Feedback Capacity of a Class of Symmetric Finite-State Markov Channels

Nevroz Şen, *Student Member, IEEE*, Fady Alajaji, *Senior Member, IEEE*, and Serdar Yüksel, *Member, IEEE*

Abstract—We consider the feedback capacity of a class of symmetric finite-state Markov channels. Here, symmetry (termed "quasi-symmetry") is defined as a generalized version of the symmetry defined for discrete memoryless channels. The symmetry yields the existence of a hidden Markov noise process that depends on the channel's state process and facilitates the channel description as a function of input and noise, where the function satisfies a desirable invertibility property. We show that feedback does not increase capacity for such class of finite-state channels and that both their nonfeedback and feedback capacities are achieved by an independent and uniformly distributed (i.u.d.) input. As a result, the channel capacity is explicitly given as a difference of output and noise entropy rates, where the output is driven by the i.u.d. input.

Index Terms—Channel capacity, channels with memory, dynamic programming, feedback capacity, finite-state Markov channels.

I. INTRODUCTION AND LITERATURE REVIEW

ALTHOUGH feedback does not increase the capacity of discrete memoryless channels (DMCs) [1], it generally increases the capacity of channels with memory. In this work, we study the feedback capacity of a class of channels with memory and show that feedback does not increase their capacity. More explicitly, we consider finite-state Markov (FSM) channels [2]–[4] which encompass symmetry in their channel transition matrices.

FSM channels have been widely used to effectively model wireless fading channels (e.g., cf. [5]–[8]). A definition of symmetric finite-state Markov channels is given in [9] and [10] and capacity without feedback is calculated where it is shown that the capacity-achieving distribution is uniform and that this distribution yields a uniform output distribution. In [11], it is shown that feedback does not increase the capacity of discrete channels with modulo additive noise. It is also shown that for any channel with memory satisfying the symmetry conditions defined in [12], feedback does not increase its capacity. Recently, it has been shown that feedback does not increase the capacity of the compound Gilbert-Elliot channel [13], which is a family of FSM channels. In a related work, the capacity of finite-state indecomposable channels with side information at the transmitter is investigated [14]. In particular, it is shown that the capacity of finite-state indecomposable Markovian channels with (modulo) additive noise, where the noise is a deterministic function of the state, is not increased with the availability of side information at the transmitter. In a more recent work, it has been shown that it is possible to formulate the computation of feedback capacity as a dynamic programming problem, and therefore, it can be solved by using the value iteration algorithm under information stability conditions [15], [16]. In [17], finite-state channels with feedback, where feedback is a time-invariant deterministic function of the output samples, is considered. It is shown that if the state of the channel is known both at the encoder and the decoder then feedback does not increase capacity. In [18] and [19], directed information is used to calculate the feedback capacity of some classes of FSM channels. In particular, the channel state is assumed in [18] to be a deterministic function of the previous state and input; whereas in [19] the channel state is assumed to be a deterministic function of the output. In [20], time varying channels are modeled as FSM channels and their capacity is studied as a function of the feedback delay assuming perfect channel state information at the receiver. In addition to these results, it has also been shown that feedback does not increase the capacity for a binary erasure channel with Markovian state [21]. Although not closely related with our result, an important insight into the use of feedback in a real time causal coding context is presented in [22]. In particular, it is shown that feedback is useful in general causal coding problems of a Markov source over a noisy channel; however, it is not useful if the channel is symmetric (as defined in [22]) and memoryless.

Considering the structure in typical communication channels and the results in the literature that we presented above, it is worth to look for the most general notion of symmetry for channels with memory under which feedback does not increase capacity. With this motivation, we study the feedback capacity of a class of symmetric FSM channels, which we call "quasi-symmetric" FSM channels, and prove that feedback does not help increase their capacity. This result is shown by demonstrating that for an FSM channel satisfying the symmetry conditions defined in the paper, its feedback capacity is achieved by an independent and uniformly distributed (i.u.d.) input which implies that its non feedback capacity is also achieved by a uniform input distribution. Along this way, we first show the existence of a hidden Markov noise process, due to the symmetry characteristics of the channel, which is conditionally independent of the input given the sate. As a result, the FSM channel can be succinctly described as a function of input and noise, where the

function is an invertible map between the noise and output alphabets for a fixed input. With this fact, the feedback capacity problem reduces to the maximization of entropy of the output process. In the second step, we show that this entropy is maximized by a uniform input distribution. It should be noted that for quasi-symmetric FSM channels, uniform inputs do not necessarily yield uniform outputs; this is a key symmetry property used in previous works for showing that feedback does not increase capacity for symmetric channels with memory (e.g., [11], [12]). This second step is solved via a dynamic programming approach which shows that it is possible to learn the channel via past feedback control actions (input distributions) that affect the future input actions by modifying the induced channel that the receiver observes. We demonstrate that, when the FSM channel satisfies the condition that the column sums of its channel transition matrices are invariant with respect to the state process, it is still possible to learn the channel via past input actions; however, the optimal input distribution remains the same even with this learning step. We also note that our result intersects with [11] and [12] when the noise process in the latter works is restricted to being Markovian, stationary and irreducible. Furthermore, a by-product contribution of this work is that the channel capacity is given as a difference of the output and noise entropy rates, where the output is driven by the i.u.d. input and is also hidden Markovian. Thus, the capacity can be easily evaluated using existing algorithms for the computation of entropy and information rates in hidden Markov channels (e.g., see [23]). Finally, although our result covers a large class of discrete channels with memory, we believe that by adopting the approach of this work, it is possible to show a similar result for more general classes of both symmetric channels and asymmetric channels whose feedback capacity is achieved by an independent and identically distributed (i.i.d.) input process, both in the single user and multiple user settings.

The paper is organized as follows. We first give the definition of *quasi-symmetric* FSM channels. This will be followed by a section on their capacity with feedback. Next, we present examples of channels that satisfy the quasi-symmetry condition and hence conclude that their capacity does not increase with feedback. Finally, we end the paper with concluding remarks.

Throughout the paper, we will use the following notations. A random variable will be denoted by an upper case letter X and its particular realization by a lower case letter x. The sequence of random variables $X_1, X_2, \ldots, X_n$ will be denoted by X^n and so its realization will be x^n. We will represent a finite-state Markov source by a pair $[\mathcal{S}, P]$, where $\mathcal{S}$ is the state set and P is the state transition probability matrix. We will also be assuming that the Markov processes in the paper are stationary, aperiodic and irreducible (hence, ergodic).

II. QUASI-SYMMETRIC FINITE STATE MARKOV CHANNEL

A finite-state Markov (FSM) channel (e.g., [9], [10]) is defined by a pentad $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, \mathcal{C}]$, where $\mathcal{X}$ is the input alphabet, $\mathcal{Y}$ is the output alphabet and the Markov process $\{S_n\}_{n=1}^{\infty}$, $S_n \in \mathcal{S}$ is represented by the pair $[\mathcal{S}, P_S]$ where $\mathcal{S}$ is the state set and P_S is the state transition probability matrix. We assume that the sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are all finite. The set $\mathcal{C}$ is

a collection of transition probability distributions, $p_{\mathcal{C}}(y|x, s)$, on $\mathcal{Y}$ for each $x \in \mathcal{X}$, $s \in \mathcal{S}$. We consider the problem of communicating message $W \in \{1, 2, \ldots, 2^{nR}\}$ over the FSM channel (without or with the use of feedback) via a code of rate R and blocklength n,[1] where W is uniformly distributed over $\{1, 2, \ldots, 2^{nR}\}$ and independent of S^n. We assume that the FSM channel satisfies the following properties under both the absence and presence of feedback:

(I) Markov Property: For any integer $i \geq 1$

$$P(s_i|s^{i-1}, y^{i-1}, x^{i-1}, w) = P(s_i|s_{i-1}). \tag{1}$$

(II) For any integer $i \geq 1$

$$P(y_i|s_i, x_i, s^{i-1}, x^{i-1}, y^{i-1}, w) = p_{\mathcal{C}}(y_i|s_i, x_i). \tag{2}$$

When the channel is without feedback, we also assume that the FSM channel satisfies:

(II.b) For any integer $i \geq 1$

$$P(y^{i-1}|x^i, s^i) = P(y^{i-1}|x^{i-1}, s^{i-1}) \tag{3}$$

where $p_{\mathcal{C}}(.|.,.)$ is defined by $\mathcal{C}$. Note that properties (II) and (II.b) imply that $P(y^n|x^n, s^n) = \prod_{i=1}^{n} p_{\mathcal{C}}(y_i|s_i, x_i)$ when the channel is without feedback. Furthermore, the nonfeedback codewords X^n at the channel input are only a function of W (which is independent of S^n); hence, in the nonfeedback scenario, the channel input $\{X_i\}$ is also independent of S^n.

In this paper, we are interested in a subclass of FSM channels where the channel transition matrices, $Q^s \triangleq [p_{\mathcal{C}}(y|s, x)]_{xy}, s \in \mathcal{S}$, carry some notion of symmetry which is similar to the symmetry defined for DMCs as in the following.

Definition 1: A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_{\mathcal{C}}(y|x)]_{xy}$ is *symmetric* if the rows of Q are permutations of each other and the columns are permutations of each other [24].

Definition 2: A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_{\mathcal{C}}(y|x)]_{xy}$ is *weakly-symmetric* if the rows of Q are permutations of each other and all the column sums $\sum_x p_{\mathcal{C}}(y|x)$ are identically equal to a constant [24].

Definition 3: A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p_{\mathcal{C}}(y|x)]_{xy}$ is *quasi-symmetric* if Q can be partitioned along its columns into *weakly-symmetric* sub-arrays, $\tilde{Q}_1, \tilde{Q}_2, \ldots, \tilde{Q}_m$, with each $\tilde{Q}_i$ having size $|\mathcal{X}| \times |\mathcal{Y}_i|$, where $\mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_m = \mathcal{Y}$ and $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset$, $\forall i \neq j$ [25]. A *weakly-symmetric* sub-array is a matrix whose rows are permutations of each other and whose column sums are all identically equal to a constant.

Note that for a quasi-symmetric DMC, the rows of its entire transition matrix, Q, are also permutations of each other. It is also worth pointing out that the above quasi-symmetry notion for DMCs encompasses Gallager's symmetry definition

---

[1]Both feedback and nonfeedback codes of rate R and blocklength n, which yield up to 2^{nR} codewords $X^n \in \mathcal{X}^n$ for transmission over the channel, are explicitly defined in Section III in terms of a pair of encoding and decoding functions.

[2, p.94].[2] A simple example of a quasi-symmetric DMC can be given by the following (stochastic, i.e., with row sums equal to 1) transition matrix, Q, for which $a_1 + a_2 = 2a_3$ and $a_4 + a_5 = 2a_6$, and it can be partitioned along its columns into two weakly-symmetric sub-arrays

$$Q = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_3 & a_2 & a_1 & a_6 & a_5 & a_4 \\ a_2 & a_1 & a_3 & a_5 & a_4 & a_6 \\ a_3 & a_1 & a_2 & a_6 & a_4 & a_5 \end{bmatrix}$$

$$\tilde{Q}_1 = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_3 & a_2 & a_1 \\ a_2 & a_1 & a_3 \\ a_3 & a_1 & a_2 \end{bmatrix}, \text{ and } \tilde{Q}_2 = \begin{bmatrix} a_4 & a_5 & a_6 \\ a_6 & a_5 & a_4 \\ a_5 & a_4 & a_6 \\ a_6 & a_4 & a_5 \end{bmatrix}.$$

We can now define similar notions of symmetry for FSM channels.

Definition 4: (e.g., [9], [10]) An FSM channel is symmetric if for each state $s \in \mathcal{S}$, the rows of Q^s are permutations of each other such that the row permutation pattern is identical for all states, and similarly, if for each $s \in \mathcal{S}$ the columns of Q^s are permutations of each other with an identical column permutation pattern across all states.

Definition 5: An FSM channel is weakly-symmetric if for each state $s \in \mathcal{S}$, Q^s is weakly-symmetric and the row permutation pattern is identical for all states.

Definition 6: An FSM channel is quasi-symmetric if for each state $s \in \mathcal{S}$, Q^s is quasi-symmetric and the row permutation pattern is identical for all states.

To illustrate these definitions, let us consider the following conditional probability matrices of a two-state quasi-symmetric FSM channel with $\mathcal{X} = \{1, 2, 3, 4\}$, $\mathcal{Y} = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{S} = \{1, 2\}$

$$Q^1 = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_3 & a_2 & a_1 & a_6 & a_5 & a_4 \\ a_2 & a_1 & a_3 & a_5 & a_4 & a_6 \\ a_3 & a_1 & a_2 & a_6 & a_4 & a_5 \end{bmatrix}$$

$$Q^2 = \begin{bmatrix} a_1' & a_2' & a_3' & a_4' & a_5' & a_6' \\ a_3' & a_2' & a_1' & a_6' & a_5' & a_4' \\ a_2' & a_1' & a_3' & a_5' & a_4' & a_6' \\ a_3' & a_1' & a_2' & a_6' & a_4' & a_5' \end{bmatrix} \qquad (4)$$

where Q^1 and Q^2 are stochastic matrices. As it can be seen, Q^1 and Q^2 have the same row permutation pattern and are both quasi-symmetric.

It directly follows by definition that symmetric and weakly symmetric FSM channels are special cases of quasi-symmetric FSM channels. Therefore, we focus on quasi-symmetric FSM channels for the sake of generality.

Let us define $\mathcal{Z}$ (which will serve as a noise alphabet) such that $|\mathcal{Y}| = |\mathcal{Z}|$, where $\mathcal{Y}$ is the output alphabet. Then for each state s, since the rows of Q^s are permutations of each other

(the FSM channel being quasi-symmetric), we can find functions $f_s(.) : \mathcal{Z} \to [0, 1]$ and $\Phi_s(., .) : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ that are onto given x (i.e., for each $x \in \mathcal{X}$, $\Phi_s(x, .) : \mathcal{Y} \to \mathcal{Z}$ is onto), such that

$$f_s(\Phi_s(x, y)) = p_{\mathcal{C}}(y|x, s). \qquad (5)$$

Note that since each function $\Phi_s(x, .) : \mathcal{Y} \to \mathcal{Z}$ is onto given x and since $|\mathcal{Y}| = |\mathcal{Z}|$, then it is also one-to-one given x; i.e., $\Phi_s(x, y) = \Phi_s(x, y') \Rightarrow y = y'$. Thus, $\Phi_s(x, .) : \mathcal{Y} \to \mathcal{Z}$ is invertible for each $x \in \mathcal{X}$.

For the sake of completeness, we herein provide an explicit construction for the functions $f_s(.)$ and $\Phi_s(., .)$. The construction is basically as follows: for each (x, y) pair having identical channel conditional probability $p_{\mathcal{C}}(y|x, s)$ under state s, $\Phi_s(x, y)$ returns the same value z with $f_s(z)$ set to equal $p_{\mathcal{C}}(y|x, s)$. More explicitly, let $\mathcal{X} = \{x_{(1)}, x_{(2)}, \ldots, x_{(k)}\}$, $\mathcal{Y} = \{y_{(1)}, y_{(2)}, \ldots, y_{(|\mathcal{Y}|)}\}$, $\mathcal{Z} = \{z_{(1)}, z_{(2)}, \ldots, z_{(|\mathcal{Y}|)}\}$, $I = \{1, 2, \ldots, k\}$ and $J = \{1, 2, \ldots, |\mathcal{Y}|\}$. For $s \in \mathcal{S}$, let $q_{i,j}^s \triangleq p_{\mathcal{C}}(y_{(j)}|x_{(i)}, s)$, $i \in I$ and $j \in J$, be the entries of Q^s. Since Q^s is quasi-symmetric, then for each $i = 1, 2, \ldots, k$, there exists a permutation $\pi_i^s : J \to J$ on the column indices of the entries of the ith row of Q^s such that the first row of Q^s is a permutation of every other row.[3] Then, $f_s(.)$ and $\Phi_s(., .)$ are given as follows: $\Phi_s(x_{(i)}, y_{(j)}) = z_{(\pi_i^s(j))}$ and $f_s(z_{(j)}) = q_{1,j}^s$, $i \in I$, $j \in J$.

Lemma 1: The function $\Phi_s(., .)$, as defined above together with $f_s(.)$ to satisfy (5), is invariant with s.

Proof: It directly follows from the above construction that $\Phi_s(x_{(i)}, y_{(j)}) = z_{(\pi_i^s(j))} = z_{(\pi_i^{\tilde{s}}(j))} = \Phi_{\tilde{s}}(x_{(i)}, y_{(j)})$, $\forall s, \tilde{s} \in \mathcal{S}$ and $\forall x_{(i)} \in \mathcal{X}, y_{(j)} \in \mathcal{Y}$ since by Definition 6, $\pi_i^s(j)$ is identical for all states. ∎

Therefore, for a quasi-symmetric FSM channel, there exists a function $\Phi(., .) : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ that is invertible given x (i.e., for each $x \in \mathcal{X}$, $\Phi(x, .) : \mathcal{Y} \to \mathcal{Z}$ is invertible) such that the random variable $Z = \Phi(X, Y)$ has the conditional distribution

$$P(z|x, s) = \frac{P(z, x, s)}{P(x, s)} = \frac{P(y, z, x, s)}{P(x, s)}$$

$$= \frac{P(z|x, y, s) p_{\mathcal{C}}(y|x, s) P(x, s)}{P(x, s)}$$

$$\overset{(a)}{=} p_{\mathcal{C}}(y|x, s) = f_s(z). \qquad (6)$$

where $y = \nu(x, z)$ and $\nu(., .) : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ is the inverse of Φ in the sense that $\nu(x, \cdot) = \Phi(x, \cdot)^{-1}$ for each $x \in \mathcal{X}$, and (a) is due to the fact that $p(z|x, y, s) = 1$. This important observation

---

[2] The capacity of a quasi-symmetric DMC is achieved by a uniform input distribution and it can be expressed via a simple closed-form formula [25]: $C = \sum_{i=1}^{m} \alpha_i C_i$ where $\alpha_i \triangleq \sum_{y \in \mathcal{Y}_i} P(y|x) = $ sum of any row in $\tilde{Q}_i$, $i = 1, \ldots, m$, and $C_i = \log_2 |\mathcal{Y}_i| - H\left(\text{any row in the matrix } \frac{1}{\alpha_i} \tilde{Q}_i\right)$, $i = 1, \ldots, m$.

[3] The row permutations are as follows. The first permutation π_1 is set as the identity function: $\pi_1^s(j) = j$ for all $j \in J$. The remaining permutations for $i = 2, \ldots, k$, are given by $\pi_i^s(1) = k$ where k is the smallest integer in J for which $q_{1,k}^s = q_{i,1}^s$, and for $j = 2, \ldots, |\mathcal{Y}|$, $\pi_i^s(j) = k'$ where k' is the smallest available (not yet assigned for values $1, 2, \ldots, j - 1$) integer in J for which $q_{1,k'}^s = q_{i,j}^s$. This assignment rule is valid whether or not the rows of Q^s contain identical entries. Specifically, if the ith row of Q^s ($i \geq 2$) has d identical entries $q_{i,j_1}^s = q_{i,j_2}^s = \cdots = q_{i,j_d}^s$ with $j_1 < j_2 < \cdots < j_d$ in J, then (by the channel's row symmetry) there exist integers $l_1 < l_2 < \cdots < l_d$ in J with $q_{i,j_1}^s = q_{i,j_2}^s = \cdots = q_{i,j_d}^s = q_{1,l_1}^s = q_{1,l_2}^s = \cdots = q_{1,l_d}^s$. In this case we set: $\pi_i^s(j_t) = l_t$ for $t = 1, 2, \ldots, d$, and $\pi_i^s(j) = \bar{k}$ where $\bar{k}$ is the *unique* integer in J for which $q_{1,\bar{k}}^s = q_{i,j}^s$ for $j \in J \setminus \{j_1, j_2, \ldots, j_d\}$.

first given in [9], reduces the set of conditional probability distributions which identifies the quasi-symmetric FSM channel to an $|\mathcal{S}| \times |\mathcal{Z}|$ matrix T defined by

$$T[s, z] = f_s(z). \tag{7}$$

Therefore, for quasi-symmetric FSM channels, we have that for any n

$$P(z_n | x_n, s_n) = P(z_n | s_n) = T[s_n, z_n]. \tag{8}$$

To make this statement explicit, let us consider the FSM channel given in (4). For this channel, we can derive the functions $z = \Phi(x, y)$ and $f_s(z)$, as explicitly shown above; for e.g., we have $\Phi(1, 1) = \Phi(2, 3) = \Phi(3, 2) = \Phi(4, 2) = 1$ and $f_1(1) = a_1$ and $f_2(1) = a_1'$. Therefore, the channel conditional probabilities for each state can now be defined by Φ and the matrix T, where

$$T = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ a_1' & a_2' & a_3' & a_4' & a_5' & a_6' \end{bmatrix}.$$

Hence, the fundamental property for quasi-symmetric FSM channels is the existence of a noise process $\{Z_n\}$ given by $Z_n = \Phi(X_n, Y_n)$ such that Z_n is independent of X_n given S_n. The class of FSM channels having this property, when there is no feedback, are termed variable noise channels [10].

The features that we have developed so far are valid for any quasi-symmetric FSM channel. However, while discussing the feedback capacity of these channels we assume that the channels also satisfy the following assumption.

Assumption 1: We assume that for a fixed $y \in \mathcal{Y}$, the column sum $\sum_x f_s(\Phi_s(x, y))$ is invariant with $s \in \mathcal{S}$: $\sum_x f_s(\Phi_s(x, y)) = \sum_x f_{s'}(\Phi_{s'}(x, y)) \ \forall s, s' \in \mathcal{S}$, where $f_s(\Phi_s(x, y)) = p_C(y | x, s)$.

In other words, the assumption requires that for each output value y, the $|\mathcal{S}|$ column sums corresponding to output y in the channel transition matrices are all identical; i.e., $\forall y \in \mathcal{Y}$

$$\sum_{x \in \mathcal{X}} p_C(y | x, s_1) = \sum_{x \in \mathcal{X}} p_C(y | x, s_2) = \cdots = \sum_{x \in \mathcal{X}} p_C(y | x, s_{|\mathcal{S}|}).$$

However, for a fixed $s \in \mathcal{S}$, $\sum_x p_C(y | x, s)$ is not necessarily invariant with $y \in \mathcal{Y}$, and as such, a uniform input does not yield a uniform output in general. This requirement will be needed in our dynamic programming approach which we use to determine the optimal feedback control action (as will be seen in the next section).[4]

III. FEEDBACK CAPACITY OF QUASI-SYMMETRIC FSM CHANNELS

In this section, we will show that feedback does not increase the capacity of quasi-symmetric FSM channels defined in the previous section. By feedback, we mean that there exists a channel from the receiver to the transmitter which is noiseless

[4]Note for our main results to hold, we require the FSM channel as defined via properties (I) and (II) to be quasi-symmetric, in addition to satisfying Assumption 1.

and delayless. Thus, at any given time, all previously received outputs are unambiguously known by the transmitter and can be used for encoding the message into the next code symbol.

A feedback code with blocklength n and rate R consists of a sequence of mappings

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

for $i = 1, 2, \ldots n$ and an associated decoding function

$$\Upsilon : \mathcal{Y}^n \to \{1, 2, \ldots, 2^{nR}\}.$$

Thus, when the transmitter wants to send message $W \in \mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$, where W is uniformly distributed over $\mathcal{W}$ and is independent of S^n, it sends the codeword X^n, where $X_1 = \psi_1(W)$ and $X_i = \psi_i(W, Y^{i-1})$, for $i = 2, \ldots, n$. In the case when there is no feedback, the codeword X^n, where $X_1 = \psi_1(W)$ and $X_i = \psi_i(W)$, for $i = 2, \ldots, n$ is transmitted, and thus, a nonfeedback code is a special case of a feedback code. For a received Y^n at the channel output, the receiver uses the decoding function to estimate the transmitted message as $\hat{W} = \Upsilon(Y^n)$. A decoding error is made when $\hat{W} \neq W$. The probability of error is given by

$$P_e^{(n)} = \frac{1}{2^{nR}} \sum_{k=1}^{2^{nR}} P\{\Upsilon(Y^n) \neq W | W = k\}.$$

It should also be observed that when communicating with feedback, property (II.b) does not hold, since X_i is a function of Y^{i-1} (in addition to W); also X^n and S^n are no longer independent as X_i causally depends on Z^{i-1}, and hence, S^{i-1}, for $i = 1, 2, \ldots, n$.

The capacity with feedback, C_{FB}, is the supremum of all admissible rates; i.e., rates for which there exists sequences of feedback codes with asymptotically vanishing probability of error. The (classical) nonfeedback capacity, C_{NFB}, is defined similarly (by replacing feedback codes with nonfeedback codes). Since a nonfeedback code is a special case of a feedback code, we always have $C_{FB} \geq C_{NFB}$.

The main result of this work is as follows.

Theorem 1: The feedback capacity of a quasi-symmetric FSM channel $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$ satisfying Assumption 1 is given by

$$C_{FB} = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$$

where $\mathcal{H}(\tilde{Y})$ is the entropy rate of the output process $\{\tilde{Y}_i\}$ driven by an i.u.d. input and $\mathcal{H}(Z)$ is the entropy rate of the channel's noise (hidden Markovian) process $\{Z_i\}_{i=1}^{\infty}$.

We devote the remainder of the section to prove this theorem and deduce that feedback does not help increase the capacity of quasi-symmetric FSM channels satisfying Assumption 1.

From Fano's inequality, we have

$$H(W | Y_n) \leq h_b(P_e^{(n)}) + P_e^{(n)} \log_2(2^{nR} - 1)$$
$$\leq 1 + P_e^{(n)} nR$$

where the first inequality holds since $h_b(P_e^{(n)}) \leq 1$, where $h_b(\cdot)$ is the binary entropy function. Since W is uniformly distributed

$$
\begin{aligned}
nR = H(W) \\
= H(W|Y^n) + I(W;Y^n) \\
\leq 1 + P_e^{(n)} nR + I(W;Y^n)
\end{aligned}
$$

where R is any admissible rate. Dividing both sides by n and taking the $\liminf$ yields

$$
C_{FB} \leq \liminf_{n \to \infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W;Y^n). \tag{9}
$$

For every coding policy with feedback $\{\psi_i, 1 \leq i \leq n\}$, there are induced maps $\{\eta_i, 1 \leq i \leq n\}$ such that

$$
\eta_i : \mathcal{X}^{i-1} \times \mathcal{Y}^{i-1} \to \mathbb{P}(\mathcal{X})
$$

with

$$
\eta_i(x^{i-1}, y^{i-1}) = \left(\beta_i(x_{(1)}), \beta_i(x_{(2)}), \ldots, \beta_i(x_{(k)})\right)
$$

and

$$
\beta_i(x_{(j)}) = \sum_{w \in \mathcal{W}} P(w|x^{i-1}, y^{i-1}) 1_{\{x_{(j)} = \psi_i(w, y^{i-1})\}}
$$

for $j = 1, 2, \ldots, k$, where $\mathcal{X} = \{x_{(1)}, x_{(2)}, \ldots, x_{(k)}\}$ with $k = |\mathcal{X}|$, $1_{\{\cdot\}}$ denotes the indicator function and $\mathbb{P}(\mathcal{X})$ denotes the space of probability distributions on $\mathcal{X}$. Every η_i can also be identified by the collection of control actions at time i:

$$
\mathcal{D}_i \triangleq \{P(x_i|x^{i-1}, y^{i-1}) : x^{i-1} \in \mathcal{X}^{i-1}, y^{i-1} \in \mathcal{Y}^{i-1}\}.
$$

In view of this discussion, following [15] (see also [16] and [26]), we have

$$
\begin{aligned}
&\liminf_{n \to \infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} I(W;Y^n) \\
&= \liminf_{n \to \infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Y_i|W,Y^{i-1})\right] \\
&= \liminf_{n \to \infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Y_i|W,Y^{i-1},X^i)\right] \\
&= \liminf_{n \to \infty} \sup_{\{\psi_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},X^i)\right] \quad (10) \\
&= \liminf_{n \to \infty} \sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},X^i)\right] \quad (11)
\end{aligned}
$$

where (10) is shown in Appendix A and (11) holds since $H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, X^{i-1})$ is a function of $\{P(X_i|x^{i-1}, y^{i-1})\}$. Note that the right-hand side of (11) is the directed information whose supremum has been shown to be the feedback capacity under information stability conditions [15].

Now, let us consider the following equation

$$
\sup_{\{\mathcal{D}_i\}_{i=1}^n} \frac{1}{n} \sum_{i=1}^n \left[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},X^i)\right]. \tag{12}
$$

We next establish three Lemmas in order to prove the main contribution of the paper. In the first Lemma, we show that the term $H(Y_i|X^i, Y^{i-1})$ is equal to $H(Z_i|Z^{i-1})$, and in the other two Lemmas we show that $\sum_{i=1}^n H(Y_i|Y^{i-1})$ is maximized by uniform feedback control actions $\{P(x_i|x^{i-1}, y^{i-1})\}_{i=1}^n$.

Lemma 2: The quasi-symmetric FSM channel satisfies

$$
H(Y_i|X^i, Y^{i-1}) = H(Z_i|Z^{i-1}), \quad \forall i = 1, \ldots, n.
$$

The proof of the above lemma is given in Appendix B.

We next show that all of the conditional output entropies $H(Y_i|Y^{i-1})$ in (12) are maximized by uniform feedback control actions. We solve this problem using dynamic programming [27]. Specifically, we recast the problem of maximizing the sum of conditional output entropies over all feedback control actions, using dynamic programming. The optimization problem can be written as

$$
\max_{\{\mathcal{D}_1, \ldots, \mathcal{D}_n\}} \{H(Y_n|Y^{n-1}) + H(Y_{n-1}|Y^{n-2}) + \cdots + H(Y_1)\}. \tag{13}
$$

Let

$$
\begin{aligned}
&V_i\left(P(y^{i-1}), \mathcal{D}_1, \ldots, \mathcal{D}_{i-1}\right) \\
&= \max_{\mathcal{D}_i} \left[H(Y_i|Y^{i-1}) + V_{i+1}\left(P(y^i), \mathcal{D}_1, \ldots, \mathcal{D}_i\right)\right]
\end{aligned}
$$

where $V_{n+1}\left(P(y^n), \mathcal{D}_1, \ldots, \mathcal{D}_n\right) = 0$ and the $V_i\left(P(y^{i-1}), \mathcal{D}_1, \ldots, \mathcal{D}_{i-1}\right)$ terms are explicitly given for $i = 1, \ldots, n$ as follows:

$$
\begin{aligned}
&V_n\left(P(y^{n-1}), \mathcal{D}_1, \ldots, \mathcal{D}_{n-1}\right) = \max_{\mathcal{D}_n} H(Y_n|Y^{n-1}) \\
&V_{n-1}\left(P(y^{n-2}), \mathcal{D}_1, \ldots, \mathcal{D}_{n-2}\right) = \\
&\quad \max_{\mathcal{D}_{n-1}} \left\{H(Y_{n-1}|Y^{n-2}) + \max_{\mathcal{D}_n}\{H(Y_n|Y^{n-1})\}\right\} \\
&V_{n-2}\left(P(y^{n-3}), \mathcal{D}_1, \ldots, \mathcal{D}_{n-3}\right) = \\
&\quad \max_{\mathcal{D}_{n-2}} \left\{H(Y_{n-2}|Y^{n-3}) + \max_{\mathcal{D}_{n-1}}\left\{H(Y_{n-1}|Y^{n-2})\right. \right. \\
&\qquad\qquad\qquad\qquad\qquad \left.\left. + \max_{\mathcal{D}_n}\{H(Y_n|Y^{n-1})\}\right\}\right\} \\
&\qquad\qquad\qquad\qquad \vdots \\
&V_1 = \max_{\mathcal{D}_1} \left\{H(Y_1) + \cdots + \max_{\mathcal{D}_{n-1}}\left\{H(Y_{n-1}|Y^{n-2})\right.\right. \\
&\qquad\qquad\qquad\quad \left.\left. + \max_{\mathcal{D}_n}\{H(Y_n|Y^{n-1})\} \cdots\right\}\right\}. \tag{14}
\end{aligned}
$$

Here, $V_{i+1}\left(P(y^i), \mathcal{D}_1, \ldots, \mathcal{D}_i\right)$ denotes the reward-to-go at time i, which is the future reward generated by the control action at time i.

Thus, (13) is given by V_1 in (14), which indicates that the optimization problem is nested and dynamic. It is nested since the actions and the action outcomes, that is the realizations of the channel inputs and outputs, are available in future time stages. It is dynamic, since the control actions applied at time k affects the future reward value realizations at time stages $i > k$. Thus, an optimal selection of the actions, should maximize both the current reward $H(Y_i|Y^{i-1})$ and the reward-to-go $V_{i+1}\left(P(y^i), \mathcal{D}_1, \ldots, \mathcal{D}_i\right)$ [see (14)].

Therefore, the optimization problem turns out to be finding the best induced policies $\{\eta_i, 1 \leq i \leq n\}$; that is the best collection of functions used to generate the set of control actions $\{\mathcal{D}_i, 1 \leq i \leq n\}$ which achieve V_1. We next show that the optimal set of control actions achieving V_1 is composed of uniform input distributions for $i = 1, \ldots, n$. Toward this goal, we find a condition such that the control actions taken at times $(i-1), \ldots, 1$ do not affect the reward value attained at time i, when the control action at time i is uniform. Specifically, we find that a sufficient condition to manage this problem is requiring $\sum_x f_s(\Phi_s(x, y))$ to be invariant with $s \in \mathcal{S}$, i.e., Assumption 1. This will be explicitly shown in Lemma 4. We first have the following.

Lemma 3: For the quasi-symmetric FSM channel, each conditional output entropy $H(Y_i|Y^{i-1})$, $i = 1, \ldots, n$ in (12), given the past sets of control actions $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{i-1}$, is maximized by uniform feedback control actions:

$$\mathcal{D}_i^\star \triangleq \underset{\mathcal{D}_i}{\arg\max}\, H(Y_i|Y^{i-1}) =$$
$$\left\{ P(x_i|x^{i-1}, y^{i-1}) = \frac{1}{|\mathcal{X}|}, \forall x^{i-1} \in \mathcal{X}^{i-1}, \forall y^{i-1} \in \mathcal{Y}^{i-1} \right\} (15)$$

for all $x_i \in \mathcal{X}$ and for all $i = 1, \ldots, n$.

The proof of the above lemma is provided in Appendix C. With this Lemma, we have shown that for each i, $H(Y_i|Y^{i-1})$ is maximized by the uniform input distribution. However, this is not sufficient to conclude that the optimal set of control actions attaining V_1, i.e., the optimal set of control actions maximizing $\sum_{i=1}^{n} H(Y_i|Y^{i-1})$, consists of a sequence of uniform input distributions for $i = 1, \ldots, n$. This is because Lemma 3 only maximizes the current conditional entropy via a uniform input (that is it is optimal in a myopic sense); however, it is still possible that a nonuniform input might result in a higher value function through the rewards-to-go. Let us now look at $P(y_i|y^{i-1})$ when we apply a uniform distribution at time i (current time). We obtain using (23) that

$$P(y_i|y^{i-1})$$
$$= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i) P(x_i|x^{i-1}, y^{i-1})$$
$$\qquad P(s_i|s^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1})$$
$$\overset{(i)}{=} \frac{1}{|\mathcal{X}|} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i) P(s_i|s^{i-1})$$
$$\qquad P(x^{i-1}, s^{i-1}|y^{i-1})$$
$$= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} \sum_{s^{i-1}} p_C(y_i|x_i, s_i) P(s_i|s^{i-1})$$
$$\qquad \sum_{x^{i-1}} P(x^{i-1}, s^{i-1}|y^{i-1})$$
$$= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} \sum_{s^{i-1}} p_C(y_i|x_i, s_i) P(s_i|s^{i-1}) P(s^{i-1}|y^{i-1})$$
$$= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} p_C(y_i|x_i, s_i) P(s_i|y^{i-1})$$

where (i) is valid since $P(x_i|x^{i-1}, y^{i-1})$ is uniform. Note that the dependency on past input control actions comes through

$P(s_i|y^{i-1})$ which includes transition probabilities between states, on which we have no control.

Lemma 4: Assume that the feedback control action $P(x_i|x^{i-1}, y^{i-1})$, at (current) time i, is uniform. Then the value of the conditional entropy $H(Y_i|Y^{i-1})$, at time i, is independent of past feedback control actions at times $(i-1), \ldots, 1$ if $\sum_x f_s(\Phi_s(x, y))$ is invariant with $s \in \mathcal{S}$ (i.e., if Assumption 1 holds).

Proof: We have the following:

$$P(y_i|y^{i-1}) = \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} p_C(y_i|x_i, s_i) P(s_i|y^{i-1})$$
$$= \frac{1}{|\mathcal{X}|} \sum_{s_i} P(s_i|y^{i-1}) \sum_{x_i} p_C(y_i|x_i, s_i)$$
$$= \frac{1}{|\mathcal{X}|} \sum_{s_i} P(s_i|y^{i-1}) \underbrace{\sum_{x_i} f_s(\Phi(x_i, y_i))}.$$

Since the underbraced term is invariant with s, the proof is complete as the final sum will be $\frac{1}{|\mathcal{X}|} \sum_{x_i} f_s(\Phi(x_i, y_i))$. ∎

We have so far shown that $H(Y_i|X^i, Y^{i-1}) = H(Z_i|Z^{i-1})$ and that $\sum_{i=1}^{n} H(Y_i|Y^{i-1})$ is maximized by uniform input distributions. With these results in hand, we have thus shown the following upperbound for the feedback capacity

$$C_{FB} \leq \liminf_{n \to \infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)] \qquad (16)$$

where $H(\tilde{Y}^n)$ is the output entropy when the input is uniform.

Let us now define a Hidden Markov Process (HMP) [28] which we will use while discussing the ergodicity of the noise and output processes. An HMP is denoted by a quadruple $[\mathcal{S}, P, \mathcal{Z}, T]$ in which $[\mathcal{S}, P]$ is a Markov process and T is the observation matrix defined by (7). The non-Markov process $\{Z_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{Z}$ is called HMP and it is the noisy version of the state process observed through a DMC determined by T.

Lemma 5: For the quasi-symmetric FSM channel with feedback, the noise process is an HMP with parameters $[\mathcal{S}, P, \mathcal{Z}, T]$.

Proof: To show this result, it suffices to show that

$$P(z_i|s_i, z^{i-1}) = P(z_i|s_i).$$

Since $\{S_i\}_{i=1}^{\infty}$ is Markovian, it directly implies that $P(s_i|s_{i-1}, z^{i-1}) = P(s_i|s_{i-1})$.

Note that

$$P(z_i|s_i, z^{i-1})$$
$$= \sum_{x^{i-1}} \sum_{\{(x_i, y_i): z_i = \Phi(x_i, y_i)\}} P(y_i, x_i, x^{i-1}|s_i, z^{i-1})$$
$$\overset{(i)}{=} \sum_{x^{i-1}} P(x^{i-1}|s_i, z^{i-1})$$
$$\qquad \sum_{\{(x_i, y_i): z_i = \Phi(x_i, y_i)\}} p_C(y_i|x_i, s_i) P(x_i|x^{i-1}, s_i, z^{i-1})$$
$$\overset{(ii)}{=} \sum_{x^{i-1}} P(x^{i-1}|s_i, z^{i-1})$$

$$\sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} f_{s_i}(\Phi(x_i,y_i))P(x_i|x^{i-1},z^{i-1})$$

$$= \sum_{x^{i-1}} P(x^{i-1}|s_i,z^{i-1})f_{s_i}(z_i)$$

$$\left(\sum_{\{(x_i,y_i):z_i=\Phi(x_i,y_i)\}} P(x_i|x^{i-1},z^{i-1}) \right)$$

$$\stackrel{(iii)}{=} f_{s_i}(z_i) \stackrel{(iv)}{=} P(z_i|s_i) \tag{17}$$

where (i) follows from (2) of property (II) and the fact that $y^{i-1} = \nu(x^{i-1},z^{i-1})$ is one-to-one with z^{i-1} given x^{i-1}, (ii) is valid by (5) and by the fact that feedback input depends on (x^{i-1},z^{i-1}), (iii) is valid since each z_i is satisfied by $|\mathcal{X}|$ number of (x_i,y_i) pairs where each x_i is different and (iv) follows from (5), (6) and (8). ∎

It should also be noted that, the output process, $\{\tilde{Y}_i\}_{i=1}^\infty$, for an i.u.d. input $\{X_i\}_{i=1}^\infty$ is also an HMP since

$$P(\tilde{y}_i|s_i,\tilde{y}^{i-1}) = \sum_{x_i} P(\tilde{y}_i,x_i|s_i,\tilde{y}^{i-1})$$

$$\stackrel{(a)}{=} \sum_{x_i} p_{\mathcal{C}}(\tilde{y}_i|x_i,s_i)P(x_i|s_i,\tilde{y}^{i-1})$$

$$\stackrel{(b)}{=} \sum_{x_i} p_{\mathcal{C}}(\tilde{y}_i|x_i,s_i)P(x_i|s_i) = P(\tilde{y}_i|s_i)$$

where (a) is due to (2) and (b) is due to the fact that X_i is uniformly distributed. The channel associated with the HMP is memoryless and as such it is stationary. Therefore, since the state process is stationary and ergodic both the output and noise processes are stationary and ergodic; this is stated in the following lemma.

Lemma 6: For the quasi-symmetric FSM channel $[\mathcal{X},\mathcal{Y},\mathcal{S},P_S,Z,T,\Phi]$, the noise process is stationary ergodic. Also the output process is stationary ergodic under an i.u.d. input.

We can now complete the proof of Theorem 1 and conclude that feedback does not increase capacity for the class of quasi-symmetric FSM channels satisfying Assumption 1.

Proof of Theorem 1: With (16) we already have a converse for the feedback capacity. We need to show that this bound is achievable. We first note that by Lemma 6 the noise and output processes are stationary which imply that

$$C_{FB}$$

$$\leq \liminf_{n\to\infty} \sup_{\{P(x_i|x^{i-1},y^{i-1})\}_{i=1}^n} \frac{1}{n}\sum_{i=1}^n H(Y_i|Y^{i-1})$$

$$\qquad - H(Y_i|Y^{i-1},X^i)$$

$$= \liminf_{n\to\infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)]$$

$$= \lim_{n\to\infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)] = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z). \tag{18}$$

It is sufficient to show that the bound in (18) is achievable. We now remark that there exists a coding policy which achieves this bound. Note that since the noise process is stationary and

ergodic, it can be shown that $\mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$ is an admissible rate (e.g., see [15, Theorem 5.3] and [29, Theorem 2]). Thus

$$C_{FB} \geq \lim_{n\to\infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)] = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$$

and this completes the proof.

Corollary 1: Feedback does not increase capacity of quasi-symmetric FSM channels satisfying Assumption 1 (i.e., for which $\sum_x f_s(\Phi_s(x,y))$ is invariant with $s \in \mathcal{S}$).

Proof: The result follows by noting that a nonfeedback code is a special case of a feedback code and that the nonfeedback capacity is also achieved by uniform input distributions. This can be shown more explicitly as follows:

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n}H(\tilde{Y}^n) - \lim_{n\to\infty} \frac{1}{n}H(Z^n)$$

$$\stackrel{(i)}{=} \lim_{n\to\infty} \frac{1}{n}H(Y^n)\Big|_{P(x^n)=\frac{1}{|\mathcal{X}|^n}} - \lim_{n\to\infty} \frac{1}{n}H(Z^n)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} \sup_{P(x^n)} I(X^n;Y^n) = C_{NFB}$$

where C_{NFB} is the nonfeedback capacity and (i) is valid since the input process is i.u.d. Finally, since $C_{FB} \geq C_{NFB}$, we obtain that $C_{FB} = C_{NFB}$. ∎

IV. EXAMPLES OF QUASI-SYMMETRIC FINITE STATE MARKOV CHANNELS

In this section, we present examples of quasi-symmetric FSM channels which satisfy Assumption 1 and hence have identical feedback and nonfeedback capacities. We also provide their feedback capacity expression which, when not given in single-letter form, can be computed using existing algorithms (e.g., see [23]) for the computation of entropy rates of HMPs.

A. Gilbert-Elliot Channel (e.g., [3])

One of the widely used FSM channels is the Gilbert-Elliot channel denoted by $[\mathcal{X},\mathcal{Y},\mathcal{S},\mathcal{P},\mathcal{C}]$, where $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0,1\}$. The two states are called "bad" state and "good" state, respectively, and the state transition matrix is given by

$$P = \begin{bmatrix} 1-g & g \\ b & 1-b \end{bmatrix}$$

where $0 < g < 1, 0 < b < 1$ and in either of these two states, the channel is a binary symmetric channel (BSC) with the following transition matrixes for states $s = 0$ and $s = 1$, respectively

$$Q^0 = \begin{bmatrix} 1-p_G & p_G \\ p_G & 1-p_G \end{bmatrix}, Q^1 = \begin{bmatrix} 1-p_B & p_B \\ p_B & 1-p_B \end{bmatrix}.$$

From the above channel transition matrixes, it can be observed that the Gilbert-Elliot channel is a symmetric FSM channel by Definition 4. Then, there exists a random variable $Z = \Phi(X,Y)$ with alphabet $\mathcal{Z} = \{0,1\}$ and a function $f_s(z)$ such that, $f_0(0) = 1 - p_G$ and $f_0(1) = p_G$,

$f_1(0) = 1 - p_B$ and $f_0(1) = p_B$. Therefore, one can define the $T[s, z]$ matrix for this channel as

$$T = \begin{bmatrix} 1 - p_G & p_G \\ 1 - p_B & p_B \end{bmatrix}$$

and we obtain that $\Phi(X, Y) = X \oplus Y$, where $\oplus$ represents modulo-2 addition, and $T[s, z]$ defined above. By Corollary 1, feedback does not increase the capacity of the Gilbert-Elliot channel and it should be noted that this result is a special case of [11] and [13]. Since $|\mathcal{X}| = 2$, the feedback capacity of the Gilbert-Elliot channel can be found as

$$C_{FB} = C_{NFB} = 1 - \mathcal{H}(Z)$$

where $\mathcal{H}(Z)$ is the entropy rate of the HMP $\{Z_i\}_{i=1}^{\infty}$ and can be computed as shown in [3] or [23].

B. Discrete Modulo Additive Channel With Markovian Noise

Consider the discrete channel with a common alphabet $\mathcal{A} = \{0, 1, \ldots, q - 1\}$ for the input, output and noise processes. The channel is described by the modulo-q additive equation $Y_n = X_n \oplus Z_n$, for $n = 1, 2, 3, \ldots$, and Y_n, X_n and Z_n denotes the output, input and noise processes respectively. The noise process, $\{Z_n\}_{n=1}^{n=\infty}$, is Markovian and it is independent of the input process. It is straightforward to see that the channel transition matrix for this channel is symmetric for each state, where the state is given by the previous noise variable: $S_i = Z_{i-1}$. For simplicity, let us assume that $q = 3$ and let $\zeta_{s_i, j} \triangleq P(Z_i = j | Z_{i-1} = s_i)$ for $j \in \{0, 1, 2\}$. Then, the channel transition matrix at state s_i, Q^{s_i}, will be as follows:

$$Q^{s_i} = \begin{bmatrix} \zeta_{s_i,0} & \zeta_{s_i,1} & \zeta_{s_i,2} \\ \zeta_{s_i,2} & \zeta_{s_i,0} & \zeta_{s_i,1} \\ \zeta_{s_i,1} & \zeta_{s_i,2} & \zeta_{s_i,0} \end{bmatrix}.$$

For each state, the channel transition matrix will still be symmetric with the same row permutation order. Furthermore, it also satisfies Assumption 1 since column sums are always one. Therefore, the discrete modulo additive channel is a symmetric FSM channel with $\mathcal{A} = \{0, 1, 2\}$ and $\Phi(X, Y) = X \oplus Y$. Hence, by Corollary 1, feedback does not increase the capacity of the discrete modulo additive channel with Markovian noise. Note that for this channel uniform input gives uniform output, and therefore, feedback capacity of this channel is $C_{FB} = C_{NFB} = \log 3 - \mathcal{H}(Z) = H(Z_2|Z_1)$ where $\mathcal{H}(Z) = H(Z_2|Z_1)$ is the entropy rate of Markov noise $\{Z_i\}_{i=1}^{\infty}$. This example can be readily extended for the case of Mth order Markovian noise; in that case the state S_i is given by $S_i = (Z_{i-1}, \ldots, Z_{i-M})$ and the noise entropy rate is $\mathcal{H}(Z) = H(Z_{M+1}|Z^M)$.

This result is a special case of [11]. It has been recently extended to finite-state multiple access channels in [30].

C. Symmetric Discrete Channel With Markovian Noise

Consider a discrete, not necessarily additive, channel with Markovian noise [12]. More precisely, consider the channel given by $Y_i = f(X_i, Z_i)$ for $i = 1, 2, \ldots$ where X_i, Z_i and Y_i are the input, noise and output of the channel, respectively, and $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ is a given function. Assume also that $\{X_i\}$ and

$\{Z_i\}$ are independent from each other and the channel satisfies the following properties.[5]

1. $|\mathcal{X}| = |\mathcal{Y}| = |\mathcal{Z}| = q$.
2. Given the input x, $f(x, .)$ is one-to-one; i.e., $\forall x \in \mathcal{X}$ $f(x, z) = f(x, \bar{z}) \Rightarrow z = \bar{z}$.
3. f^{-1} exists such that $z = f^{-1}(x, y)$ and given y, $f^{-1}(., y)$ is one-to-one; i.e., $\forall y \in \mathcal{Y}$ $f^{-1}(x, y) = f^{-1}(\bar{x}, y) \Rightarrow x = \bar{x}$.

We note that a channel satisfying these conditions has a symmetric channel transition matrix for each state, where the state is given by the previous noise variable: $S_i = Z_{i-1}$. Therefore, this channel is a symmetric FSM channel with the same permutation order determined by the function f. It also satisfies Assumption 1 as the column sums are one for each state. Therefore, by Corollary 1, feedback does not increase the capacity of these channels. This result is first shown in [12], where the noise process may be non-Markovian and nonergodic in general. Similar to the previous example, uniform input yields uniform output for this channel, and therefore, feedback capacity of this channel is $C_{FB} = C_{NFB} = \log q - \mathcal{H}(Z) = \log q - H(Z_2|Z_1)$. As in the previous example, this example can be extended for the case of Mth order Markov noise.

We next present two different channels which illustrate the result of the paper when the column sums for each state are different than one.

D. Binary Channel With Erasures, Errors and Markovian State

Consider the two-state channel given by $\mathcal{X} = \{0, 1\}$, $\mathcal{S} = \{s_1, s_2\}$, where $\{S_i\}$ is Markovian, $\mathcal{Y} = \{0, E, 1\}$ with the following channel transition matrices

$$Q^{s_1} = \begin{bmatrix} 1 - \varepsilon - \xi & \xi & \varepsilon \\ \varepsilon & \xi & 1 - \varepsilon - \xi \end{bmatrix}$$

$$Q^{s_2} = \begin{bmatrix} 1 - \varepsilon' - \xi' & \xi' & \varepsilon' \\ \varepsilon' & \xi' & 1 - \varepsilon' - \xi' \end{bmatrix}$$

where $0 < \varepsilon, \xi, \varepsilon', \xi' < 1$ are fixed. We first note that this channel is a two-state quasi-symmetric FSM channel, since we can partition Q^{s_1} and Q^{s_2} in two symmetric sub-arrays given by

$$\tilde{Q}_{\mathcal{Y}_1}^{s_1} = \begin{bmatrix} 1 - \varepsilon - \xi & \varepsilon \\ \varepsilon & 1 - \varepsilon - \xi \end{bmatrix}, \quad \tilde{Q}_{\mathcal{Y}_2}^{s_1} = \begin{bmatrix} \xi \\ \xi \end{bmatrix}$$

and

$$\tilde{Q}_{\mathcal{Y}_1}^{s_2} = \begin{bmatrix} 1 - \varepsilon' - \xi' & \varepsilon' \\ \varepsilon' & 1 - \varepsilon' - \xi' \end{bmatrix}, \quad \tilde{Q}_{\mathcal{Y}_2}^{s_2} = \begin{bmatrix} \xi' \\ \xi' \end{bmatrix}$$

respectively, where $\mathcal{Y}_1 = \{0, 1\}$ and $\mathcal{Y}_1 = \{E\}$ with identical permutation order between states. For this channel, if we set $\xi = \xi'$, then we automatically satisfy Assumption 1 since the column sums in both Q^{s_1} and Q^{s_2} will be $1 - \xi$, 2ξ and $1 - \xi$ respectively. In other words, although the error probabilities are different across the states ($\epsilon \neq \epsilon'$ in general), we still have identical column sums. Therefore, by Corollary 1, feedback does not increase the capacity of this channel. Furthermore, since both the output and noise process are HMPs the value of feedback capacity can be computed using [23].

---

[5]In [12], it is stated that $|\mathcal{X}| = |\mathcal{Z}| = q$. However, following the proof, it can be evidently seen that $|\mathcal{Y}| = q$ is also assumed.

E. Nonbinary Noise Discrete Channel With Markovian Noise

We now present a binary-input 2^q-ary output communication channel with memory which was recently introduced in [31] (in the absence of feedback) with the objective of capturing both the statistical memory and the soft-decision information of time-correlated fading channels modulated via binary phase-shift keying (BPSK) and coherently demodulated with an output quantizer of resolution q. This channel, which we refer to as the nonbinary noise channel (NBNDC), has a straightforward structure and useful properties and it can help in the design of coding/decoding schemes for soft-decision demodulated channels with memory that result in superior performance over coding systems that ignore the channel's memory (via interleaving) and/or soft-decision information (via hard demodulation) [31]. The NBNDC model is explicitly described by the following equation:

$$Y_k = (2^q - 1)X_k + (-1)^{X_k} Z_k \qquad (19)$$

for $k = 1, 2, \ldots$, where $X_k \in \mathcal{X} = \{0, 1\}$ is the input, $Y_k, Z_k \in \mathcal{Z} = \mathcal{Y} = \{0, 1, \ldots, 2^q - 1\}$ is the output and the noise processes, respectively. The noise and input processes are independent from each other and we assume that the noise process is Markovian (an Mth order Markov process can also be considered as examined in [31] for modeling the underlying fading channel). For the sake of simplicity, we consider the NBNDC channel with $q = 2$. Let $\Lambda = [\lambda_{s_i, j}]_{i=1,\ldots,4; j=1,\ldots,4}$, where $\lambda_{s_i, j} \triangleq P(Z_i = j | Z_{i-1} = s_i)$, denotes the transition probability matrix of the noise process. Then, with the state $S_i = Z_{i-1}$, the channel transition matrix at state s_i, Q^{s_i}, is given by

$$Q^{s_i} = \begin{bmatrix} \lambda_{s_i,0} & \lambda_{s_i,1} & \lambda_{s_i,2} & \lambda_{s_i,3} \\ \lambda_{s_i,3} & \lambda_{s_i,2} & \lambda_{s_i,1} & \lambda_{s_i,0} \end{bmatrix}.$$

Note that NBNDC is a quasi-symmetric FSM channel but it does not necessarily satisfy Assumption 1. However, it can be easily shown that for any Λ satisfying that both $\sum_{j=0,3} \lambda_{i,j}$ and $\sum_{j=1,2} \lambda_{i,j}$ do not change with different i values, Assumption 1 is satisfied; therefore, by Corollary 1, feedback does not increase capacity of such NBNDC channels. Furthermore, the nonfeedback capacity of NBNDC is given in [31] as $C_{NFB} = 1 + \mathcal{H}(W) - H(Z_2|Z_1)$, where $\mathcal{H}(W)$ is the entropy rate of the process $\{W_k\}$ which is defined on the alphabet $\mathcal{W} = \{0, 1, \ldots, 2^{q-1} - 1\}$ with $W_k = \min\{Z_k, 2^q - 1 - Z_k\}$. Therefore, if Λ satisfies the condition that both $\sum_{j=0,3} \lambda_{i,j}$ and $\sum_{j=1,2} \lambda_{i,j}$ do not change with different i values, we then have $C_{FB} = C_{NFB} = 1 + \mathcal{H}(W) - H(Z_2|Z_1)$. Note that $\{W_k\}$ is an HMP and as such $\mathcal{H}(W)$ can be computed as shown in [23].

There is one more quasi-symmetric FSM channel that needs further attention. We now investigate how its channel properties directly satisfy the condition that the previous feedback control actions do not affect the current value of the conditional output entropy. In other words, the example below satisfies Lemma 4 without having the condition that the column sums are identical among different states, (i.e., it does not satisfy Assumption 1).

F. Simplified Binary Erasure Channel With Markovian State

Consider the following binary erasure channel [21], which is a simplified (special) case of the erasure channel of Example D

and has been used to model packet losses in a packet communication network, such as the Internet. The channel has binary input and ternary output; $\mathcal{X} = \{0, 1\}, \mathcal{Y} = \{0, E, 1\}$. Let S_i denote the state of the erasure channel when the packet i arrives such that when $S_i = 1$, the packet is erased, and when $S_i = 0$, the packet gets through. For a given input, the channel output is identical to the input if there is no erasure, and it is equal to the erasure symbol (E) if an erasure occurs. Therefore, the channel transition matrices at states $0, 1$ will be as follows:

$$Q^0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad Q^1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

This channel can be considered as a special case of deletion channel in which the erased packet is assumed to be known by the decoder. Therefore, in an erasure channel, the receiver has also the side information about the state. In [21], this channel is considered as a finite buffer queue, which can be viewed as an FSM channel, and the state of the finite buffer channel is determined by the state of the buffer and it is shown that feedback does not increase the capacity of this channel. We herein note that the approach presented in the paper gives the same result.

Proposition 1: Feedback does not increase capacity of simplified binary erasure channel with Markovian state and the feedback capacity is achieved by an i.u.d. input.

Proof: We first note that since the channel is quasi-symmetric for each state, the conditional output entropy is maximized by uniform input distributions. What we further need to show is the independence of the value attained by $H(Y_i|Y^{i-1} = y^{i-1})$ from previous input control actions. In particular, we need to show that $P(s_i|y^{i-1})$ is independent of past input control actions (see Lemma 4). It should be noted that

$$P(s_i|y^{i-1}) = \sum_{s_{i-1}} P(s_i|s_{i-1}) P(s_{i-1}|y^{i-1}).$$

Thus, given y^{i-1}, s_{i-1} is deterministic and independent of x^{i-1}. Integrating this fact in our approach proves the desired result. ∎

It has been shown that [21, Proposition 3.1] the capacity of this channel, with and without feedback, is given by $C_{FB} = C_{NFB} = (1 - p_e)$ where p_e is the erasure probability.

This particular example has the benefit of learning the state deterministically by only observing the output. We should remark that availability of both the state information and output feedback has also been considered within different setups in some other works and the situations for which feedback does not help increasing capacity are determined (see [17, Theorem 19] and [20]).

V. CONCLUSION

In this work, we presented a class of symmetric channels which encapsulates a variety of discrete channels with memory. Motivated by several results in the literature, we established a class of symmetric finite-state Markovian channels for which feedback does not increase their capacity. We showed this result by first reformulating the optimization problem in terms of dynamic programming and then proving that, under feedback, the capacity achieving distribution is uniform. An important observation should be highlighted again: when feedback exists,

one can learn the channel via the past control actions and as such may apply a nonuniform distribution which will result in a higher output entropy and capacity. We present a sufficient condition, Assumption 1, under which it is still possible to learn the channel via these past control actions; however, this learning does not affect the optimal distribution. It is also worth observing that even though we have emphasized finite-state channels with Markovian state (i.e., FSM channels) due to their wide use in the literature, our result also holds when the state process is not Markovian but still stationary ergodic.[6] Finally, although this result covers a large class of discrete channels with memory, we believe that by adopting the approach of this work, it is possible to show a similar result for a further general class of both symmetric and asymmetric channels whose feedback capacity is achieved by an i.i.d. input, both in the single user and multiple user settings.

APPENDIX A
PROOF OF (10)

1) Proof: We need to show that $P(y_i|w, x^i, y^{i-1}) = P(y_i|x^i, y^{i-1})$ for $i = 1, 2, \ldots, n$. Note that

$$
\begin{aligned}
&P(y_i|w, x_i, x^{i-1}, y^{i-1}) \\
&\overset{(a)}{=} \sum_{s_i} p_C(y_i|x_i, s_i) \frac{P(x_i|x^{i-1}, y^{i-1})P(s_i, w, x^{i-1}, y^{i-1})}{P(x_i|x^{i-1}, y^{i-1})P(w, x^{i-1}, y^{i-1})} \\
&= \sum_{s_i} p_C(y_i|x_i, s_i) P(s_i|w, x^{i-1}, y^{i-1}) \\
&\overset{(b)}{=} \sum_{s_i} p_C(y_i|x_i, s_i) P(s_i|x^{i-1}, y^{i-1}) \\
&= P(y_i|x^i, y^{i-1})
\end{aligned}
$$

where (a) follows from property (II), and (b) is valid since

$$
\begin{aligned}
P(s_i|w, x^{i-1}, y^{i-1}) &\overset{(i)}{=} \sum_{s_{i-1}} P(s_i|s_{i-1}) P(s_{i-1}|w, x^{i-1}, y^{i-1}) \\
&\overset{(ii)}{=} \sum_{s_{i-1}} P(s_i|s_{i-1}) P(s_{i-1}|x^{i-1}, y^{i-1}) \\
&= P(s_i|x^{i-1}, y^{i-1})
\end{aligned}
$$

where (i) holds by the channel Markovian property (I). Furthermore, (ii) can be shown recursively as follows:

$$
\begin{aligned}
P(s_1|w, x_1, y_1) &= \frac{P(s_1, w, x_1, y_1)}{\sum_{s_1} P(s_1, w, x_1, y_1)} \\
&= \frac{P(y_1|x_1, s_1)P(x_1, s_1, w)}{\sum_{s_1} P(y_1|x_1, s_1)P(x_1, s_1, w)} \\
&\overset{(iii)}{=} \frac{P(y_1|x_1, s_1)P(s_1)P(x_1, w)}{\sum_{s_1} P(y_1|x_1, s_1)P(s_1)P(x_1, w)} \\
&= \frac{P(y_1|x_1, s_1)P(s_1|x_1)}{\sum_{s_1} P(y_1|x_1, s_1)P(s_1|x_1)} \\
&= P(s_1|x_1, y_1) \quad\quad\quad\quad (20)
\end{aligned}
$$

---

[6] In this case, property (I) is modified by replacing $P(s_i|s_{i-1})$ with $P(s_i|s^{i-1})$ and the noise process is no longer an HMP but remains stationary ergodic.

where (iii) is valid since s_1 is independent of w and x_1 (as x_1 is only a function of w). Similarly

$$
\begin{aligned}
&P(s_2|w, x^2, y^2) \\
&= \frac{P(s_2, w, x^2, y^2)}{\sum_{s_2} P(s_2, w, x^2, y^2)} \\
&= \frac{P(y_2|x_2, s_2)P(x_1, x_2, y_1, s_2, w)}{\sum_{s_2} P(y_2|x_2, s_2)P(x_1, x_2, y_1, s_2, w)} \\
&\overset{(iv)}{=} \frac{P(y_2|x_2, s_2)P(x_2|x_1, y_1, w)P(s_2, x_1, y_1, w)}{\sum_{s_2} P(y_2|x_2, s_2)P(x_2|x_1, y_1, w)P(s_2, x_1, y_1, w)} \\
&= \frac{P(y_2|x_2, s_2)P(s_2|x_1, y_1, w)P(x_1, y_1, w)}{\sum_{s_2} P(y_2|x_2, s_2)P(s_2|x_1, y_1, w)P(x_1, y_1, w)} \\
&\overset{(v)}{=} \frac{P(y_2|x_2, s_2)\sum_{s_1} P(s_2|s_1)P(s_1|x_1, y_1)}{\sum_{s_2} P(y_2|x_2, s_2)\sum_{s_1} P(s_2|s_1)P(s_1|x_1, y_1)} \\
&= P(s_2|x^2, y^2)
\end{aligned}
$$

where (iv) is valid since x_2 is a function of x_1, y_1 and w and (v) is due to (20). Using these steps recursively for $i = 1, 2, \ldots, n$ yields (ii) and completes the proof. ∎

APPENDIX B
PROOF OF LEMMA 2

1) Proof: The proof is composed of two steps. In the first step we show that $H(Y_i|X^i, Y^{i-1}) = H(Z_i|Z^{i-1}, X^i)$. Following this, we show that $Z_i \to Z^{i-1} \to X^i$ form a Markov chain. Observe that

$$
P(y_i|x^i, y^{i-1}) \overset{(i)}{=} P(z_i|x^i, y^{i-1}) \overset{(ii)}{=} P(z_i|x^i, y^{i-1}, z^{i-1})
$$
$$
\overset{(iii)}{=} P(z_i|x^i, z^{i-1})
$$

where (i) and (ii) is valid since $z^i = \Phi(x^i, y^i)$ and (iii) is valid since $y_i = \nu(x_i, z_i)$ where $\nu(x, \cdot) = \Phi^{-1}(x, \cdot)$. This completes the first step. We next show that $Z_i \to Z^{i-1} \to X^i$ form a Markov chain. Note that

$$
\begin{aligned}
P(z_i|x^i, z^{i-1}) &= \frac{P(z_i, x^i, z^{i-1})}{P(x^i, z^{i-1})} \\
&\overset{(iv)}{=} \frac{P(x_i|x^{i-1}, z^{i-1})P(x^{i-1}, z_i, z^{i-1})}{P(x_i|x^{i-1}, z^{i-1})P(x^{i-1}, z^{i-1})} \\
&= P(z_i|x^{i-1}, z^{i-1})
\end{aligned}
$$

where (iv) is valid since the feedback input depends (causally) only on (x^{i-1}, y^{i-1}), or equivalently on (x^{i-1}, z^{i-1}). Similarly, we get

$$
\begin{aligned}
P(z_i|x^{i-1}, z^{i-1}) &= \frac{P(x_{i-1}|x^{i-2}, z^{i-2})P(x^{i-2}, z_i, z^{i-1})}{P(x_{i-1}|x^{i-2}, z^{i-2})P(x^{i-2}, z^{i-1})} \\
&= P(z_i|x^{i-2}, z^{i-1}).
\end{aligned}
$$

Using these steps recursively, we get

$$
P(z_i|x^{i-2}, z^{i-1}) = \frac{P(z_i, x^{i-2}, z^{i-1})}{P(x^{i-2}, z^{i-1})}
$$

$$
\overset{(v)}{=} \frac{P(x_{i-2}|x^{i-3}, z^{i-3})P(z_i, x^{i-3}, z^{i-1})}{P(x_{i-2}|x^{i-3}, z^{i-3})P(x^{i-3}, z^{i-1})}
$$

$$
\vdots
$$

$$
\overset{(vi)}{=} \frac{P(x_2|x_1, z_1)P(z_i, x_1, z^{i-1})}{P(x_2|x_1, z_1)P(x_1, z^{i-1})}
$$

$$
\overset{(vii)}{=} \frac{P(x_1)P(z_i, z^{i-1})}{P(x_1)P(z^{i-1})} = P(z_i|z^{i-1})
$$

where (v), (vi) and (vii) are valid due to the same reasoning above. ∎

APPENDIX C
PROOF OF LEMMA 3

1) Proof: Let us first write the conditional output entropy $H(Y_i|Y^{i-1})$ as

$$
H(Y_i|Y^{i-1}) = \sum_{y^{i-1}} P(y^{i-1}) H(Y_i|Y^{i-1} = y^{i-1}) \quad (21)
$$

where

$$
H(Y_i|Y^{i-1} = y^{i-1}) = -\sum_{y_i} P(y_i|y^{i-1}) \log P(y_i|y^{i-1}). \quad (22)
$$

To show that $H(Y_i|Y^{i-1})$ in (21) is maximized by a uniform input distribution, it is enough to show that such a uniform distribution maximizes each of the $H(Y_i|Y^{i-1} = y^{i-1})$ terms.

We now expand $P(y_i|y^{i-1})$ as follows:

$$
P(y_i|y^{i-1})
$$
$$
= \sum_{x_i} \sum_{x^{i-1}} \sum_{s_i} \sum_{s^{i-1}} P(y_i, x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1})
$$
$$
= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i, x^{i-1}, s^{i-1}, y^{i-1})
$$
$$
\qquad P(x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1})
$$
$$
\overset{(i)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i) P(x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1})
$$
$$
= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i) P(x_i, x^{i-1}, s^{i-1}|y^{i-1})
$$
$$
\qquad P(s_i|x_i, x^{i-1}, s^{i-1}, y^{i-1})
$$
$$
\overset{(ii)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i) P(s_i|s^{i-1})
$$
$$
\qquad P(x_i|x^{i-1}, s^{i-1}, y^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1})
$$
$$
\overset{(iii)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} p_C(y_i|x_i, s_i)
$$
$$
\qquad P(x_i|x^{i-1}, y^{i-1}) P(s_i|s^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1}) \quad (23)
$$

where (i) follows by (2), (ii) is valid due to the property (I) and finally (iii) is due to the fact that the feedback input depends only on (x^{i-1}, y^{i-1}).

The key observation in (23) is the existence of an equivalent channel. More specifically, $\sum_{s_i} P(y_i|x_i, s_i) P(s_i|s^{i-1})$ actually represents a quasi-symmetric channel transition matrix such that its entries are determined by the entries of the channel transition matrices of each state and the transition distribution of state probabilities. To continue, by (5)

$$
P(y_i|y^{i-1}) = \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} f_{s_i}(\Phi_{s_i}(x_i, y_i)) P(s_i|s^{i-1})
$$
$$
\qquad P(x_i|x^{i-1}, y^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1}). \quad (24)
$$

By definition of quasi-symmetry, there exists m weakly symmetric sub-arrays in the channel transition matrix at each state s_i. Among these sub-arrays, let us pick $\tilde{Q}_j^{s_i}$ of size $|\mathcal{X}| \times |\mathcal{Y}_j|$. (We assume that the partition of $\mathcal{Y}$ is identical across all states.) Let y_{j_t}, for $t = 1, \ldots, |\mathcal{Y}_j|$, denote the output values in sub-array j. Therefore, we obtain

$$
P(y_{j_t}|y^{i-1}) = \sum_{x_i, x^{i-1}, s_i, s^{i-1}} f_{s_i}(\Phi_{s_i}(x_i, y_{j_t})) P(s_i|s^{i-1})
$$
$$
\qquad P(x_i|x^{i-1}, y^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1}). \quad (25)
$$

We desire to maximize (21) over the feedback control actions $P(X_i|X^{i-1}, Y^{i-1})$. To be more precise in the following lines of equations, for $\mathcal{X} = \{x_{(1)}, \ldots, x_{(k)}\}$ with $k = |\mathcal{X}|$, let $\kappa(i-1) = P(s_i|s^{i-1})$, $\chi(i-1) = P(x^{i-1}, s^{i-1}|y^{i-1})$ and denote the feedback control actions by

$$
P(X_i = x_{(l)}|x^{i-1}, y^{i-1}) = \varphi_i(x_{(l)}), \ \text{ for } \ l = 1, \ldots, k. \quad (26)
$$

Then, for $t = 1, \ldots, |\mathcal{Y}_j|$ we can write

$$
P(y_{j_1}|y^{i-1}) = \sum_{s^{i-1}, x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1)
$$
$$
\{\varphi_i(x_{(1)}) f_{s_i}(\Phi_{s_i}(x_{(1)}, y_{j_1})) + \cdots +
$$
$$
\qquad \varphi_i(x_{(k)}) f_{s_i}(\Phi_{s_i}(x_{(k)}, y_{j_1}))\}.
$$
$$
P(y_{j_2}|y^{i-1}) = \sum_{s^{i-1}, x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1)
$$
$$
\{\varphi_i(x_{(1)}) f_{s_i}(\Phi_{s_i}(x_{(1)}, y_{j_2})) + \cdots +
$$
$$
\qquad \varphi_i(x_{(k)}) f_{s_i}(\Phi_{s_i}(x_{(k)}, y_{j_2}))\}
$$
$$
P(y_{j_{|\mathcal{Y}_j|}}|y^{i-1}) = \sum_{s^{i-1}, x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1)
$$
$$
\{\varphi_i(x_{(1)}) f_{s_i}(\Phi_{s_i}(x_{(1)}, y_{j_{|\mathcal{Y}_j|}})) + \cdots +
$$
$$
\qquad \varphi_i(x_{(k)}) f_{s_i}(\Phi_{s_i}(x_{(k)}, y_{j_{|\mathcal{Y}_j|}}))\}.
$$

It should be noted that, each $f_{s_i}(\Phi_{s_i}(x_{(l)}, y_{j_t}))$ in the equations above corresponds to an entry in the channel transition matrix $\tilde{Q}^{s_i}$ at state s_i. We also know that, the rows of the sub-array $\tilde{Q}_j^{s_i}$ are permutations of each other. In other words, each $f_{s_i}(\Phi_{s_i}(x_{(l)}, y_{j_t}))$ value appears exactly k times (once in each row) in the sub-array $Q_j^{s_i}$. Thus, the feedback control action $\varphi_i(x_{(l)})$ is multiplied by a different $f_{s_i}(\Phi_{s_i}(x_{(l)}, y_{j_t}))$ value for

each $t = 1, \ldots, |\mathcal{Y}_j|$ in the $P(y_{j_t}|y^{i-1})$ given above. Therefore, $\sum_{t=1}^{|\mathcal{Y}_j|} P(Y_i = y_{j_t}|y^{i-1})$ is equal to

$$\sum_{t=1}^{|\mathcal{Y}_j|} P(Y_i = y_{j_t}|y^{i-1})$$

$$= \sum_{s^{i-1},x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{l=1}^{k} \varphi_i(x_{(l)})$$
$$\sum_{t=1}^{|\mathcal{Y}_j|} f_{s_i}(\Phi_{s_i}(x_{(l)}, y_{j_t}))$$

$$= \sum_{s^{i-1},x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{l=1}^{k} \varphi_i(x_{(l)})$$
$$\sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i) \qquad (27)$$

$$= \sum_{s^{i-1},x^{i-1}} \chi(i-1) \sum_{s_i} \kappa(i-1) \sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i) \quad (28)$$

where (27) follows since $f_s(\Phi_s(x,y)) = p_{\mathcal{C}}(y|x,s)$ and (28) is valid since each rows in the channel transition matrix are permutations of each other and as such $\sum_{t=1}^{|\mathcal{Y}_j|} p_c(y_{j_t}|x_{(l)}, s_i)$ is identical for each $x_{(l)}$, and finally noting that $\sum_{l=1}^{k} \varphi_i(x_{(l)}) = 1$ verifies (28). The critical observation is that the value attained by (28) is independent of the feedback control actions. Similarly, for all the other $m - 1$ sub-arrays, their conditional output sums will be independent of the feedback control actions. Let us denote these sums by $\Omega_1, \ldots, \Omega_m$. More specifically for sub-array j, let $\Omega_j = \sum_{t=1}^{|\mathcal{Y}_j|} P(Y_i = y_{i_t}|y^{i-1})$. Then the maximization of (22) now becomes

$$\underset{\Omega_{j,t}}{\operatorname{argmax}} - \sum_{j=1}^{m} \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t} \qquad (29)$$

where $\sum_{j=1}^{m} \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} = 1$ and $\Omega_{j,t} = P(Y_i = y_{j_t}|y^{i-1})$, $t = 1, \ldots, |\mathcal{Y}_j|$, $j = 1, \ldots, m$. For each sub-array j, we need to find the $\Omega_{j,t}$ values that maximize $\sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t}$. By the log-sum inequality, we have that

$$- \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \Omega_{j,t} \leq - \sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t} \log \frac{\sum_{t=1}^{|\mathcal{Y}_j|} \Omega_{j,t}}{|\mathcal{Y}_j|} \qquad (30)$$

with equality if and only if

$$\Omega_{j,t} = \Omega_{s,w} \quad \forall s, w \in \{1, \ldots, |\mathcal{Y}_j|\}. \qquad (31)$$

In other words, for the sub-array j, the conditional entropy is maximized if and only if the conditional output probabilities in this sub-array are identical. Since this fact is valid for the other sub-arrays, to maximize the conditional entropy we need to (31) to be valid for all sub-arrays.

At this point, we have shown that the conditional output entropy is maximized if the conditional output probabilities are identical for each sub-array. In order to complete this step, we have to show that this is achieved by uniform input distributions.

Now, let us consider two conditional output probabilities, $P(Y_i = y_{j_s}|y^{i-1})$ and $P(Y_i = y_{j_t}|y^{i-1})$, in sub-array j. Then $P(Y_i = y_{j_s}|y^{i-1}) = P(Y_i = y_{j_t}|y^{i-1})$ which implies that

$$\sum_{l=1}^{k} \varphi_i(x_{(l)}) f_{s_i}(\Phi(x_{(l)}, y_{j_s})) = \sum_{l=1}^{k} \varphi_i(x_{(l)}) f_{s_i}(\Phi(x_{(l)}, y_{j_t})). \qquad (32)$$

However, for a fixed output $\sum_{l=1}^{k} f_{s_i}(\Phi(x_{(l)}, y_{j_s}))$ is equal to the sum of the column corresponding to output y_{j_s} (similarly for y_{j_t}) and since sub-array j is weakly symmetric, the column sums are identical. Therefore, (32) can be achieved if $\varphi_i(x_{(l)}) = \varphi_i(x_{(m)}) = \frac{1}{k}$ $\forall$ $l, m = 1, \ldots, k$, by which we get $P(Y_i = y_{j_s}|y^{i-1}) = P(Y_i = y_{j_t}|y^{i-1}) = \frac{1}{|\mathcal{X}|} \sum_{l=1}^{k} f_{s_i}(\Phi(x_{(l)}, y_{j_s}))$. Thus, for other sub-arrays since they are also weakly-symmetric, the uniform feedback control action will also satisfy the equivalence of conditional output probabilities. ∎

REFERENCES

[1] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, pp. 8–19, 1956.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] M. Mushkin and I. B. David, "Capacity and coding for the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.

[4] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 221–227, Apr. 1967.

[5] W. Turin, *Performance Analysis of Digital Transmission Systems*. New York: Computer Science Press, 1990.

[6] H. S. Wang and N. Moayeri, "Finite-state Markov channel-A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[7] C. Pimentel, T. H. Falk, and L. Lisbôa, "Finite-state Markov modeling of correlated Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1491–1501, Sep. 2004.

[8] L. Zhong, F. Alajaji, and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 79–89, Jan. 2008.

[9] M. Rezaeian, "Symmetric characterization of finite state Markov channels," presented at the IEEE Int. Symp. Information Theory, Seattle, WA, Jul. 2006.

[10] A. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[11] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 546–549, Mar. 1995.

[12] F. Alajaji and T. Fuja, "Effect of feedback on the capacity of discrete additive channels with memory," presented at the IEEE Int. Symp. Information Theory, Trondheim, Norway, Jun. 1994.

[13] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3629–3644, Aug. 2009.

[14] F. Jelinek, "Determination of capacity achieving input probabilities for a class of finite state channels with side information," *Inform. Control*, vol. 9, pp. 101–129, 1966.

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

[16] S. C. Tatikonda, "Control Under Communication Constraints," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 2000.

[17] H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

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

[19] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780–798, Mar. 2005.

[20] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.

[21] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, 3, no. 3, pp. 1226–1237, Mar. 2006.

[22] J. C. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 814–820, Nov. 1983.

[23] D. M. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3507, Aug. 2006.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[25] F. Alajaji, *Lecture Notes in Information Theory*. Kingston, ON, Canada: Queen's Univ., 1999.

[26] G. Kramer, "Directed Information for Channels with Feedback," Ph.D. dissertation, ETH, Zurich, Switzerland, 1998.

[27] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Belmont, MA: Athena Scientific, 2001.

[28] Y. Ephraim and N. Merhav, "Hidden Markov Processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.

[29] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[30] H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455–2477, Jun. 2009.

[31] C. Pimentel and F. Alajaji, "A discrete channel model for capturing memory and soft-decision information: A capacity study," in *Proc. IEEE Int. Conf. Commun.*, Dresden, Germany, Jun. 2009, pp. 1–6.

Nevroz Şen (S'03) received the B.Sc. degree in electrical and electronics engineering from the Middle East Technical University, Turkey, 2003, and the M.Eng. degree in mathematics and engineering from Queen's University, Kingston, ON, Canada, in 2009, where he is currently pursing the Ph.D. degree in mathematics and engineering.

In 2004–2008, he was with the research and development department of Nortel NETAŞ, Turkey, where he worked as a software design engineer. His research interests include information theory and stochastic control.

Fady Alajaji (S'90–M'94–SM'00) received the B.E. degree with distinction from the American University of Beirut, Lebanon, and the M.Sc. and Ph.D. degrees from the University of Maryland, College Park, all in electrical engineering, in 1988, 1990, and 1994, respectively. He held a postdoctoral appointment in 1994 at the Institute for Systems Research, University of Maryland.

In 1995, he joined the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, where he is currently a Professor of mathematics and engineering. Since 1997, he has also been cross-appointed in the Department of Electrical and Computer Engineering at the same university. His research interests include information theory, digital communications, error control coding, joint source-channel coding, and data compression.

Dr. Alajaji currently serves as Area Editor and Editor for Source and Source-Channel Coding for the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as co-chair of the 1999 Canadian Workshop on Information Theory, co-chair of the Technical Program Committee (TPC) of the 2004 Biennial Symposium on Communications, and a TPC member of several international conferences and workshops. He received the Premier's Research Excellence Award from the Province of Ontario.

Serdar Yüksel (M'11) received the B.Sc. degree in electrical and electronics engineering from Bilkent University, Turkey, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana-Champaign, Urbana, in 2003 and 2006, respectively.

He was a postdoctoral researcher at Yale University, New Haven, CT, for a year before joining Queen's University, Kingston, ON, Canada, as an Assistant Professor of mathematics and engineering in the Department of Mathematics and Statistics. His research interests are in stochastic and decentralized control, information theory, and applied probability.

Dr. Yüksel serves on the IFAC (International Federation of Automatic Control) Committee on Stochastic Systems.