-,Q}(x|x^-, q_{i-1})\, \pi_{X^-|Q}(x^-|q_{i-1})}$$

$$\overset{(b)}{=} \pi_{X^-|Q}(0 \mid q_i),$$

where (b) can be verified directly from the known expressions for $p_{Y|X}$, $p^*_{X|X^-,Q}$ and $\pi_{X^-|Q}$. The induction hypothesis has been used in (a) above to validate the construction of $\widetilde{\mathcal{J}}_i$ from $\mathcal{J}_i$ via message splitting, which yields the fact that $\sum_{\mathsf{J} \in \widetilde{\mathcal{J}}_i : x_i(\mathsf{J})=0} \tilde{s}_i(\mathsf{J}) = 1 - \pi_{X,X^-|Q}(1,0 \mid q_{i-1})$. ∎

Remark 3. The PMS construction described in this section and its analysis to come in the next section are also valid for non-optimal input distributions and for a broader class of channels called unifilar finite state channels (with feedback) [11], provided that one condition is met. This condition, which may easily go unnoticed within the details of the PMS construction, is that of Lemma V.4: the sum of the lengths of all the messages intervals with the same history bit x_i^- is a function of the Q-state q_{i-1} only. This condition is essentially equivalent to the BCJR-invariant property that was introduced in [26, Section III.A]. We emphasize that this property is immediately satisfied when an input distribution satisfies the Bellman equation in the corresponding DP problem. However, this property can also be verified directly, as has been done in Lemma V.4.

VI. PMS Analysis: Proof of Theorem V.1

The statement to be proved concerns the probability of $s_{n-\zeta_n}(J_{n-\zeta_n}) < \xi$, conditioned on the occurrence of the event $s_i(J_i) \le S_{\min}$, $i = 1, 2, \ldots, n - \zeta_n - 1$. Thus, throughout this section, we assume that $s_i(J_i) \le S_{\min}$, $i = 1, 2, \ldots, n - \zeta_n - 1$, holds. All probabilities and expectations in this section are implicitly conditioned on this event. This results in a simplified analysis of the PMS in Phase I, since, under this assumption, the true message interval J_i cannot split into more than two children at any point of the PMS, as is evident from the description of message-interval splitting in Section V-D.

Since this proof is concerned only with the sequence of true message intervals J_i, $i = 1, 2, 3, \ldots$, we will use some simplified notation: $S_{i+1} = s_{i+1}(J_{i+1})$, $\tilde{S}_i = \tilde{s}_i(J_{i+1})$, $X_i = x_i(J_i)$, $X_i^- = x_i^-(J_i) = x_{i-1}(J_{i-1})$.

A. Preliminaries

Define for $\rho \in [0, 1)$,

$$\phi_{q,x,q^+}(\rho) \triangleq \left(\frac{p_{Y|X}(y|x)}{\pi_{Y|Q}(y|q)} \right)^{-\rho}, \tag{30}$$

where y is the unique solution to the equation $q^+ = g(q, y)$. We also use $\phi_{(x,q)_{i-1}^i}(\rho)$ as a shorthand for $\phi_{q_{i-1}, x_i, q_i}(\rho)$. Throughout the analysis, it is assumed that $p_{Y|X}(y|x) > 0$,

otherwise, define $\phi_{q,x,q^+}(\rho) = 0$ and the derivations can be easily repeated.

Define also

$$\psi^s_{x^-,q,x,q^+}(\rho)$$
$$\triangleq \mathbb{E}\left[\left(\frac{S_{i+1}}{S_i} \right)^{-\rho} \middle| (S_i, X_i^-, Q_{i-1}, X_i, Q_i) = (s, x^-, q, x, q^+) \right].$$

Indeed, there ought also be a time index in the notation ψ, but it will be shown in the proof of Lemma VI.1 below that the expected value does not depend on the time index i.

B. Analysis

The following lemma comprises the core of our PMS analysis:

Lemma VI.1. *For all $\delta > 0$, there exists $s^*(\delta)$ such that*

$$\psi^s_{x^-,q,x,q^+}(\rho) \le \phi_{q,x,q^+}(\rho) 2^\delta,$$

for all $s \le s^*(\delta)$, $0 \le \rho < 1$, and all (x^-, q, x, q^+).

Proof: In this proof we show that $\psi^s_{x^-,q,x,q^+}(\rho)$ can be made arbitrarily close to $\phi_{q,x,q^+}(\rho)$ if we take s to be small enough. From Lemma V.3, we have

$$\psi^s_{x^-,q,x,q^+}(\rho)$$
$$= \mathbb{E}\left[\left(\frac{S_{i+1}}{S_i} \right)^{-\rho} \middle| (S_i, X_i^-, Q_{i-1}, X_i, Q_i) = (s, x^-, q, x, q^+) \right]$$
$$= \left(\frac{p_{Y|X}(y|x)}{\pi_{Y|Q}(y|q)} \right)^{-\rho}$$
$$\cdot \mathbb{E}\left[\left(\frac{\tilde{S}_i}{S_i} \right)^{-\rho} \middle| (S_i, X_i^-, Q_{i-1}, X_i, Q_i) = (s, x^-, q, x, q^+) \right]$$
$$= \phi_{q,x,q^+}(\rho)$$
$$\cdot \mathbb{E}\left[\left(\frac{\tilde{S}_i}{S_i} \right)^{-\rho} \middle| (S_i, X_i^-, Q_{i-1}, X_i, Q_i) = (s, x^-, q, x, q^+) \right], \tag{31}$$

where in the second equality above, y is the unique solution of $q^+ = g(q, y)$. Our interest is in showing an upper bound on the expected value in (31). The simpler case is when $X_i^- = 1$, since there is no split (i.e., $\tilde{S}_i = S_i$), so that $\psi^s_{1,q,x,q^+}(\rho) = \phi_{q,x,q^+}(\rho)$.

To deal with the other case, $X_i^- = 0$, we need the conditional probability density function $f_{\tau_i, V_i | S_i, Q_{i-1}, X_i^-, X_i, Q_i}(u, v | s, q, 0, x, q^+)$, where $\tau_i \triangleq t_i(J_i) + U_i(q) \bmod \pi_{X^-|Q}(0|q)$ denotes the left end-point of the message interval J_{i-1} after cyclic shifting by $U_i(q)$. Observe that

$$f_{\tau_i, V_i | S_i, Q_{i-1}, X_i^-, X_i, Q_i}(u, v \mid s, q, 0, x, q^+)$$
$$\overset{(a)}{=} \frac{\Pr[X_i = x \mid S_i = s, X_i^- = 0, Q_{i-1} = q, \tau_i = u, V_i = v]}{\Pr[X_i = x \mid S_i = s, X_i^- = 0, Q_{i-1} = q]}$$
$$\cdot f_{U_i(q), V_i}(u, v)$$
$$\overset{(b)}{=} \frac{\Pr[X_i = x \mid S_i = s, X_i^- = 0, Q_{i-1} = q, \tau_i = u, V_i = v]}{\Pr[X_i = x \mid X_i^- = 0, Q_{i-1} = q]}$$

$$\cdot f_{U_i(q),V_i}(u,v)$$

$$\overset{(c)}{=} \frac{\Pr[X_i = x \mid S_i = s, X_i^- = 0, Q_{i-1} = q, \tau_i = u, V_i = v]}{P_{x,0,q}}$$

$$\cdot f_{U,V}\left(\frac{u}{\pi_{X^-|Q}(0|q)}, v\right), \tag{32}$$

where (a) follows from the Markov chain $Q_i - (X_i, Q_{i-1}) - (S_i, X_i^-, \tau_i, V_i)$ and the fact that τ_i is distributed uniformly on $[0, \pi_{X^-|Q}(0|q)]$, (b) follows from the Markov chain $X_i - (X_i^-, Q_{i-1}) - S_i$ and (c) is due to replacement of the random variable $U_i(q)$ with U_i, and the notation $P_{x,0,q} \triangleq \pi_{X,X^-|Q}(x, 0|q)$. The Markov chain $Q_i - (X_i, Q_{i-1}) - (S_i, X_i^-, \tau_i, V_i)$ follows from the fact that Q_i is a function of (Y_i, Q_{i-1}) and the memoryless property. The second Markov chain $X_i - (X_i^-, Q_{i-1}) - S_i$ is shown in two steps: first, if $X_i^- = 1$, then $X_i = 0$. For the other case, $X_i^- = 0$, note that W_i, given $X_i^- = 0$ and $Q_{i-1} = q$, is distributed uniformly on $[0, \pi_{X^-|Q}(0|q)]$, and X_i is a function of W_i.

Note that the numerator of (32) is an indicator function of the event that the point $(u + sv) \bmod \pi_{X^-|Q}(0|q)$ is mapped to $X = x$. Also, we have $f_{U,V}\left(\frac{u}{\pi_{X^-|Q}(0|q)}, v\right) = 1$ for all $(u,v) \in [0, \pi_{X^-|Q}(0|q)] \times [0,1]$. Thus, the density $f_{\tau_i, V_i | S_i, Q_{i-1}, X_i^-, X_i, Q_i}$ does not depend on the time index i. Since the expected value in (31) is determined from this density function, the time index can be omitted from the notation of $\psi^s_{x^-, q, x, q^+}(\rho)$.

We begin with the calculation of the expected value in (31) for the case $X_i = 0$:

$$\mathbb{E}[\tilde{S}_i^{-\rho}/S_i^{-\rho} \mid S_i = s, X_i^- = 0, Q_{i-1} = q, X_i = 0, Q_i = q^+]$$

$$\overset{(a)}{=} s^\rho \int_{[0,\pi_{X^-|Q}(0|q)]} \int_{[0,1]}$$

$$f_{\tau_i, V_i | S_i, X_i^-, Q_{i-1}, X_i, Q_i}(u, v \mid s, 0, q, 0, q^+)\tilde{s}(q, u, s, 0)^{-\rho} dv\,du$$

$$\overset{(b)}{=} \frac{s^\rho}{P_{0,0,q}} \int_{I_1 \cup I_2 \cup I_3} \int_{I_{V(u,s)}} \tilde{s}(q, u, s, 0)^{-\rho} dv\,du, \tag{33}$$

where

(a) follows by defining $\tilde{s}(q, u, s, 0)$ to be the length of the new true message interval after a possible split, which is a function of $(Q_{i-1} = q, \tau_i = u, S_i = s, X_i = 0)$ only;

(b) follows from substituting (32) and restricting the integration over u, v to be on the domains where $\Pr[X_i = 0 \mid S_i = s, X_i^- = 0, Q_{i-1} = q, \tau_i = u, V_i = v] = 1$, and are illustrated in Fig. 7. The u-intervals are given by

$$I_1 \triangleq \left[\pi_{X,X^-|Q}(1, 0|q), \pi_{X^-|Q}(0|q) - s\right]$$
$$I_2 \triangleq \left[\pi_{X,X^-|Q}(1, 0|q) - s, \pi_{X,X^-|Q}(1, 0|q)\right]$$
$$I_3 \triangleq \left[\pi_{X^-|Q}(0|q) - s, \pi_{X^-|Q}(0|q)\right], \tag{34}$$

and the corresponding v-intervals that depend on u, s are:

$$I_{V(u,s)} \triangleq \begin{cases} [0,1] & \text{for } u \in I_1 \\ \left[\frac{-u + P_{0,0,q}}{s}, 1\right] & \text{for } u \in I_2 \\ \left[0, \frac{-u + P_{0,0,q} + P_{1,0,q}}{s}\right] & \text{for } u \in I_3. \end{cases} \tag{35}$$
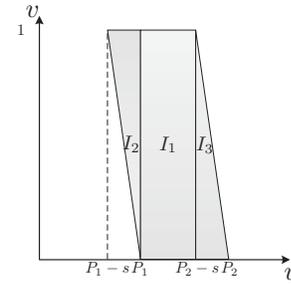
Fig. 7. Illustration of the intervals in (35). We use $P_1 \triangleq P_{1,0,q}$, $P_2 \triangleq P_{0,0,q} + P_{1,0,q}$. The shaded area corresponds to values of u and v for which $\Pr[X_i = 0 \mid S_i = s, X_i^- = 0, Q_{i-1} = q, \tau_i = u, V_i = v] = 1$.

For all $u \in I_1$ and all $v \in [0, 1]$, $\tilde{s}(q, u, s, 0) = s$, so

$$\frac{s^\rho}{P_{0,0,q}} \int_{I_1} \int_{[0,1]} \tilde{s}(q, u, s, 0)^{-\rho} dv\,du = \frac{P_{0,0,q} - s}{P_{0,0,q}}.$$

For the interval I_2,

$$\frac{s^\rho}{P_{0,0,q}} \int_{I_2} \int_{I_{V(u,s)}} \tilde{s}(q, u, s, 0)^{-\rho} dv\,du$$

$$\overset{(a)}{=} \frac{s^\rho}{P_{0,0,q}} \frac{1}{s} \int_{I_2} (u - P_{0,0,q} + s)(u - P_{0,0,q} + s)^{-\rho} du$$

$$= \frac{s^\rho}{P_{0,0,q}} \frac{1}{s} \int_{[0,s]} u^{-\rho+1} du$$

$$= \frac{1}{P_{0,0,q}} \frac{s}{(-\rho + 2)},$$

where (a) follows from $\tilde{s}(q, u, s, 0) = u - P_{0,0,q} + s$ and $|I_{V(u,s)}| = (u - P_{0,0,q} + s)$ for all $u \in I_2$. The calculation for the third interval, I_3, is similar to that for I_2, and results in the same value for the integral.

To conclude, we have shown that

$$\psi^s_{0,q,x,q^+}(\rho) = \phi_{q,x,q^+}(\rho)\left[\frac{P_{x,0,q} - s}{P_{x,0,q}} + 2\frac{1}{P_{x,0,q}}\frac{s}{(-\rho + 2)}\right]$$

$$\triangleq \phi_{q,x,q^+}(\rho)2^{\delta_{q,x}(s)}$$

for $\rho \in [0, 1)$. The last step is to define $\delta(s) \triangleq \max_{q,x} \delta_{q,x}(s)$, which goes to zero when $s \to 0$. Now, it is clear that $\psi^s_{x^-,q,x,q^+}(\rho) \leq \phi_{q,x,q^+}(\rho)2^{\delta(s)}$ for all (x^-, q, x, q^+), $0 \leq \rho < 1$ and $s \leq S_{\min}$, as required. ∎

Proof of Theorem V.1: Throughout this proof, ζ_n is a sequence of integers that satisfies $\frac{\zeta_n}{n} \to 0$. Our aim is to show that for $|\mathcal{M}| = 2^{nR}$, there exists $\xi > 0$ such that $\Pr[S_{n-\zeta_n} \leq \xi] \to 0$ as n increases. For convenience, we replace the variable n by $n + \zeta_n$, so that the analysis is made for a message set of size $|\mathcal{M}| = 2^{(n+\zeta_n)R}$ and the probability analysis is for $\Pr[S_n \leq \xi]$.

From Lemma VI.1, for all $\delta > 0$, there exists $s^*(\delta)$, such that for all $s \leq s^*(\delta)$

$$\psi^s_{x^-,q,x,q^+}(\rho) \leq \phi_{q,x,q^+}(\rho)2^\delta. \tag{36}$$

We will utilize (36) to provide a vanishing upper bound on $\mathbb{E}[\Lambda(S_n)]$, with $\Lambda(s) = s^{-\rho}$ for a judiciously chosen $\rho > 0$. This, by the Markov inequality, will imply that the probability that $\Lambda(S_n)$ is above a certain threshold is vanishingly small. Since $\Lambda(s)$ is a decreasing function of s, we then obtain

that the probability that S_n lies below a certain threshold is vanishingly small, as desired.

We introduce some convenient notation for the upcoming analysis. Set $\Delta_i = (X_i, Q_i, S_i)$, and let $(X, Q)_i^j$ stand for (X_i^j, Q_i^j) when $i < j$. Consider the following chain of inequalities:

$$\mathbb{E}_{\Delta_n}[\Lambda(S_n)]$$

$$\overset{(a)}{=} \mathbb{E}_{\Delta_{n-1}}[\Lambda(S_{n-1})\,\mathbb{E}_{\Delta_n|\Delta_{n-1}}[\Lambda(S_n/S_{n-1})|\Delta_{n-1}]]$$

$$= \mathbb{E}_{\Delta_{n-1}}[\Lambda(S_{n-1})\,\mathbb{E}_{(X_n,Q_n)|\Delta_{n-1}}[$$
$$\mathbb{E}_{S_n|(X_n,Q_n),\Delta_{n-1}}[\Lambda(S_n/S_{n-1})|(X_n,Q_n),\Delta_{n-1}]]]$$

$$\overset{(b)}{\leq} 2^{\delta}\,\mathbb{E}_{\Delta_{n-1}}[\Lambda(S_{n-1})\,\mathbb{E}_{(X_n,Q_n)|\Delta_{n-1}}[\phi_{(X,Q)_{n-1}^n}(\rho)]]$$

$$= 2^{\delta}\,\mathbb{E}_{(X,Q)_{n-1}^n}[\phi_{(X,Q)_{n-1}^n}(\rho)$$
$$\cdot\,\mathbb{E}_{S_{n-1}|(X,Q)_{n-1}^n}[\Lambda(S_{n-1})|(X,Q)_{n-1}^n]]$$

$$\overset{(a)}{=} 2^{\delta}\,\mathbb{E}_{(X,Q)_{n-1}^n}[\phi_{(X,Q)_{n-1}^n}(\rho)\,\mathbb{E}_{\Delta_{n-2}|(X,Q)_{n-1}^n}[\Lambda(S_{n-2})$$
$$\cdot\,\mathbb{E}_{S_{n-1}|(X,Q)_{n-1}^n,\Delta_{n-2}}[\Lambda(S_{n-1}/S_{n-2})|(X,Q)_{n-1}^n,\Delta_{n-2}]]]$$

$$\overset{(c)}{=} 2^{\delta}\,\mathbb{E}_{(X,Q)_{n-1}^n}[\phi_{(X,Q)_{n-1}^n}(\rho)\,\mathbb{E}_{\Delta_{n-2}|(X,Q)_{n-1}^n}[\Lambda(S_{n-2})$$
$$\cdot\,\mathbb{E}_{S_{n-1}|(X_{n-1},Q_{n-1}),\Delta_{n-2}}[\Lambda(S_{n-1}/S_{n-2})|(X_{n-1},Q_{n-1}),\Delta_{n-2}]]]$$

$$\overset{(b)}{\leq} 2^{2\delta}\,\mathbb{E}_{(X,Q)_{n-1}^n}[\phi_{(X,Q)_{n-1}^n}(\rho)\,\mathbb{E}_{\Delta_{n-2}|(X,Q)_{n-1}^n}[$$
$$\Lambda(S_{n-2})\phi_{(X,Q)_{n-2}^{n-1}}(\rho)]]$$

$$= 2^{2\delta}\,\mathbb{E}_{(X,Q)_{n-2}^n}[\phi_{(X,Q)_{n-1}^n}(\rho)\phi_{(X,Q)_{n-2}^{n-1}}(\rho)$$
$$\cdot\,\mathbb{E}_{S_{n-2}|(X,Q)_{n-2}^n}[\Lambda(S_{n-2})|(X,Q)_{n-2}^n]]$$

$$\overset{(d)}{\leq} 2^{n\delta}\,\mathbb{E}_{(X,Q)_1^n}\left[\prod_{i=1}^{n}\phi_{(X,Q)_{i-1}^i}(\rho)\mathbb{E}_{S_0|(X,Q)_1^n}[\Lambda(S_0)|(X,Q)_1^n]\right]$$

$$\overset{(e)}{=} 2^{n\delta}2^{(n+\zeta_n)R\rho}\,\mathbb{E}_{(X,Q)_1^n}\left[\prod_{i=1}^{n}\phi_{(X,Q)_{i-1}^i}(\rho)\right],$$

where:

(a) follows from the law of total expectation;
(b) follows from (36) since $\psi_{X_{n-1},Q_{n-1},X_n,Q_n}^{S_{n-1}}(\rho) = \mathbb{E}_{S_n|(X_n,Q_n),\Delta_{n-1}}[\Lambda(S_n/S_{n-1})|(X_n,Q_n),\Delta_{n-1}]$;
(c) follows from the Markov chain $S_i - (\Delta_{i-1},(X_i,Q_i)) - (X,Q)_{i+1}^n$ for all i, and specifically for $i = n-1$;[7]
(d) follows from applying the above steps $n-2$ times; and
(e) follows from the fact that $|\mathcal{M}| = 2^{(n+\zeta_n)R}$.

The expectation above can be decomposed into non-typical and typical sequences with respect to the Markov distribution $p(q^+, x|q, x^-) := \sum_y \mathbb{1}\{q^+ = g(q,y)\}p_{Y|X}(y|x)p_{X|X^-,Q}^*(x|x^-,q)$. With some abuse of notation, since q and q^+ determine a unique y such that $q^+ = g(q,y)$, we refer to $\phi_{(X,Q)_{i-1}^i}(\rho)$ as $\phi_{Q_{i-1},X_i,Y_i}(\rho)$. Consider

$$\mathbb{E}_{\Delta_n}[\Lambda(S_n)]$$
$$\overset{(a)}{\leq} 2^{n(R\rho(1+\frac{\zeta_n}{n})+\delta)}$$

$$\cdot\left[\epsilon_n[\max_{q,x,y}\phi_{q,x,y}(\rho)]^n + \prod_{q,x,y}\phi_{q,x,y}(\rho)^{n\pi_{Q,X,Y}(q,x,y)+\kappa_n}\right]$$

$$= 2^{n(R\rho+R\rho\frac{\zeta_n}{n}+\delta+K\rho)}\epsilon_n$$
$$+ 2^{n(R\rho+R\rho\frac{\zeta_n}{n}+\delta)}\prod_{q,x,y}2^{(n\pi_{Q,X,Y}(q,x,y)+\kappa_n)\log\phi_{q,x,y}(\rho)}$$

$$= 2^{n(R\rho+R\rho\frac{\zeta_n}{n}+\delta+K\rho)}\epsilon_n$$
$$+ 2^{n(R\rho+R\rho\frac{\zeta_n}{n}+\delta)}2^{-\rho\sum_{q,x,y}(n\pi_{Q,X,Y}(q,x,y)+\kappa_n)\log\left(\frac{p_{Y|X}(y|x)}{\pi_{Y|Q}(y|q)}\right)}$$

$$\overset{(b)}{\leq} 2^{n(R\rho+R\rho\frac{\zeta_n}{n}+\delta+K\rho)}\epsilon_n + 2^{-n\rho(I(X;Y|Q)-R(1+\frac{\zeta_n}{n})-\frac{\delta}{\rho}+\frac{\kappa'_n}{n})}, \tag{37}$$

where (a) follows from separating the contributions made to the expected value by non-typical and typical sequences: we let ϵ_n denote the probability that a sequence is not in the typical set, K stands for $\max_{q,x,y}\log\left(\frac{\pi_{Y|Q}(y|q)}{p_{Y|X}(y|x)}\right)$ and, finally, κ_n denotes the maximum deviation of the empirical distribution from $\pi_{Q,X,Y}(q,x,y)$ for a typical sequence. Item (b) follows from the notation $\kappa'_n \triangleq \kappa_n|\mathcal{Q}||\mathcal{X}||\mathcal{Y}|K$. Now, since ϵ_n decreases exponentially with n, there exists a choice of (ρ^*, δ^*) such that $2^{n(R\rho^*+\frac{\zeta_n}{n}+\delta^*+K\rho^*)}\epsilon_n$ is arbitrarily small, while R can be made arbitrarily close to $I(X;Y|Q)$.

Finally, the main result can be derived with δ^* and ρ^*:

$$\Pr[S_n \leq s^*(\delta^*)] \overset{(a)}{=} \Pr[\Lambda(S_n) \geq \Lambda(s^*(\delta^*))]$$
$$\overset{(b)}{\leq} \frac{\mathbb{E}[\Lambda(S_n)]}{\Lambda(s^*(\delta^*))}$$
$$\overset{(c)}{\to} 0, \tag{38}$$

where (a) follows from the fact that $\Lambda(\cdot)$ is a decreasing function, (b) follows from Markov's inequality and (c) follows from (37). ∎

VII. DP FORMULATION AND SOLUTION

This section covers the formulation of feedback capacity as DP and its solution. The solution of the DP problem implies almost immediately the derivations of feedback capacity and optimal input distribution, which were stated earlier as separate results in Theorems III.1 and IV.2 (which are proved at the end of this section). We begin with presenting the family of DP problems termed infinite-horizon with average reward.

A. Average reward DP

Each DP is defined by the tuple $(\mathcal{Z}, \mathcal{U}, \mathcal{W}, F, P_Z, P_W, g)$. We consider a discrete-time dynamical system evolving according to:

$$z_t = F(z_{t-1}, u_t, w_t), \; t = 1, 2, \ldots \tag{39}$$

Each state, z_t, takes values in a Borel space $\mathcal{Z}$, each action, u_t, takes values in a compact subset $\mathcal{U}$ of a Borel space, and each disturbance, w_t, takes values in a measurable space $\mathcal{W}$. The initial state, z_0, is drawn from the distribution P_Z, and the disturbance, w_t, is drawn from $P_{W|Z_{t-1},U_t}$. The history, $h_t = (z_0, w_1, \ldots, w_{t-1})$, summarizes all the information available to the controller at time t. The controller at time t chooses

the action, u_t, by a function μ_t that maps histories to actions, i.e., $u_t = \mu_t(h_t)$. The collection of these functions is called a policy and is denoted as $\pi = \{\mu_1, \mu_2, \dots\}$. Note that given a policy, π, and the history, h_t, one can compute the actions vector, u^t, and the states of the system, $z_1, z_2, \dots, z_{t-1}$.

Our objective is to maximize the average reward given a bounded reward function $r : \mathcal{Z} \times \mathcal{U} \to \mathbb{R}$. The average reward for a given policy π is given by:

$$\rho_\pi = \liminf_{N \to \infty} \frac{1}{N} \mathbb{E}_\pi \left[\sum_{t=1}^{N} r(Z_{t-1}, \mu_t(h_t)) \right],$$

where the subscript indicates that actions u_t are subject to the policy π. The optimal average reward is defined as

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

Having defined the DP problem, we are ready to show the formulation of feedback capacity as DP.

B. Formulation of capacity as DP

The state of the DP, z_{t-1}, is defined as the conditioned probability vector $\beta_{t-1}(x_{t-1}) \triangleq p(x_{t-1}|y^{t-1})$. The action space, $\mathcal{U}$, is the set of stochastic matrices, $p_{X_t|X_{t-1}}$, such that $p_{X_t|X_{t-1}}(1|1) = 0$. For a given policy and an initial state, the encoder at time $t-1$ can calculate the state, β_{t-1}, since the tuple y^{t-1} is available from the feedback. The disturbance is taken to be the channel output, $w_t = y_t$, and the reward gained at time $t-1$ is chosen as $I(Y_t; X_t|y^{t-1})$. These definitions imply that the optimal reward of this DP is equal to the feedback capacity given in Theorem IV.1.

It can also be shown that the DP states satisfy the following recursive relation,

$$\begin{aligned} \beta_t(x_t) &= p(x_t|y^t) \\ &= \frac{\sum_{x_{t-1}} \beta_{t-1}(x_{t-1}) u_t(x_t, x_{t-1}) p_{Y|X}(y_t|x_t)}{\sum_{x_t, x_{t-1}} \beta_{t-1}(x_{t-1}) u_t(x_t, x_{t-1}) p_{Y|X}(y_t|x_t)}, \end{aligned}$$
(40)

where $u_t(x_t, x_{t-1})$ corresponds to $p(x_t|x_{t-1}, y^{t-1})$, the dependence on y^{t-1} being left out of the notation for u_t. In [13], this formulation was shown to satisfy the Markov nature required in DP problems and it was also shown that the optimal average reward is exactly the capacity expression in Theorem IV.1. Note that this formulation is valid for any memoryless channel with our input constraint; moreover, minor variations can also yield a similar formulation with different input constraints.

C. The DP for the BIBO channel

Here, each element in the formulation above will be calculated for the BIBO channel; the DP state at time $t - 1$, z_{t-1}, is the probability vector $[p_{X_{t-1}, Y^{t-1}}(0|y^{t-1}), p_{X_{t-1}|Y^{t-1}}(1|y^{t-1})]$. Since the components of this vector sum to 1, the notation can be abused as $z_{t-1} \triangleq p_{X_{t-1}|Y^{t-1}}(0|y^{t-1})$, i.e., the first

TABLE I
THE CONDITIONAL DISTRIBUTION $p(y_t, x_t, x_{t-1}|z_{t-1}, u_t)$

x_{t-1}	x_t	$y_t = 0$	$y_t = 1$
0	0	$\bar{\alpha} z_{t-1} u_t(1,1)$	$\alpha z_{t-1} u_t(1,1)$
0	1	$\bar{\beta} z_{t-1} u_t(1,2)$	$\beta z_{t-1} u_t(1,2)$
1	0	$\bar{\alpha}(1 - z_{t-1})$	$\alpha(1 - z_{t-1})$

component will be the DP state. Each action, u_t, is a constrained 2×2 stochastic matrix, $p_{X_t|X_{t-1}}$, of the form:

$$u_t = \begin{bmatrix} p_{X_t|X_{t-1}}(0|0) & p_{X_t|X_{t-1}}(1|0) \\ 1 & 0 \end{bmatrix}.$$

The disturbance w_t is the channel output, y_t, and thus, it can take values from $\{0, 1\}$.

The notation $\delta_t \triangleq z_{t-1} p_{X_t|X_{t-1}}(1|0)$ is useful and implies the constraint $0 \le \delta_t \le z_{t-1}$, since u_t, by definition, must be a stochastic matrix. Furthermore, given z_{t-1}, u_t can be recovered from δ_t for all $z_{t-1} \ne 0$. For the case $z_{t-1} = 0$, we will see that $p_{X_t|X_{t-1}}(1|0)$ has no effect on the DP, so it can be fixed to zero. The system equation can then be calculated from (40):

$$z_t = \begin{cases} \dfrac{\bar{\alpha} \bar{\delta}_t}{(1-\alpha)(1-\delta_t) + \beta \delta_t} & \text{if } w_t = 0, \\[2mm] \dfrac{\alpha \bar{\delta}_t}{\alpha(1-\delta_t) + (1-\beta)\delta_t} & \text{if } w_t = 1. \end{cases}$$
(41)

The conditional distribution, $p(x_t, x_{t-1}, y_t|z_{t-1}, u_t)$, is described in Table I, so one can calculate the reward

$$\begin{aligned} r(z_{t-1}, u_t) &= I(Y_t; X_t|z_{t-1}, u_t) \\ &= H_2(\bar{\alpha}\bar{\delta}_t + \beta \delta_t) - (1 - \delta_t)H_2(\alpha) - \delta_t H_2(\beta). \end{aligned}$$

Before computing the DP operator, it is convenient to define

$$p_{\alpha,\beta}(\delta) = \alpha \bar{\delta} + \bar{\beta} \delta$$

$$\text{arg1}_{\alpha,\beta}(\delta) = \frac{\bar{\alpha} \bar{\delta}}{1 - p_{\alpha,\beta}(\delta)}$$
(42)

$$\text{arg2}_{\alpha,\beta}(\delta) = \frac{\alpha \bar{\delta}}{p_{\alpha,\beta}(\delta)}.$$

We will omit the subscripts α, β in the notation above when it is clear from the context. The DP operator is then given by:

$$\begin{aligned} (Th_{\alpha,\beta})(z) = \max_{0 \le \delta \le z} & H_2(p(\delta)) - (1 - \delta)H_2(\alpha) - \delta H_2(\beta) \\ & + (1 - p(\delta))h_{\alpha,\beta}(\text{arg1}(\delta)) + p(\delta)h_{\alpha,\beta}(\text{arg2}(\delta)), \end{aligned}$$
(43)

for all functions $h_{\alpha,\beta} : [0, 1] \to \mathbb{R}$, parameterized by (α, β).

Now when the DP problem for the BIBO channel is well-defined, the Bellman equation which can verify the optimality of rewards, can be used to obtain an analytic solution. However, the Bellman equation cannot be easily solved, and therefore, numerical algorithms are required to estimate the Bellman components. The numerical study of DP problems is not within the scope of this paper, and the reader may find [11], [12] to be suitable references for learning this topic in the context of feedback capacities. Therefore, we proceed directly to the statement and the solution of the Bellman equation.

D. The Bellman Equation

In DP, the Bellman equation suggests a sufficient condition for average reward optimality. This equation establishes a mechanism for verifying that a given average reward is optimal. The next result encapsulates the Bellman equation:

Theorem VII.1 (Theorem 6.2, [30]). *If $\rho \in \mathbb{R}$ and a bounded function $h : \mathcal{Z} \to \mathbb{R}$ satisfies for all $z \in \mathcal{Z}$:*

$$\rho + h(z) = \sup_{u \in \mathcal{U}} r(z, u) + \int p(w|z, u) h(F(z, u, w)) dw, \quad (44)$$

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

This result is a direct consequence of Theorem 6.2 in [30]; specifically, the triplet $(\rho, h(\cdot), \mu(\cdot))$ is a canonical triplet by Theorem 6.2, since it satisfies (44). Now, because a canonical triplet defines for all N the N-stage optimal reward and policy under terminal cost $h(\cdot)$, it can be concluded that a canonical triplet also defines the optimal reward and policy in the infinite horizon regime, since in this case, the bounded terminal cost has a negligible effect.

Define two constants

$$\tilde{\rho}_{\alpha,\beta} = \max_{0 \le z \le 1} R_{\alpha,\beta}(z)$$
$$z_{\alpha,\beta}^{\text{opt}} = \arg\max_{0 \le z \le 1} R_{\alpha,\beta}(z), \quad (45)$$

where $R_{\alpha,\beta}(z)$ was defined in (2) and has a unique maximizer from Lemma III.1.

Also, define the functions:

$$h_1^{\alpha,\beta}(z) = H_2(p(z)) - (1-z)H_2(\alpha) - zH_2(\beta)$$
$$X^{\alpha,\beta}(z) = H_2(p(z)) - (1-z)H_2(\alpha) - zH_2(\beta) - p(z)\tilde{\rho}_{\alpha,\beta}$$
$$h_2^{\alpha,\beta}(z) = \frac{X^{\alpha,\beta}(z) + p(z)X^{\alpha,\beta}(\arg2_{\alpha,\beta}(z))}{1 - \alpha\bar{\beta}}, \quad (46)$$

for $z \in [0, 1]$. The concatenation of the above functions can be defined:

$$\tilde{h}_{\alpha,\beta}(z) = \begin{cases} h_1^{\alpha,\beta}(z); & \text{if } 0 \le z \le z_1^{\alpha,\beta} \\ h_2^{\alpha,\beta}(z); & \text{if } z_1^{\alpha,\beta} < z \le z_2^{\alpha,\beta} \\ \tilde{\rho}_{\alpha,\beta} & \text{if } z_2^{\alpha,\beta} < z \le 1, \end{cases}$$

where $z_1^{\alpha,\beta}$ and $z_2^{\alpha,\beta}$ were defined in (15). With these definitions, we are ready to state the fundamental theorem of this section.

Theorem VII.2. *The function $\tilde{h}_{\alpha,\beta}(z)$ and the constant $\tilde{\rho}_{\alpha,\beta}$ satisfy the Bellman equation, i.e.,*

$$\tilde{h}_{\alpha,\beta} + \tilde{\rho}_{\alpha,\beta} = T\tilde{h}_{\alpha,\beta},$$

for all $[\alpha, \beta] \in [0, 1] \times [0, 1]$ satisfying $\alpha + \beta \le 1$. Moreover, the maximum in $T\tilde{h}_{\alpha,\beta}$ is achieved when $\delta^*(z) = z$ for $z \in [0, z_2^{\alpha,\beta}]$, and $\delta^*(z) = z_2^{\alpha,\beta}$ otherwise.

The computations needed to verify that $\tilde{h}_{\alpha,\beta}(z)$ and $\tilde{\rho}_{\alpha,\beta}$ indeed satisfy the Bellman equation are given in Appendix C. To derive this solution in the first place, we started with the numerical techniques used in similar contexts in the prior

works [11]–[13]. The numerical results indicated that the optimal actions are linear in parts and that the number of visited DP states is 4. These observations were sufficient for some good guesswork that helped us to find the Bellman equation solution. For further reading on solving the feedback capacity when actions may be non-linear functions but the DP visits a finite number of states, the reader is referred to [26]. Loosely speaking, the paper [26] includes upper and lower bounds on the feedback capacity that match if the DP (under optimal actions) visits a finite number of states.

As consequences of Theorem VII.2, we obtain the facts that the feedback capacity and the optimal input distribution of the BIBO channel are as stated in Theorems III.1 and IV.2.

Proof of Theorem III.1: By Theorem VII.2, the DP optimal average reward is $\tilde{\rho}_{\alpha,\beta}$, which is the same as the capacity expression in the statement of Theorem III.1. ∎

Proof of Theorem IV.2: We first show that the initial DP state may be assumed to be such that the optimal policy visits only a finite set of DP states. From this, we infer the form of the optimal input distribution given in the theorem statement. It is then straightforward to check that $\{(X_i, Q_i)\}_{i \ge 1}$ forms an irreducible and aperiodic Markov chain. It follows from this that the average reward, i.e., the feedback capacity, can also be expressed as $I(X; Y|Q)$.

Recall the optimal actions from Theorem VII.2:

$$\delta^*(z) = \begin{cases} z & \text{if } 0 \le z \le z_2^{\alpha,\beta} \\ z_2^{\alpha,\beta} & \text{if } z_2^{\alpha,\beta} < z \le 1. \end{cases}$$

The DP state evolution in (41) can be described using the $\arg j_{\alpha,\beta}$ functions in (42). It is easy to check that the set $\mathcal{Z}_* \triangleq \{z_i^{\alpha,\beta} : i = 1, 2, 3, 4\}$, with $z_i^{\alpha,\beta}$ as defined in (15), is closed under the composite function $\arg j_{\alpha,\beta} \circ \delta^*$, i.e., $\arg j_{\alpha,\beta}(\delta^*(z_i^{\alpha,\beta})) \in \mathcal{Z}_*$ for all i, j. The functions $\arg j_{\alpha,\beta}(\delta^*(z))$, $j = 1, 2$, create a sink, meaning that there is always a positive probability for a transition being made from any DP state $z \in \mathcal{Z}$ to a state in $\mathcal{Z}_*$, and zero probability of leaving the set $\mathcal{Z}_*$. Therefore, we can assume that the initial DP state z_0 is from $\mathcal{Z}_*$. Note also that for the function $g(\cdot, \cdot)$ that describes the transitions in the Q-graph in Fig. 4, we have $\ell = g(i, j-1)$ iff $z_\ell^{\alpha,\beta} = \arg j_{\alpha,\beta}(\delta^*(z_i^{\alpha,\beta}))$, for all $i, \ell \in \{1, 2, 3, 4\}$ and $j \in \{1, 2\}$. Therefore, we may identify the set $\mathcal{Z}_*$ with the set of Q-states $\mathcal{Q} = \{1, 2, 3, 4\}$, so that the evolution of the DP states can be described on the Q-graph of Fig. 4. The form given for the optimal input distribution in the theorem statement follows directly from this observation.

We next verify the first-order Markov property of $(X_i, Q_i)_{i \ge 1}$. Observe that

$$p(x_i, q_i | x^{i-1}, q^{i-1})$$
$$= \sum_{y_i} p(y_i, x_i, q_i | x^{i-1}, q^{i-1})$$
$$\stackrel{(a)}{=} \sum_{y_i} p(q_i | q_{i-1}, y_i) p(y_i | x_i) p_{X|X^-,Q}^*(x_i | x_{i-1}, q_{i-1}),$$
$$\quad (47)$$

where (a) follows from the structure of $p_{X|X^-,Q}^*$ given in (16), the memoryless channel property, and the fact that Q_i is a function of (Q_{i-1}, Y_i). It can be verified that the Markov chain

$(X_i, Q_i)_{i \geq 1}$ is irreducible and aperiodic, and hence, ergodic. It is then a straightforward, albeit tedious, exercise to verify that its (unique) stationary distribution is given by $\pi_{X^-,Q} = \pi_{X^-|Q}\pi_Q$, as in the statement of the theorem.

Any trajectory of states z_t, $t \geq 0$, followed by the DP under the optimal policy accumulates an average reward of $\liminf_N \frac{1}{N}\sum_{t=1}^{N} r(z_t, \delta^*(z_t))$. Since we assume that $z_0 \in \mathcal{Z}_*$, each term in the sum is of the form $r(z_i^{\alpha,\beta}, \delta^*(z_i^{\alpha,\beta}))$ for some $i \in \mathcal{Q}$, which is equal to $I(Y; X | Q = i)$, the reward at the DP state $z_i^{\alpha,\beta}$. It then follows from the ergodicity of the Markov chain $(X_i, Q_i)_{i \geq 1}$ that, as $N \to \infty$, the time average $\frac{1}{N}\sum_{t=1}^{N} r(z_t, \delta^*(z_t))$ converges almost surely to $\sum_{i=1}^{4} I(Y; X | Q = i)\pi_Q(i) = I(X; Y | Q)$. ∎

VIII. SUMMARY AND CONCLUDING REMARKS

The capacity of the BIBO channel with input constraints was derived using a corresponding DP problem. A by-product of the DP solution is the optimal input distribution, which can be described compactly using Q-graphs. For the S-channel, we were able to derive a capacity-achieving coding scheme with simple and intuitive analysis for the achieved rate. For the general BIBO channel, we provided a PMS construction that includes the new element of history bits that captures the memory embedded in the setting. Furthermore, to ease the analysis of the scheme, a message-interval splitting operation was introduced so as to keep a Markov property of the involved random variables. With these ideas, we showed that the constructed PMS achieves the capacity of the BIBO channel.

The elements that were presented here for the PMS in the input-constrained BIBO channel setting may be exploited to derive a PMS for a broader class of finite-state channels (FSC) with feedback. Specifically, a FSC is *unifilar* if the channel state s_t is a deterministic function of the previous channel state s_{t-1}, input x_t and output y_t. Though several works have proposed the PMS approach for this class, the assertion that this approach is optimal (in the sense of being feedback-capacity achieving) remains to be proved [31], [32]. The idea of history bits that was presented in this paper can be extended to "history states" for unifilar channels, since the encoder can determine the channel state at each time t, assuming knowledge of the initial state s_0. Moreover, for all unifilar channels with simple capacity expressions, their optimal input distributions have a Q-graph representation [26] so that they satisfy the BCJR-invariant property that is crucial for the PMS construction. Thus, the steps of the construction and the analysis can be repeated in order to show that the corresponding PMS operates successfully in the sense of Theorem V.1.

APPENDIX A
PROOF OF LEMMA III.1

In this appendix, we show that the argument that achieves the maximum of $R_{\alpha,\beta}(z)$ is unique and lies within $[z_L, z_U] = \left[\frac{\sqrt{\alpha}}{\sqrt{\alpha}+\sqrt{\bar{\beta}}}, \frac{\sqrt{\alpha}}{\sqrt{\alpha}+\sqrt{\beta}}\right]$.

Let $p(z) = \alpha\bar{z} + \bar{\beta}z$ and denote by p' the derivative of $p(z)$. After some simplifications, the derivative equals:

$$\frac{d}{dz}R(z)$$
$$= \frac{1}{(1+p(z))^2}\left\{(1-\alpha\bar{\beta})(H_2(\alpha) - H_2(\beta)) + (\bar{\beta} - \alpha)\right.$$
$$\left. \cdot [2\log(1-p(z)) - \log(p(z) - \alpha\bar{\beta})(1+\alpha\bar{\beta}) + \alpha\bar{\beta}\log\alpha\bar{\beta}]\right\}$$

The above derivative equals zero when the function

$$f_{\alpha,\beta}(z) \triangleq (1-\alpha\bar{\beta})[H_2(\alpha) - H_2(\beta)] + (\bar{\beta} - \alpha)$$
$$\cdot [2\log(1-p(z)) - \log(p(z) - \alpha\bar{\beta})(1+\alpha\bar{\beta}) + \alpha\bar{\beta}\log\alpha\bar{\beta}]$$

equals zero. It is easy to note that the function $f_{\alpha,\beta}(z)$ is a decreasing function of its argument and, therefore, it can be zero at one point only.

We will show two facts:

$$f_{\alpha,\beta}(p(z_L)) \geq 0 \tag{48}$$
$$f_{\alpha,\beta}(p(z_U)) \leq 0,, \tag{49}$$

from which we can conclude that $R_{\alpha,\beta}(z)$ attains its maximum at a unique point in $z \in [z_L, z_U]$. For the BSC, it needs to be shown that $f_{\alpha,\alpha}(0.5) \leq 0$ which can be verified immediately.

We begin with an explicit calculation of $f_{\alpha,\beta}(p(z_L))$:

$$f_{\alpha,\beta}(p(z_L))$$
$$\overset{(a)}{=} (1-\alpha\bar{\beta})[H_2(\alpha) - H_2(\beta)] + (\bar{\beta} - \alpha) \cdot$$
$$\cdot [2\log(1-\sqrt{\alpha\bar{\beta}}) - \log(\sqrt{\alpha\bar{\beta}} - \alpha\bar{\beta})(1+\alpha\bar{\beta}) + \alpha\bar{\beta}\log\alpha\bar{\beta}]$$
$$= (1-\alpha\bar{\beta})\left[H_2(\alpha) - H_2(\beta) + (\bar{\beta} - \alpha)\log\left(\frac{1-\sqrt{\alpha\bar{\beta}}}{\sqrt{\alpha\bar{\beta}}}\right)\right], \tag{50}$$

where (a) follows from $p(z_L) = \sqrt{\alpha\bar{\beta}}$. Since $1 - \alpha\bar{\beta} \geq 0$, we need to show that $\frac{f_{\alpha,\beta}(p(z_L))}{1-\alpha\bar{\beta}} \geq 0$.

We now show that the minimal value of (50) is 0. Consider the first derivative, with respect to α, of $\frac{f_{\alpha,\beta}(p(z_L))}{1-\alpha\bar{\beta}}$:

$$\frac{d}{d\alpha}\left[H_2(\alpha) - H_2(\beta) + (\bar{\beta} - \alpha)\log\left(\frac{1-\sqrt{\alpha\bar{\beta}}}{\sqrt{\alpha\bar{\beta}}}\right)\right]$$
$$= \log\left(\frac{(1-\alpha)\sqrt{\alpha\bar{\beta}}}{\alpha(1-\sqrt{\alpha\bar{\beta}})}\right) - \frac{\bar{\beta} - \alpha}{2\alpha(1-\sqrt{\alpha\bar{\beta}})}$$
$$\leq \frac{(1-\alpha)\sqrt{\alpha\bar{\beta}}}{\alpha(1-\sqrt{\alpha\bar{\beta}})} - 1 - \frac{\bar{\beta} - \alpha}{2\alpha(1-\sqrt{\alpha\bar{\beta}})}$$
$$= \frac{-(\sqrt{\bar{\beta}} - \sqrt{\alpha})^2}{2\alpha(1-\sqrt{\alpha\bar{\beta}})}$$
$$\leq 0, \tag{51}$$

where the first inequality follows from $\log x < x - 1$ for all $x > 0$ with $x = \frac{(1-\alpha)\sqrt{\alpha\bar{\beta}}}{\alpha(1-\sqrt{\alpha\bar{\beta}})}$.

Therefore, for each β, the function is non-increasing in α, so the function can only be decreased if we substitute $\alpha = \bar{\beta}$. Since $f_{\bar{\beta},\beta}(p(z_L)) = 0$, inequality (48) is proven.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2018.2809554, IEEE Transactions on Information Theory

18

We now use a similar methodology to show (49). The inequality that needs to be shown is

$$f_{\alpha,\beta}(p(z_U))$$
$$= (1 - \alpha\bar{\beta})[H_2(\alpha) - H_2(\beta)] + (\bar{\beta} - \alpha) \cdot$$
$$\cdot (2\log\sqrt{\bar{\alpha}\beta} + \alpha\bar{\beta}\log\alpha\bar{\beta} - \log(1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta})(1 + \alpha\bar{\beta}))$$
$$\leq 0. \tag{52}$$

Because it is difficult to prove straightforwardly, we write inequality (52) as a sum of simpler components, i.e., $-f_{\alpha,\beta}(p(z_U)) = F^1_{\alpha,\beta} + F^2_{\alpha,\beta}$, and we show that $F^1_{\alpha,\beta}$ and $F^2_{\alpha,\beta}$ are always non-negative. The functions are:

$$F^1_{\alpha,\beta} = \alpha\bar{\beta}\left[(\bar{\beta} - \alpha)\log\left(\frac{1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta}}{\alpha\bar{\beta}}\right) + H_2(\alpha) - H_2(\beta)\right]$$

$$F^1_{\alpha,\beta} = (\bar{\beta} - \alpha)\log\left(\frac{1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta}}{\bar{\alpha}\beta}\right) - (H_2(\alpha) - H_2(\beta))$$

As before, we take the first derivative of $F^1_{\alpha,\beta}$:

$$\frac{d}{d\alpha}\left[\frac{F^1_{\alpha,\beta}}{\alpha\bar{\beta}}\right]$$
$$= (\bar{\beta} - \alpha)\left[-\frac{1}{\alpha} + \frac{-\bar{\beta} + \frac{\beta}{2\sqrt{\bar{\alpha}\beta}}}{1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta}}\right] + \log\left(\frac{\bar{\alpha}\beta}{1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta}}\right)$$
$$\leq (\bar{\beta} - \alpha)\left[-\frac{1}{\alpha} + \frac{-\bar{\beta} + \frac{\beta}{2\sqrt{\bar{\alpha}\beta}}}{1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta}}\right] + \left(\frac{\bar{\alpha}\beta}{1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta}}\right) - 1$$
$$= \frac{(\sqrt{\bar{\alpha}} - \sqrt{\bar{\beta}})\sqrt{\beta}[\sqrt{\bar{\alpha}}(\alpha\sqrt{\beta} - \bar{\beta}\sqrt{\alpha}) + (\alpha\beta - \bar{\alpha}\bar{\beta})]}{2\alpha\sqrt{\bar{\alpha}\beta}(1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta})}$$
$$\leq 0,$$

where the first inequality follows from $\log x \leq x - 1$ with $x = \frac{\bar{\alpha}\beta}{1 - \sqrt{\bar{\alpha}\beta} - \alpha\bar{\beta}}$. The last inequality follows from $\alpha \leq \bar{\beta}$, which implies, in turn, $\alpha\sqrt{\beta} - \bar{\beta}\sqrt{\alpha} \leq 0$ and $\alpha\beta - \bar{\alpha}\bar{\beta} \leq 0$. We thus conclude that $F^1_{\alpha,\beta}$ is non-increasing in α, and therefore, if we take α to be $1 - \beta$, we get its minimal value. Note that $F^1_{\bar{\beta},\beta} = 0$, so we have $F^1_{\alpha,\beta} \geq 0$.

Now we take the derivative of $F^2_{\alpha,\beta}$ with respect to β:

$$\frac{d}{d\beta}F^2_{\alpha,\beta}$$
$$= \frac{(\beta - \bar{\alpha})\sqrt{\bar{\alpha}\beta}(-\beta + 2\sqrt{\beta\bar{\alpha}})}{2\beta^2(1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta})} + \log\left(\frac{\bar{\alpha}\beta}{1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta}}\right)$$
$$\leq \frac{(\beta - \bar{\alpha})\sqrt{\bar{\alpha}}(-\sqrt{\beta} + 2\sqrt{\bar{\alpha}})}{2\beta(1 - \alpha\bar{\beta} - \sqrt{\beta\bar{\alpha}})} + \left(\frac{\bar{\alpha}\beta}{1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta}}\right) - 1$$
$$= \frac{(\sqrt{\beta} - \sqrt{\bar{\alpha}})}{2\beta(1 - \alpha\bar{\beta} - \sqrt{\bar{\alpha}\beta})}(\bar{\alpha}\sqrt{\beta} + 2\beta\sqrt{\beta} + \sqrt{\bar{\alpha}}(2\bar{\alpha} - \beta))$$
$$\leq 0.$$

The last inequality follows from $\sqrt{\beta} - \sqrt{\bar{\alpha}} \leq 0$. Repeating the same steps as was done for $F^1_{\alpha,\beta}$, we find that $F^2_{\alpha,\beta} \geq 0$, which, in turn, gives that $-f_{\alpha,\beta}(p(z_U)) \geq 0$ as required. ∎

APPENDIX B
PROOF OF THEOREM III.3

Throughout this section, we use $x = \alpha\bar{\alpha}$, and x', x'' to stand for the first and second derivatives of x, respectively.

Recall that p_α in Corollary 1 is the solution for $(\alpha\bar{\alpha})\log(\alpha\bar{\alpha}) + 2\log(1-p) = (1 + \alpha\bar{\alpha})\log(p - \alpha\bar{\alpha})$, and let p'_α denote its first derivative. The next lemma concerns p'_α and is the foundation for the proof of Theorem III.3.

Lemma B.1. *The first derivative of p_α is:*

$$p'_\alpha = (1 - 2\alpha)\frac{(1 - p_\alpha)(p_\alpha - \alpha\bar{\alpha})}{(\alpha\bar{\alpha} - 1)(1 + p_\alpha)}\left[\log\left(\frac{p_\alpha - \alpha\bar{\alpha}}{\bar{\alpha}}\right) - \log\alpha\right]$$
$$+ (1 - 2\alpha)\frac{1 - p_\alpha}{1 - \alpha\bar{\alpha}}$$
$$\triangleq K_2(\alpha) - K_1(\alpha)\log\alpha,$$

with the defined functions:

$$K_1(\alpha) \triangleq (1 - 2\alpha)\frac{(1 - p_\alpha)(p_\alpha - \alpha\bar{\alpha})}{(\alpha\bar{\alpha} - 1)(1 + p_\alpha)}$$

$$K_2(\alpha) \triangleq K_1(\alpha)\log\left(\frac{p_\alpha - \alpha\bar{\alpha}}{\bar{\alpha}}\right) + (1 - 2\alpha)\frac{1 - p_\alpha}{1 - \alpha\bar{\alpha}}.$$

Note that $p_0 = 2 - \lambda$, so $K_1(\alpha)$ and $K_2(\alpha)$ are defined at $\alpha = 0$:

$$K_1(0) = -\frac{p_0(1 - p_0)}{1 + p_0}$$
$$K_2(0) = K_1(0)\log p_0 + 1 - p_0.$$

Proof of Lemma B.1: We calculate the first derivative for each side of $2\log(1 - p_\alpha) = (1 + \alpha\bar{\alpha})\log(p_\alpha - \alpha\bar{\alpha}) - (\alpha\bar{\alpha})\log(\alpha\bar{\alpha})$ so we have:

$$\frac{-2}{1 - p_\alpha}p'_\alpha = x'\left[\log\left(\frac{p_\alpha - x}{x}\right) - \frac{1 + p_\alpha}{p_\alpha - x}\right] + \frac{1 + x}{p_\alpha - x}p'_\alpha. \tag{53}$$

Arranging both sides of (53) complete the proof. ∎

The next lemma is technical and is made to shorten the proof of Theorem III.3:

Lemma B.2. *Define $K_3(\alpha) = \frac{x'}{1 + p_\alpha}$, then it can be expressed as*

$$K_3(\alpha) = \frac{1}{1 + p_0} + N\alpha\log\alpha + o(\alpha\log\alpha),$$

where N is a constant.

The proof of Lemma B.2 appears in Appendix B-A. We are now ready to prove the main result of this section.

Proof of Theorem III.3: Consider the next chain of equalities:

$$C^{\text{BSC}}(\alpha) + H_2(\alpha) - C^{\text{BSC}}(0) + K_3(\alpha)\alpha\log\alpha$$
$$\overset{(a)}{=} \log(1 - p_\alpha) - \log(p_\alpha - x) - C^{\text{BSC}}(0) + K_3(\alpha)\alpha\log\alpha$$
$$\overset{(b)}{=} \alpha\left[\frac{-p'_\alpha}{1 - p_\alpha} - \frac{p'_\alpha - x'}{p_\alpha - x} + K'_3(\alpha)\alpha\log\alpha\right.$$
$$\left. + K_3(\alpha)[1 + \log\alpha]\right]_{\alpha=0} + o(\alpha)$$
$$\overset{(c)}{=} \alpha\left[\frac{p'_\alpha(x - 1)}{(1 - p_\alpha)(p_\alpha - x)} + \frac{x'}{p_\alpha - x} + K_3(\alpha)[1 + \log\alpha]\right]_{\alpha=0}$$
$$+ o(\alpha)$$
$$\overset{(d)}{=} \alpha\left[\frac{(K_2(\alpha) - K_1(\alpha)\log\alpha)(x - 1)}{(1 - p_\alpha)(p_\alpha - x)} + \frac{x'}{p_\alpha - x}\right.$$

$$\left. + K_3(\alpha)[1 + \log \alpha]\right]_{\alpha=0} + o(\alpha)$$

$$\stackrel{(e)}{=} M\alpha + \alpha \left[\frac{K_1(\alpha) \log \alpha (1 - x)}{(1 - p_\alpha)(p_\alpha - x)} + K_3(\alpha) \log \alpha \right]_{\alpha=0} + o(\alpha)$$

$$\stackrel{(f)}{=} M\alpha + o(\alpha) \tag{54}$$

where:

(a) follows from Corollary 1;

(b) follows from the Taylor series approximation $f(\alpha) = f(0) + f'(0)\alpha + O(\alpha^2)$;

(c) follows from Lemma B.2, specifically, $\lim_{\alpha \to 0} K_3'(\alpha)\alpha \log \alpha = 0$;

(d) follows from Lemma B.1, specifically, $p_\alpha' = K_2(\alpha) - K_1(\alpha) \log \alpha$;

(e) follows from the notation $M \triangleq K_2(0) \frac{-1}{(1-p_0)(p_0)} + \frac{1}{p_0} + K_3(0)$;

(f) follows from $\frac{K_1(\alpha)(1-x)}{(1-p_\alpha)(p_\alpha - x)} + K_3(\alpha) = 0$.

Thus, we have from (54) that $C^{\mathrm{BSC}}(\alpha) + H_2(\alpha) - C^{\mathrm{BSC}}(0) + K_3(\alpha)\alpha \log \alpha = M\alpha + o(\alpha)$.

The derivation is completed with the following equalities:

$$C^{\mathrm{BSC}}(\alpha)$$
$$= C^{\mathrm{BSC}}(0) - K_3(\alpha)\alpha \log \alpha - H_2(\alpha) + M\alpha + o(\alpha)$$
$$\stackrel{(a)}{=} \log \lambda - [K_3(0) + N\alpha \log \alpha + o(\alpha \log \alpha)]\alpha \log \alpha$$
$$\quad - H_2(\alpha) + M\alpha + o(\alpha)$$
$$\stackrel{(b)}{=} \log \lambda + \frac{2 - \lambda}{3 - \lambda}\alpha \log \alpha + \left(\frac{\log(2 - \lambda) - (2 - \lambda)}{3 - \lambda} \right)\alpha$$
$$\quad + O(\alpha^2 \log^2 \alpha)$$

where:

(a) follows from $C^{\mathrm{BSC}}(0) = \log \lambda$ and Lemma B.2;

(b) follows from $H_2(\alpha) = -\alpha \log \alpha + \alpha + o(\alpha)$ and arranging the equation.

■

A. Proof of Lemma B.2

By a Taylor series approximation, we have

$$\frac{x'}{1 + p_\alpha} - \frac{x'K_1(\alpha)}{(1 + p_\alpha)^2}\alpha \log \alpha - \frac{1}{1 + p_0}$$
$$= \alpha \left[\frac{x''}{1 + p_\alpha} - p_\alpha' \frac{x'}{(1 + p_\alpha)^2} - \left(\frac{x'K_1(\alpha)}{(1 + p_\alpha)^2} \right)' \alpha \log \alpha \right.$$
$$\quad \left. - \frac{x'K_1(\alpha)}{(1 + p_\alpha)^2}(1 + \log \alpha) \right]_{\alpha=0} + o(\alpha)$$
$$\stackrel{(a)}{=} C\alpha + \alpha \left[-p_\alpha' \frac{x'}{(1 + p_\alpha)^2} - \frac{x'K_1(\alpha)}{(1 + p_\alpha)^2}\log \alpha \right]_{\alpha=0} + o(\alpha)$$
$$\stackrel{(b)}{=} C\alpha + \alpha \left[K_2(\alpha)\frac{-x'}{(1 + p_\alpha)^2} \right]_{\alpha=0} + o(\alpha)$$
$$\stackrel{(c)}{=} \tilde{C}\alpha + o(\alpha),$$

where (a) follows from the fact that $\lim_{\alpha \to 0} \left(\frac{x'K_1(\alpha)}{(1 + p_\alpha)^2} \right)' \alpha \log \alpha = 0$ and the notation $C = \frac{-2}{1 + p_\alpha} - \frac{K_1(0)}{(1 + p_0)^2}$, (b) follows from $p_\alpha' = K_2(\alpha) - K_1(\alpha)\alpha \log \alpha$, and finally, (c) follows from the notation $\tilde{C} = C - \frac{K_2(0)}{(1 + p_0)^2}$.

TABLE II
THE FUNCTIONS ARG1(z) AND ARG2(z)

	domain	$\tilde{h}(\mathrm{arg1}(z))$	$\tilde{h}(\mathrm{arg2}(z))$
I	$[0, z_1]$	$\tilde{\rho}_{\alpha,\beta}$	$\tilde{\rho}_{\alpha,\beta}$
II	$[z_1, z_2]$	$\tilde{\rho}_{\alpha,\beta}$	$h_2^{\alpha,\beta}(\mathrm{arg2}(z))$
III	$[z_2, z_3]$	$\tilde{\rho}_{\alpha,\beta}$	$h_1^{\alpha,\beta}(\mathrm{arg2}(z))$
IV	$[z_3, z_4]$	$h_2^{\alpha,\beta}(\mathrm{arg1}(z))$	$h_1^{\alpha,\beta}(\mathrm{arg2}(z))$
V	$[z_4, 1]$	$h_1^{\alpha,\beta}(\mathrm{arg1}(z))$	$h_1^{\alpha,\beta}(\mathrm{arg2}(z))$

So, we have that

$$\frac{x'}{1 + p_\alpha} = \frac{1}{1 + p_0} + \tilde{C}\alpha + o(\alpha) + \frac{x'K_1(\alpha)}{(1 + p_\alpha)^2}\alpha \log \alpha.$$

Applying the Taylor series approximation once again on $\frac{x'K_1(\alpha)}{(1+p_\alpha)^2}$ gives that:

$$\frac{x'K_1(\alpha)}{(1 + p_\alpha)^2} = \frac{K_1(0)}{(1 + p_0)^2} + h(\alpha),$$

where $h(\alpha)$ is some function such that $\lim_{\alpha \to 0} h(\alpha) = 0$.

Combining the last two derivations, we have the required equality, i.e.,

$$\frac{x'}{1 + p_\alpha} = \frac{1}{1 + p_0} + \tilde{C}\alpha + o(\alpha) + \left[\frac{K_1(0)}{(1 + p_0)^2} + h(\alpha) \right]\alpha \log \alpha$$
$$= \frac{1}{1 + p_0} + N\alpha \log \alpha + o(\alpha \log \alpha),$$

where $N = \frac{K_1(0)}{(1 + p_0)^2}$. ■

APPENDIX C
PROOF OF THEOREM VII.2

The following lemma is technical and is useful for understanding the structure of $\tilde{h}_{\alpha,\beta}(z)$.

Lemma C.1. *For all* $[\alpha, \beta] \in [0, 1] \times [0, 1]$ *s.t.* $\alpha + \beta \leq 1$,

1) *The function* $\tilde{h}_{\alpha,\beta}(z)$ *is continuous on* $[0, 1]$.
2) *The function* $\tilde{h}_{\alpha,\beta}(z)$ *is concave on* $[0, 1]$.
3) *The only maximum of* $h_2^{\alpha,\beta}(z)$ *is attained at* $z = z_2^{\alpha,\beta}$, *and its value is* $\tilde{\rho}_{\alpha,\beta}$.
4) *The first derivative of* $h_1^{\alpha,\beta}(z)$ *is non-negative for* $z \in [0, z_1^{\alpha,\beta}]$.

The proof of Lemma C.1 appears in Appendix C-A.

Proof of Theorem VII.2: The function $\tilde{h}_{\alpha,\beta}(z)$ is defined as a concatenation of $h_1^{\alpha,\beta}(z)$, $h_2^{\alpha,\beta}(z)$, and $\tilde{\rho}_{\alpha,\beta}$; to simplify the calculation of $(T\tilde{h}_{\alpha,\beta})(z)$, the unit interval is partitioned into non-intersecting sub-intervals, where each sub-interval uniquely determines the function $\tilde{h}_{\alpha,\beta}(\mathrm{arg}i(z))$ to be $h_1^{\alpha,\beta}(z)$, $h_2^{\alpha,\beta}(z)$ or $\tilde{\rho}_{\alpha,\beta}$, for $i = 1, 2$. Since there are two concatenation points, $z_1^{\alpha,\beta}$ and $z_2^{\alpha,\beta}$, the unit interval is partitioned at the set of points that satisfy,

$$\mathrm{arg1}_{\alpha,\beta}(z) = z_i^{\alpha,\beta}$$
$$\mathrm{arg2}_{\alpha,\beta}(z) = z_i^{\alpha,\beta}, \tag{55}$$

for $i = 1, 2$.

Calculation of the points in (55) reveals that the unit interval should be partitioned at $z_1^{\alpha,\beta}$, $z_2^{\alpha,\beta}$, $z_3^{\alpha,\beta}$, $z_4^{\alpha,\beta}$ from (15). Fig. 8 illustrates the argument functions and the partitions when
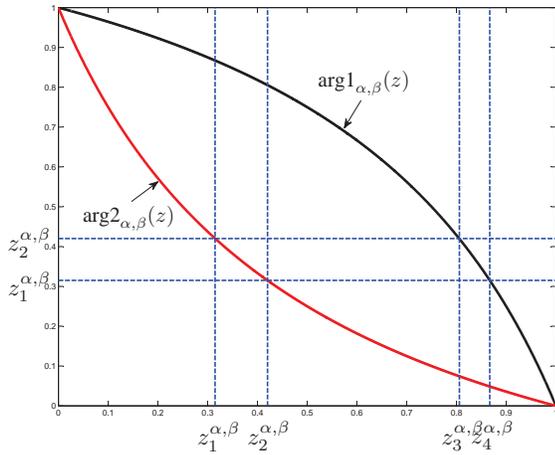
Fig. 8. Illustration of the argument functions as a function of z, for $\alpha = \beta = 0.25$.

$\alpha = \beta = 0.25$. As can be seen from Fig. 8, five segments need to be considered when calculating $(T\tilde{h}_{\alpha,\beta})(z)$. The relevant segments are summarized in Table II together with $\tilde{h}_{\alpha,\beta}(\text{arg}i(z))$ for $i = 1, 2$ for each sub-interval.

Now, the operator $T\tilde{h}_{\alpha}(z)$ can be calculated, such that in each calculation, we restrict actions to one sub-interval from Table II. For the interval I, i.e., $z \in [0, z_1^{\alpha,\beta})$,

$$
\begin{aligned}
&(T\tilde{h}_{\alpha,\beta})(z) \\
&= \sup_{0 \leq \delta \leq z} H_2(p(\delta)) - (1-\delta)H_2(\alpha) - \delta H_2(\beta) \\
&\quad + (1 - p(\delta))h_{\alpha,\beta}(\text{arg}1(\delta)) + p(\delta)h_{\alpha,\beta}(\text{arg}2(\delta)) \\
&\stackrel{(a)}{=} \sup_{0 \leq \delta \leq z} H_2(p(\delta)) - (1-\delta)H_2(\alpha) - \delta H_2(\beta) \\
&\quad + (1 - p(\delta))\tilde{\rho}_{\alpha,\beta} + p(\delta)\tilde{\rho}_{\alpha,\beta} \\
&\stackrel{(b)}{=} \sup_{0 \leq \delta \leq z} h_1(\delta) + \tilde{\rho}_{\alpha,\beta} \\
&\stackrel{(c)}{=} h_1(z) + \tilde{\rho}_{\alpha,\beta}
\end{aligned} \tag{56}
$$

where (a) follows from the restriction of $z \in [0, z_1]$ and substituting the functions from Table II, (b) follows from the definition of $h_1(\delta)$ in (46) and (c) follows from Item 4) of Lemma C.1, specifically, $h_1^{\alpha,\beta}(z)$ is non-decreasing on $[0, z_1^{\alpha,\beta}]$. Note that the maximizer is $\delta(z) = z$.

The operator with actions restricted to interval II, i.e., $\delta \in [z_1^{\alpha,\beta}, z]$ for $z \in [z_1^{\alpha,\beta}, z_2^{\alpha,\beta}]$ is:

$$
\begin{aligned}
&\sup_{z_1 \leq \delta \leq z} H_2(p(\delta)) - (1-\delta)H_2(\alpha) - \delta H_2(\beta) \\
&\quad + (1 - p(\delta))\tilde{h}_{\alpha,\beta}(\text{arg}1(\delta)) + p(\delta)\tilde{h}_{\alpha,\beta}(\text{arg}2(\delta)) \\
&\stackrel{(a)}{=} \sup_{z_1 \leq \delta \leq z} X(\delta) + p(\delta)\tilde{\rho}_{\alpha,\beta} + (1 - p(\delta))\tilde{\rho}_{\alpha,\beta} \\
&\quad + \left[\frac{p(\delta)X(\text{arg}2(\delta)) + \alpha\bar{\beta}X(\delta)}{1 - \alpha\bar{\beta}} \right] \\
&\stackrel{(b)}{=} \sup_{z_1 \leq \delta \leq z} h_2(\delta) + \tilde{\rho}_{\alpha,\beta} \\
&\stackrel{(c)}{=} h_2(z) + \tilde{\rho}_{\alpha,\beta},
\end{aligned} \tag{57}
$$

where (a) follows from the definition of $X(\delta)$ in (46) and Table II, (b) follows the expression for $h_2(\delta)$ in (46) and (c) follows from Item 3) in Lemma C.1, where it was shown that $h_2(z)$ is increasing on $[0, z_2^{\alpha,\beta}]$.

To conclude the calculation of $(T\tilde{h}_{\alpha,\beta})(z)$ for $z \in [z_1^{\alpha,\beta}, z_2^{\alpha,\beta}]$, consider

$$
\begin{aligned}
(T\tilde{h}_{\alpha,\beta})(z) &\stackrel{(a)}{=} \max(\sup_{z \in [0, z_1]} h_1(z) + \tilde{\rho}_{\alpha,\beta}, \sup_{z \in [z_1, z]} h_2(z) + \tilde{\rho}_{\alpha,\beta}) \\
&\stackrel{(b)}{=} \max(h_1(z_1), h_2(z)) + \tilde{\rho}_{\alpha,\beta} \\
&\stackrel{(c)}{=} h_2(z) + \tilde{\rho}_{\alpha,\beta},
\end{aligned} \tag{58}
$$

where (a) follows from (56) and (57), and both (b) and (c) follow from Items 3) and 4) in Lemma C.1. Note also here that the maximizer of $(T\tilde{h}_{\alpha,\beta})(z)$ for z on sub-interval II is $\delta(z) = z$.

For actions that are restricted to interval III, i.e., $\delta \in [z_2^{\alpha,\beta}, z]$ with $z \in [z_2^{\alpha,\beta}, z_3^{\alpha,\beta}]$, consider

$$
\begin{aligned}
&\sup_{z_2 \leq \delta \leq z} H_2(p(\delta)) - (1-\delta)H_2(\alpha) - \delta H_2(\beta) \\
&\quad + (1 - p(\delta))\tilde{h}_{\alpha,\beta}(\text{arg}1(\delta)) + p(\delta)\tilde{h}_{\alpha,\beta}(\text{arg}2(\delta)) \\
&\stackrel{(a)}{=} \sup_{z_2 \leq \delta \leq z} X(\delta) + \tilde{\rho}_{\alpha,\beta} + p(\delta)\left[X(\text{arg}2_{\alpha,\beta}(\delta)) + \left(\frac{\alpha\bar{\beta}}{p(\delta)} \right)\tilde{\rho}_{\alpha,\beta} \right] \\
&\stackrel{(b)}{=} \tilde{\rho}_{\alpha,\beta} + \tilde{\rho}_{\alpha,\beta},
\end{aligned} \tag{59}
$$

where (a) follows from the definition of $X(\delta)$ in (46) and Table II and (b) follows from Item 3) in Lemma C.1, specifically, $h_2(z)$ achieves its maximum value at $z = z_2$. Note from (58) and (59) that the operator on III satisfies $(T\tilde{h}_{\alpha,\beta})(z) = 2\tilde{\rho}_{\alpha,\beta}$ and that the maximizer is $\delta(z) = z_2$.

For the action restricted on interval IV, i.e., $\delta \in [z_3, z]$ with $z \in [z_3, z_4]$, consider

$$
\begin{aligned}
&\sup_{z_3 \leq \delta \leq z} H_2(p(\delta)) - (1-\delta)H_2(\alpha) - \delta H_2(\beta) \\
&\quad + (1 - p(\delta))\tilde{h}_{\alpha,\beta}(\text{arg}1(\delta)) + p(\delta)\tilde{h}_{\alpha,\beta}(\text{arg}2(\delta)) \\
&\stackrel{(a)}{=} \sup_{z_3 \leq \delta \leq z} X(\delta) + p(\delta)\tilde{\rho}_{\alpha,\beta} + (1 - p(\delta))h_2(\text{arg}1_{\alpha,\beta}(\delta)) \\
&\quad + p(\delta)\left[X(\text{arg}2(\delta)) + \frac{\alpha\bar{\beta}}{p(\delta)}\tilde{\rho}_{\alpha,\beta} \right] \\
&\stackrel{(b)}{\leq} \tilde{\rho}_{\alpha,\beta} + \tilde{\rho}_{\alpha,\beta},
\end{aligned} \tag{60}
$$

where (a) follows from the definition of $X(\delta)$ in (46) and Table II and (b) follows from $h_2(z) \leq \tilde{\rho}_{\alpha,\beta}$ shown in Item 3), Lemma C.1.

The calculation of the last interval, $[z_4, 1]$, is omitted here, but it follows the same repeated arguments, so we have $(T\tilde{h}_{\alpha,\beta})(z) \leq 2\tilde{\rho}_{\alpha,\beta}$. Now, Item 3) in Lemma C.1 together with (60) gives us that $(T\tilde{h}_{\alpha,\beta})(z) = 2\tilde{\rho}_{\alpha,\beta}$ also for $z \in [z_2, z_4]$. To conclude, we have shown that $(T\tilde{h}_{\alpha,\beta})(z) = \tilde{h}_{\alpha,\beta}(z) + \tilde{\rho}_{\alpha,\beta}$.

A. Proof of Lemma C.1

Throughout this section, we use z_i as shorthand for $z_i^{\alpha,\beta}$ and p^{opt} stands for $p(z_2^{\alpha,\beta})$.

Continuity: Each of the functions defining $\tilde{h}_{\alpha,\beta}(z)$ is continuous, and therefore, one should verify that the concatenation points satisfy

$$h_1^{\alpha,\beta}(z_1) = h_2^{\alpha,\beta}(z_1)$$
$$h_2^{\alpha,\beta}(z_2) = \tilde{\rho}_{\alpha,\beta}. \tag{61}$$

The second equality in (61) is verified as follows:

$$(1-\alpha\bar{\beta})h_2^{\alpha,\beta}(z_2)$$
$$= X(z_2) + p^{\text{opt}} X(\arg_2(z_2))$$
$$= H_2(p^{\text{opt}}) + p^{\text{opt}} H_2\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) - (\bar{z}_2 + \bar{\beta}z_2)H_2(\alpha)$$
$$- (z_2 + \alpha\bar{z}_2)H_2(\beta) - (p^{\text{opt}} + \alpha\bar{\beta})\tilde{\rho}_{\alpha,\beta}$$
$$= (1-\alpha\bar{\beta})\tilde{\rho}_{\alpha,\beta},$$

and since $(1-\alpha\bar{\beta}) \neq 0$, it follows that $h_2^{\alpha,\beta}(z_2) = \tilde{\rho}_{\alpha,\beta}$. This derivation also gives us that

$$\tilde{\rho}_{\alpha,\beta} = h_2^{\alpha,\beta}(z_2)$$
$$\stackrel{(a)}{=} \frac{X(z_2) + p^{\text{opt}}X(z_1)}{1-\alpha\bar{\beta}}, \tag{62}$$

where (a) follows from the fact that $z_1 = \arg 2(z_2)$.

The value of $h_2^{\alpha,\beta}(z_1)$ is

$$h_2^{\alpha,\beta}(z_1) = \frac{X(z_1) + \frac{\alpha\bar{\beta}}{p^{\text{opt}}}X(z_2)}{1-\alpha\bar{\beta}}$$
$$\stackrel{(a)}{=} \frac{1}{p^{\text{opt}}}[\tilde{\rho}_{\alpha,\beta} - X(z_2)]$$
$$= H_2\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) - \frac{\bar{\beta}z_2}{p^{\text{opt}}}H_2(\alpha) - \frac{\alpha\bar{z}_2}{p^{\text{opt}}}H_2(\beta)$$
$$= H_2\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) - \bar{z}_1 H_2(\alpha) - z_1 H_2(\beta) \tag{63}$$

where (a) follows from (62).

From (46), we have that

$$h_1^{\alpha,\beta}(z_1) = H_2\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) - \bar{z}_1 H_2(\alpha) - z_1 H_2(\beta)], \tag{64}$$

which together with (63) concludes the continuity of $\tilde{h}_{\alpha,\beta}(z)$ at $z = z_1$.

Concavity: First, we show that each element in $\tilde{h}_{\alpha,\beta}(z)$ is concave and then we argue that the concatenation of these functions is also concave. The function $h_1^{\alpha,\beta}(z)$ is concave since it is a composition of the binary entropy function, which is concave, with an affine function. The function $h_2^{\alpha,\beta}(z)$ can be written explicitly from (46), and then all of its elements are linear except for the entropy function which is concave and the expression $p(z)H_2\left(\frac{\alpha\bar{z}}{p(z)}\right)$. The latter expression is also concave since it is known that the perspective of the concave function $H_2(z)$, that is, $tf\left(\frac{x}{t}\right)$ for $t > 0$ is also concave. Therefore, each element comprises $\tilde{h}_{\alpha,\beta}(z)$ is concave.

It was shown in [12, Lemma 5] that a continuous concatenation of concave functions is concave if the one-sided derivatives at the concatenation points are equal. Therefore, $\tilde{h}_{\alpha,\beta}(z)$ is concave if the following conditions are satisfied:

$$h_1'(z_1) = h_2'(z_1) \tag{65}$$

$$h_2'(z_2) = 0. \tag{66}$$

An auxiliary relation can be derived using the fact that $R_{\alpha,\beta}(z)$ has a unique maximizer. Specifically, by taking the first derivative of $R_{\alpha,\beta}(z)$ and equating it to zero, we have

$$p'H_2'(p) + \left[pH_2\left(\frac{\alpha\bar{\beta}}{p}\right)\right]' - p'\tilde{\rho}_{\alpha,\beta}\Big|_{p=p^{\text{opt}}}$$
$$= -\beta H_2(\alpha) + \bar{\alpha} H_2(\beta) \tag{67}$$

The following calculations are also necessary:

$$X'(z) = p'H_2'(p) + H_2(\alpha) - H_2(\beta) - p'\tilde{\rho}_{\alpha,\beta}$$

$$X(\arg 2(z)) = H_2\left(\frac{\alpha\bar{\beta}}{p}\right) - \overline{\arg 2(z)}H_2(\alpha) - \arg 2(z)H_2(\beta)$$
$$- \frac{\alpha\bar{\beta}}{p}\tilde{\rho}_{\alpha,\beta}$$

$$X'(\arg 2(z)) = p'H_2'\left(\frac{\alpha\bar{\beta}}{p}\right) + H_2(\alpha) - H_2(\beta) - p'\tilde{\rho}_{\alpha,\beta},$$

where derivatives are taken with respect to z.

The first derivative of $h_2(z)$ is:

$$\frac{d}{dz}((1-\alpha\bar{\beta})h_2^{\alpha,\beta}(z))$$
$$= X'(z) + p'X(\arg 2(z)) + p\arg 2'(z)X'(\arg 2(z))$$
$$= p'H_2'(p) - p'\tilde{\rho}_{\alpha,\beta} + p'H_2\left(\frac{\alpha\bar{\beta}}{p}\right) - \frac{\alpha\bar{\beta}}{p}p'H_2'\left(\frac{\alpha\bar{\beta}}{p}\right)$$
$$+ p'\left[-\overline{\arg 2(z)}H_2(\alpha) - \arg 2(z)H_2(\beta)\right]$$
$$+ \left(1 - \frac{\alpha\bar{\beta}}{p}\right)[H_2(\alpha) - H_2(\beta)]. \tag{68}$$

Substituting $z = z_2$ into (68) and using (67), we obtain the desired equality in 66

$$(1-\alpha\bar{\beta})\frac{d}{dz}h_2^{\alpha,\beta}(z)\Big|_{z=z_2} = 0.$$

For the other condition, (65), one can show that $p(z_1^{\alpha,\beta}) = \frac{\alpha\bar{\beta}}{p(z_2)}$, which results in

$$(1-\alpha\bar{\beta})\frac{d}{dz}h_2^{\alpha,\beta}(z)\Big|_{z=z_1}$$
$$= (1-\alpha\bar{\beta})[p'H_2'\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) + H_2(\alpha) - H_2(\beta)]$$

The derivative of $h_1^{\alpha,\beta}(z)$ at $z = z_1$ is:

$$\frac{d}{dz}h_1^{\alpha,\beta}(z_1) = p'H_2'\left(\frac{\alpha\bar{\beta}}{p^{\text{opt}}}\right) + H_2(\alpha) - H_2(\beta),$$

and this concludes the concavity of $\tilde{h}_{\alpha,\beta}(z)$.

The last two items in Lemma C.1 follow from the concavity of $\tilde{h}_{\alpha,\beta}(z)$ and the fact that $z_1 \leq z_2$: since the maximum is at z_2, then the derivative of $h_2^{\alpha,\beta}(z)$ at z_1 is positive and equals the derivative of $h_1^{\alpha,\beta}(z)$ at z_1. ∎

REFERENCES

[1] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 9, no. 3, pp. 136–143, Jul. 1963.

[2] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1186–1222, Mar. 2011.

[3] P. Vontobel, A. Kavcic, D. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[4] G. Han and B. Marcus, "Asymptotics of entropy rate in special families of hidden Markov chains," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1287–1295, Mar. 2010.

[5] E. Zehavi and J. Wolf, "On runlength codes," *IEEE Trans. Inf. Theory*, vol. 34, no. 1, pp. 45–54, Jan. 1988.

[6] G. Han and B. Marcus, "Concavity of the mutual information rate for input-restricted memoryless channels at high SNR," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1534–1548, Mar. 2012.

[7] Y. Li and G. Han, "Input-constrained erasure channels: Mutual information and capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 3072–3076.

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

[9] S. C. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[10] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[11] H. H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2009.

[12] 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.

[13] O. Sabag, H. Permuter, and N. Kashyap, "The feedback capacity of the binary erasure channel with a no-consecutive-ones input constraint," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 8–22, Jan 2016.

[14] J. Wu and A. Anastasopoulos, "On the capacity of the chemical channel with feedback," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 295–299.

[15] A. Sharov and R. M. Roth, "On the capacity of generalized ising channels," *IEEE Trans. on Inf. Theory*, vol. 63, no. 4, pp. 2338–2356, Apr. 2017.

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

[17] C. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[18] G. Han and B. Marcus, "Asymptotics of input-constrained binary symmetric channel capacity," *Ann. Appl. Probab.*, vol. 19, no. 3, pp. 1063–1091, 2009.

[19] D. Shaviv, A. Ozgur, and H. Permuter, "Can feedback increase the capacity of the energy harvesting channel?" in *Proc. IEEE Information Theory Workshop (ITW)*, 2015.

[20] Y. Li and G. Han, "Asymptotics of input-constrained erasure channel capacity," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, pp. 1–1, 2017.

[21] J. P. M. Schalkwijk and T. Kailath, "Coding scheme for additive noise channels with feedback I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. 12, pp. 172–182, 1966.

[22] T. P. Coleman, "A stochastic control viewpoint on posterior matching feedback communication schemes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2009, pp. 1520–1524.

[23] M. Naghshvar, T. Javidi, and M. Wigger, "Extrinsic Jensen-Shannon divergence: Applications to variable-length coding," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2148–2164, April 2015.

[24] C. T. Li and A. E. Gamal, "An efficient feedback coding scheme with low error probability for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 2953–2963, June 2015.

[25] O. Shayevitz and M. Feder, "A simple proof for the optimality of randomized posterior matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3410–3418, June 2016.

[26] O. Sabag, H. Permuter, and H. Pfister, "A single-letter upper bound on the feedback capacity of unifilar finite-state channels," *IEEE Trans. on Inf. Theory*, vol. PP, no. 99, pp. 1–1, Dec. 2016.

[27] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, pp. 73–77, 1973.

[28] A. Thangaraj, "Dual capacity upper bounds for noisy runlength constrained channels," in *Proc. IEEE Information Theory Workshop (ITW), 2016 IEEE*, 2016, full version: arxiv.org/abs/1609.00189.

[29] O. Sabag, H. Permuter, and N. Kashyap, "The feedback capacity of the binary symmetric channel with a no-consecutive-ones input constraint," in *Proc. Allerton Conference Communication, Control, and Computing*, 2015.

[30] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. Marcus, "Discrete time controlled Markov processes with average cost criterion - a survey," *SIAM Journal of Control and Optimization*, vol. 31, no. 2, pp. 282–344, 1993.

[31] A. Anastasopoulos, "A sequential transmission scheme for unifilar finite-state channels with feedback based on posterior matching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2012, pp. 2914–2918.

[32] J. Wu and A. Anastasopoulos, "Zero-rate achievability of posterior matching schemes for channels with memory," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2016, pp. 2256–2260.

Oron Sabag (S'14) received his B.Sc. (cum laude) degree and his M.Sc. (summa cum laude) in Electrical and Computer Engineering from the Ben-Gurion University of the Negev, Israel, in 2013 and 2016, respectively. He is currently pursuing his Ph.D in the direct track for honor students in Electrical and Computer Engineering at the same institution.

Oron is a recipient of several awards, among them are the Dean's List Award, the Lachish Fellowship, SPCOM-2016 best student paper award, ISIT-2017 best student paper award, and a Feder Family Award for outstanding research in communications (3rd prize).

Haim H. Permuter (M'08-SM'13) 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 a professor, Luck-Hille Chair in Electrical Engineering, and head of the communication systems track.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and 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.

Navin Kashyap (S'97-M'02-SM'07) received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Bombay, in 1995, the M.S. degree in Electrical Engineering from the University of Missouri-Rolla in 1997, and the M.S. degree in Mathematics and the Ph.D. degree in Electrical Engineering from the University of Michigan, Ann Arbor, in 2001. From November 2001 to November 2003, he was a postdoctoral research associate at the University of California, San Diego. From 2004 to 2010, he was on the faculty of the Department of Mathematics and Statistics at Queen's University, Kingston, Ontario. In January 2011, he joined the Department of Electrical Communication Engineering at the Indian Institute of Science, where is currently a Professor. His research interests lie primarily in the application of combinatorial and probabilistic methods in information and coding theory. Prof. Kashyap served on the editorial board of the IEEE Transactions on Information Theory during the period 2009–2014. He is at present an Associate Editor for the SIAM Journal on Discrete Mathematics.