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

Nevroz Şen, Fady Alajaji and Serdar Yüksel
Department of Mathematics and Statistics
Queen's University
Kingston, ON K7L 3N6, Canada
Email: nsen, fady, yuksel@mast.queensu.ca

***Abstract*—In this paper we consider the feedback capacity of a class of symmetric finite-state Markov channels. For this type of channels, symmetry is a generalized version of the symmetry defined for discrete memoryless channels. We show that feedback does not increase capacity for such class of finite-state channels. We indeed demonstrate that for such channels, both non-feedback and feedback capacities are achieved by a uniform i.i.d. input distribution.**

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 and the feedback capacity for a wide class of channels still encapsulates many open problems. Among them, we herein study finite-state Markov (FSM) channels. This work is mainly motivated by the results on the feedback capacity of channels which exhibit some notion of symmetry. Alajaji in [2], showed that feedback does not increase the capacity of discrete channels with modulo additive noise. It was also shown that for any channel satisfying the symmetry conditions defined in [3], feedback does not increase its capacity.

A definition of symmetric finite-state Markov channels is given in [4] and [5] and capacity without feedback is calculated. It is shown that the capacity-achieving distribution is uniform and that this distribution yields a uniform output. Recently, it has been shown that feedback does not increase the capacity of the compound Gilbert-Elliot channel [6], which is a family of FSM channels, where the capacity is achieved by applying a uniform input. In a closely related work, Jelinek investigated the capacity of finite-state indecomposable channels with side information at the transmitter [7]. In particular, he showed that the capacity of finite-state Markovian indecomposable 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 feedback capacity as a dynamic programming problem and therefore an approximate solution can be found by using the value iteration algorithm [8]. In [9], finite-state channels with feedback, where feedback is a time-invariant deterministic function of the output samples, is considered. It was shown that if the state of the channel is known both at the encoder and the decoder then feedback does not increase capacity. In addition to these results, it has also been shown that feedback does not increase the capacity for the binary erasure channel [10] and a discrete binary-input non-binary output channel [11] used to model soft-decision demodulated discrete fading channels with memory [12].

From control theory point of view, Walrand and Varaiya [13] presented an important insight in the use of feedback in a real time casual coding context. In particular, they showed that feedback is useful in general casual coding problems, however, it is not useful if the channel is symmetric as defined in their paper and memoryless.

These results prompt one to look for the most general case of symmetry under which feedback does not increase capacity. With this motivation, we study the feedback capacity of a class of (point-to-point) symmetric FSM channels and prove that feedback does not increase their capacity.

In the rest of this paper, we first give the definition of "quasi-symmetric" FSM channels. This will be followed by a section on their capacity with feedback. Then, we discuss some channels that satisfy all the conditions presented in the paper and hence their capacity does not increase with feedback. Following this, we present two types of channels which do not fully satisfy the conditions presented in the paper yet the presented approach in the paper is still applicable. We end the paper with concluding remarks.

Throughout the paper, we will be using 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 $X_i, X_{i+1}, ..., X_n$ will be denoted by ${X_i}^n$ and so its realization will be ${x_i}^n$. We will represent a 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 and irreducible.

II. Quasi-Symmetric Finite State Markov Channel

A finite-state Markov channel (FSMC) [4] 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

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{S}$ are all finite. $\mathcal{C}$ defines a transition probability distribution, $p_{\mathcal{C}}(y|x,s)$, over $\mathcal{Y}$ for each $x \in \mathcal{X}$, $s \in \mathcal{S}$. We also assume that the FSM channel satisfies the following three properties when there is no feedback:

I1 States and inputs are independent of each other:

$$P(s_n|x_n) = P(s_n) \;\; \forall n \geq 1. \tag{1}$$

I2 State is Markovian,

$$P(s_n|s^{n-1}, y^{n-1}, x^{n-1}) = P(s_n|s_{n-1}) \;\; \forall n \geq 1. \tag{2}$$

I3 For any integer $n \geq 1$,

$$P(y^n|s^n, x^n) = \prod_{i=1}^{n} p_{\mathcal{C}}(y_i|s_i, x_i) \tag{3}$$

where $p_{\mathcal{C}}(.|.,.)$ is defined by $\mathcal{C}$. It should be noted that, the property $I3$ exhibits a memoryless property for the channel if the state is known. Property $I2$ and $I3$ imply that:

$$P(y_n|s_n, x_n, s^{n-1}, x^{n-1}, y^{n-1}) = p_{\mathcal{C}}(y_n|s_n, x_n). \tag{4}$$

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

Definition 1: A DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and channel transition matrix $Q = [p(y|x)]$ is *quasi-symmetric* if Q can be partitioned along its columns into *weakly-symmetric* sub-arrays, $Q_1, Q_2, \ldots, Q_m$, with each Q_i having size $|\mathcal{X}| \times |\mathcal{Y}_i|$ where $\mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_n = \mathcal{Y}$ and $\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset \; \forall i \neq j$ [14]. 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.

It should be noted that for a quasi-symmetric DMC the rows of its entire transition matrix, Q, are also permutations of each other. A simple example of a quasi symmetric DMC can be given by the following transition matrix

$$Q = \begin{pmatrix} a & b & c & d \\ d & c & b & a \end{pmatrix}$$

which can be partitioned along its columns into two sub-arrays

$$Q_1 = \begin{pmatrix} a & d \\ d & a \end{pmatrix} \text{ and } Q_2 = \begin{pmatrix} b & c \\ c & b \end{pmatrix}.$$

We can now give various definitions of symmetry for FSM channels.

Definition 2: 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 3: 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 4: 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 make these definitions clear, let us consider the following conditional probability matrices of a two-state quasi-symmetric FSM channel:

$$Q^1 = \begin{pmatrix} a & b & c & d \\ d & c & b & a \end{pmatrix}, \;\; Q^2 = \begin{pmatrix} a' & b' & c' & d' \\ d' & c' & b' & a' \end{pmatrix}. \tag{5}$$

As it can be seen, Q^1 and Q^2 have the same permutation orders. It directly follows 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.

A question may appear regarding the partition of the channel transition matrix for a case in which there exists more than one partition. It should be noted that the unique permutation order between the states can generate a common partition on each of the the quasi-symmetric channel transition probability matrices $Q^s, \;\; s \in \mathcal{S}$.

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, the symmetry definitions above imply that for each state s, we can find functions $f_s(.) : \mathcal{Z} \to [0,1]$ and $\Phi_s(.,.) : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, such that

$$f_s(\Phi_s(x,y)) = p_{\mathcal{C}}(y|x,s). \tag{6}$$

Lemma 1: The function $\Phi_s(.,.)$ is invariant with s.

Proof: $\Phi_s(.,.)$ is a function which takes the row and column position of a matrix as the input and outputs a noise value. We need to show that $\Phi_{s_i}(x,y) = \Phi_{s_j}(x,y)$ $\forall i,j \in \{1,2,\cdots,|\mathcal{S}|\}$ and $\forall x \in \mathcal{X}, y \in \mathcal{Y}$. Rows of a quasi-symmetric channel transition matrix are permutations of each other. Therefore,

$$\Phi_{s_i}(x_1, y_{i_1}) = \Phi_{s_i}(x_2, y_{i_2}) = \cdots = \Phi_{s_i}(x_{|\mathcal{X}|}, y_{i_k}) = z^*,$$

where $k = |\mathcal{X}|$. By the unique order of row permutation between states, we have that

$$\Phi_{s_j}(x_1, y_{i_1}) = \Phi_{s_j}(x_2, y_{i_2}) = \cdots = \Phi_{s_j}(x_{|\mathcal{X}|}, y_{i_k}) = z^*$$

which implies that $\Phi_{s_j} = \Phi_{s_i}$ ■

Therefore, for a quasi-symmetric FSM channel, there exists a function $\Phi(.,.)$ such that the random variable $Z = \Phi(X,Y)$ has the conditional distribution

$$p(z|x,s) = f_s(z). \tag{7}$$

This important observation given in [4], reduces the set of conditional probabilities which identifies the quasi-symmetric FSM channel to an $|\mathcal{S}| \times |\mathcal{Z}|$ matrix T defined by

$$T[s,z] = f_s(z). \tag{8}$$

Therefore, for quasi-symmetric FSM channels besides properties [I1] to [I3], we have an additional property defined as follows:

I4 For a symmetric Markov channel, for any n, $P(z_n|x_n,s_n) = P(z_n|s_n) = T[s_n, z_n]$.

To make this statement clear, let us consider the FSM channel given above with $\mathcal{X} = \{1,2\}$, $\mathcal{Y} = \mathcal{Z} = \{1,2,3,4\}$

and $\mathcal{S}=\{1,2\}$. For this channel, we can define the function $z=\Phi(x,y)$ and $f_s(z)$ such that for each (x,y) pair, which has the same conditional probability within that state, $\Phi(x,y)$ returns the same value and the value of $f_s(z)$ at that value is the $p_{\mathcal{C}}(y_i|x_i,s_i)$ value, e.g., $\Phi(1,1)=\Phi(2,4)=1$ and $f_1(1)=a$ and $f_2(1)=a'$. Therefore, the channel conditional probabilities for each state can now be defined by Φ and the matrix T, where

$$T=\begin{pmatrix} a & b & c & d \\ a' & b' & c' & d' \end{pmatrix}.$$

Hence, the fundamental property for quasi- symmetric FSM channels is the existence of a noise process given by $Z_n=\Phi(X_n,Y_n)$ such that Z^n is independent of X^n [4]. The class of FSM channels having this property are termed variable noise channels [5].

Assumption 1: We finally assume that $\sum_x f_s(\Phi_s(x,y))$ is invariant with $s\in\mathcal{S}$.

This requirement will be needed in our dynamic programming approach which we use to determine the optimal feedback policy (as will be seen in the next section).

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, delayless and has large capacity. 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

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

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

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

Thus, when the transmitter wants to send a message, say $W\in\{1,2,...,2^{nR}\}$, it sends the codeword X^n, where $X_1=\psi_1(W)$ and $X_i=\psi_i(W,Y_1,\cdots,Y_{i-1})$, for $i=2,\cdots,n$. For a received Y^n at the channel output, the receiver uses the decoding function to estimate the transmitted message as $\hat{W}=\phi(Y^n)$. A decoding error is made when $\hat{W}\neq W$. We assume that the message W is uniformly distributed over $\{1,2,...,2^{nR}\}$. Therefore, the probability of error is given by

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

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. From Fano's inequality, we have

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

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

$$C_{FB}\leq\liminf_{n\to\infty}\sup\frac{1}{n}I(W;Y^n)$$

where the supremum is taken over all feedback encoding schemes $\{\psi_i\}_{i=1}^n$. Note that, $\sup_{\{\psi_i\}_{i=1}^n}\frac{1}{n}I(W;Y^n)=\sup\frac{1}{n}I(W;Y^n)$ where the right-hand side supremum is taken over feedback policies $\left\{P(x_i|x^{i-1},y^{i-1})\right\}_{i=1}^n$. We can write $I(W;Y^n)$ as follows

$$\begin{aligned} I(W;Y^n)= &\sum_{i=1}^n I(W;Y_i|Y^{i-1})\\ = &\sum_{i=1}^n\left(H(Y_i|Y^{i-1})-H(Y_i|W,Y^{i-1})\right). \end{aligned} \quad (9)$$

We next follow two steps in order to prove the contribution of the paper. In the first step, we show that the term $H(Y_i|W,Y^{i-1})$ is equal to $H(Z_i|Z^{i-1})$ and in the second step we will show that $\sum_{i=1}^n H(Y_i|Y^{i-1})$ is maximized by uniform feedback policies. We show the second step using a dynamic programming approach. Let us now start with the first step.

Lemma 2: A quasi-symmetric FSM satisfies

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

Proof of the lemma is given in Appendix A. As the next step, we show that all of the output conditional entropies $H(Y^i|Y^{i-1})$ in (9) are maximized by uniform i.i.d feedback policies. We solve this problem using a dynamic programming technique.

A. Entropy Optimization and Dynamic Programming

We now recast the optimization problem, maximization of the sum of conditional output entropies over all feedback policies, using a dynamic programming perspective [15]. Let us denote the feedback policies by

$$P(x_i|x^{i-1},y^{i-1})=\varphi_i,\ \text{ for }\ i=1,\ldots,n$$

and let $\pi=\{\varphi_i,1\leq i\leq n\}$. Let us recall our optimization problem:

$$\max_{\{\varphi_1,\cdots,\varphi_n\}}\{H(Y_n|Y^{n-1})+H(Y_{n-1}|Y^{n-2})+\cdots+H(Y_1)\}. \quad (10)$$

From (10), we observe that the optimization problem is nested. More explicitly, the policy at time n, i.e., φ_n, knows the previous policies and should maximize $H(Y_n|Y^{n-1})$ and on the other hand, φ_i knows previous policies but it should maximize $H(Y_i|Y^{i-1})+H(Y_{i+1}|Y^i)+\cdots+H(Y_n|Y^{n-1})$. Let us denote $V_i(Y^{i-1})=\max_{\varphi_i}[H(Y_i|Y^{i-1})+V_{i+1}(Y^i)]$ where $V_i(Y^{i-1})$ terms are explicitly given in (11) for $i=1,\cdots,n$.

$$
\begin{aligned}
V_n(Y^{n-1}) &= \max_{\varphi_n} H(Y_n|Y^{n-1}) \\
V_{n-1}(Y^{n-2}) &= \max_{\varphi_{n-1}} \left\{ H(Y_{n-1}|Y^{n-2}) + \max_{\varphi_n} \left\{ H(Y_n|Y^{n-1}) \right\} \right\} \\
&\vdots \\
V_1 &= \max_{\varphi_1} \left\{ H(Y_1) + \max_{\varphi_2} \left\{ H(Y_2|Y_1) + \cdots + \max_{\varphi_{n-1}} \left\{ H(Y_{n-1}|Y^{n-2}) + \max_{\varphi_n} \left\{ H(Y_n|Y^{n-1}) \right\} \right\} \cdots \right\} \right\} \quad (11)
\end{aligned}
$$

$V_{i+1}(Y^i)$ denotes the reward-to-go which is the future reward generated by the control policy at the current time. Therefore, the optimization problem turns out to be finding the policy, π, which achieves V_1. As the next step, we show that the policy achieving V_1 is composed of the uniform feedback policies for $i = 1, \cdots, n$. Along this way, we need the condition that the policies at times $(i-1), \cdots, 1$ should not affect the value attained by the conditional output entropy at time i when the policy at time i is uniform.

A sufficient condition to manage this problem is that $\sum_x f_s(\Phi_s(x, y))$ being invariant with $s \in \mathcal{S}$. This will be explicitly shown by a lemma in the succeeding sections.

Lemma 3: For a quasi-symmetric FSM channel, each conditional output entropy $H(Y_i|Y^{i-1}),\ i = 1, \cdots, n$ in (9) is maximized by a uniform i.i.d feedback policy:

$$
\begin{aligned}
\operatorname*{argmax}_{\varphi_i} H(Y_i|Y^{i-1}) &= \varphi^\star(x) = \frac{1}{|\mathcal{X}|},\ \forall x \in \mathcal{X} \\
& \quad \forall i = 1, \cdots, n. \qquad (12)
\end{aligned}
$$

A detailed proof of above lemma can be found in Appendix B. With this lemma, we have shown that for each i, $H(Y_i|Y^{i-1})$ is maximized by the uniform input policy. However, this is not sufficient to conclude that the optimal policy attaining V_1, i.e., the optimal policy maximizing $\sum_{i=1}^n H(Y_i|Y^{i-1})$, consists of uniform input policies. The reason for this is that Lemma 3 only maximizes the current conditional entropy with uniform input however, it is still possible that a non-uniform policy might result in a higher value function through the rewards-to-go.

Let us now look at the $P(Y_i|Y^{i-1})$ when we apply a uniform policy at time i (current time). With the uniform policy, we obtain

$$
\begin{aligned}
P(Y_i|Y^{i-1}) &= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i) P(x_i|x^{i-1}, y^{i-1}) \\
&\qquad P(s_i|s^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1}) \\
&\overset{(i)}{=} \frac{1}{|\mathcal{X}|} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i) P(s_i|s^{i-1}) P(x^{i-1}, s^{i-1}|y^{i-1}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s^i} P(y_i|x_i, s_i) P(s_i|s^{i-1}) \sum_{x^{i-1}} P(x^{i-1}, s^{i-1}|y^{i-1}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} \sum_{s^{i-1}} P(y_i|x_i, s_i) P(s_i|s^{i-1}) P(s^{i-1}|y^{i-1}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} P(y_i|x_i, s_i) P(s_i|y^{i-1})
\end{aligned}
$$

where (i) is valid since the input policy is uniform. Note that the dependency on past input policies comes through $P(s_i|y^{i-1})$ which includes transition probabilities between states, on which we have no control.

Lemma 4: The value of conditional entropy $H(Y_i|Y^{i-1})$, at time i, is independent of past feedback policies at times $(i-1), \cdots, 1$ iff $\sum_x f_s(\Phi_s(x, y))$ is invariant with $s \in \mathcal{S}$ and feedback policies are uniform i.i.d.

Proof:

$$
\begin{aligned}
P(Y_i|Y^{i-1}) &= \frac{1}{|\mathcal{X}|} \sum_{x_i} \sum_{s_i} P(y_i|x_i, s_i) P(s_i|y^{i-1}) \\
&= \frac{1}{|\mathcal{X}|} \sum_{s_i} P(s_i|y^{i-1}) \sum_{x_i} P(y_i|x_i, s_i) \\
&= \frac{1}{|\mathcal{X}|} \sum_{s_i} P(s_i|y^{i-1}) \underbrace{\sum_{x_i} f_s(\Phi(x_i, y_i))}
\end{aligned}
$$

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

Till now, we have shown that $H(Y_i|W, Y^{i-1}) = H(Z_i|Z^{i-1})$ and $\sum_{i=1}^n H(Y_i|Y^{i-1})$ is maximized by uniform input polices. With these results in hand, we have been able to show the following converse for the feedback capacity

$$
C_{FB} \le \liminf_{n \to \infty} \frac{1}{n} [H(\tilde{Y}^n) - H(Z^n)] \qquad (13)
$$

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

Lemma 5: For a quasi-symmetric FSM channel, $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$, with feedback, the noise process is a hidden Markov process with parameters $[\mathcal{S}, P, Z, T]$.

The proof can be found in [4, Lemma 1]. It should be noted that this proof is valid if the input process is i.i.d. Therefore, this proof seems to be inapplicable when feedback exists. However, for quasi-symmetric FSM channels, the capacity achieving distribution is uniform therefore the statement holds with feedback as well. Additionally, the output process, $\{\tilde{Y}_n\}_{n=1}^\infty$, for uniform X^n is also a hidden Markov process. Therefore, since the state process is stationary and ergodic both the output and noise processes are stationary and ergodic.

Theorem 1: The feedback capacity of the quasi-symmetric FSM channel $[\mathcal{X}, \mathcal{Y}, \mathcal{S}, P_S, Z, T, \Phi]$ satisfying the condition that $\sum_x f_s(\Phi_s(x, y))$ is invariant with $s \in \mathcal{S}$ is

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

where $\mathcal{H}(\tilde{Y})$ is the entropy rate of the output process for uniform i.i.d. X^n and $\mathcal{H}(Z)$ is the entropy rate of the hidden Markov process.

Proof: With (13) we already have a converse for the feedback capacity. We need to show that this bound is achievable. We first note that by Lemma 5, the noise and output processes are stationary which implies that

$$\begin{aligned} C_{FB} &\leq \liminf_{n\to\infty} \sup_{\{\varphi_1,\cdots,\varphi_n\}} \frac{1}{n} I(W;Y^n) \\ &= \liminf_{n\to\infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)] \\ &= \lim_{n\to\infty} \frac{1}{n}[H(\tilde{Y}^n) - H(Z^n)] = \mathcal{H}(\tilde{Y}) - \mathcal{H}(Z). \end{aligned} \tag{14}$$

It is sufficient to show that the bound in (14) is achievable. We now remark that there exists a coding policy which achieves this bound. Towards this end, we note that by adopting a policy which does not use feedback, it can be shown that $\mathcal{H}(\tilde{Y}) - \mathcal{H}(Z)$ is an admissible rate since the noise process is stationary and ergodic (e.g. see [8, Theorem 5.3]). 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 $\sum_x f_s(\Phi_s(x,y))$ being invariant with $s \in \mathcal{S}$.

The result follows by noting that a non-feedback code is a special case of a feedback code and that the non-feedback capacity is also achieved by uniform input policies.

IV. Examples of Quasi-Symmetric Finite State Markov Channels

In this section, we present some well-known channels that satisfy the quasi-symmetry condition presented in the paper.

Gilbert-Elliot Channel One of the widely used FSM channel 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" and "good" and the state transition matrix is given by:

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

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:

$$\begin{pmatrix} 1-p_G & p_G \\ p_G & 1-p_G \end{pmatrix}, \begin{pmatrix} 1-p_B & p_B \\ p_B & 1-p_B \end{pmatrix}.$$

From the above channel transition matrixes, it can be observed that the Gilbert-Elliot channel is a symmetric FSM channel. Namely, there exists a random variable $Z=\Phi(X,Y)$ and a function $f_s(z)$ such that, $f_0(0)=1-p_G$ and $f_0(1)=p_G$ and $f_1(0)=1-p_B$ and $f_0(1)=p_B$.

Therefore, the Gilbert-Elliot channel is a symmetric FSM channel with $\mathcal{Z}=\{0,1\}$, $\Phi(X,Y)=X\oplus Y$, where $\oplus$ represents modulo addition, and $T[s,z]$ defined above. By Corollary 1, feedback does not increase the capacity of the Gilbert-Elliot channel. It should be noted that this result is a special case of [2] and [6].

Discrete Modulo Additive Channel with Markovian Noise The second example that we consider is the discrete modulo additive channel with Markovian noise. Consider a 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 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, $\{Z_n\}_{n=1}^{\infty}$, is Markovian and it is independent from the input process. It is straightforward to see that the channel transition matrix for this channel is symmetric where each state is given by Z_{i-1}. For simplicity, let us assume that $q=3$. Then, the channel transition matrix at state i, C^i, will be as follows:

$$\begin{pmatrix} P(Z_i=0|z_{i-1}) & P(Z_i=1|z_{i-1}) & P(Z_i=2|z_{i-1}) \\ P(Z_i=2|z_{i-1}) & P(Z_i=0|z_{i-1}) & P(Z_i=1|z_{i-1}) \\ P(Z_i=1|z_{i-1}) & P(Z_i=2|z_{i-1}) & P(Z_i=0|z_{i-1}) \end{pmatrix}.$$

At different states, the channel transition matrix will still be symmetric with the same row permutation order. Therefore, the discrete modulo additive channel is a symmetric FSM channel with $\mathcal{A}=\{0,1,2\}$, $\Phi(X,Y)=X\oplus Y$ and state is Z_{i-1}. Therefore, by Corollary 1, feedback does not increase the capacity of the discrete modulo additive channel with Markovian noise.

This result is first shown in [2].

We should note that both examples satisfy the condition that the column sums of channel transition matrix is identical across all states as in fact they are all equal to one.

V. Two Specific Quasi-Symmetric FSM Channels

There are two classes of quasi-symmetric FSM channels that need further attention. In this section, we briefly mention how their channel properties involve the condition that the previous feedback policies either do not affect the value of the conditional output entropy at current time or the previous input policy should also be uniform to maximize the current conditional output entropy.

A. Non-Binary Noise Discrete Channel with Markovian Noise

The first example is a discrete binary-input 2^q-ary output communication channel (NBNDC), with memory introduced in [12]. The NBNDC model is described by the following equation

$$Y_k = (2^q-1)X_k + (-1)^{X_k} Z_k \tag{15}$$

for $k=1,2,\cdots$, where $X_k \in \mathcal{X}=\{0,1\}$ is the input, $Y_k, Z_k \in \mathcal{Z}=\mathcal{Y}=\{0,1,\cdots,2^q-1\}$ is the output and the noise processes respectively. The noise and input processes are independent from each other. The channel is a quasi-symmetric FSM channel with the following transition matrix, assuming that $q=2$, at state i

$$\begin{pmatrix} \epsilon_0 & \epsilon_1 & \epsilon_2 & \epsilon_3 \\ \epsilon_3 & \epsilon_2 & \epsilon_1 & \epsilon_0 \end{pmatrix}$$

where $\epsilon_j = P(Z_i = j|z_{i-1})$ and z_{i-1} is the state process. Note that, $\sum_x f_s(\Phi_s(x,y))$ varies with $s \in \mathcal{S}$. However, it can be shown that $\sum_{i=1}^n H(Y_i|Y^{i-1})$ is maximized if all the past policies are uniform as well.

B. Binary Erasure Channel with Markovian State

The second example is the binary erasure channel. In [10], this channel is considered as a finite buffer queue, which can be viewed as an FSM channel where the state of the finite buffer channel is determined by the state of the buffer. The channel has binary input and ternary output: $\mathcal{X} = \{0,1\}, \mathcal{Y} = \{0,1,E\}$. Let S_i denote the state of the channel when the packet i arrives such that $S_i = 1$: packet is erased and $S_i = 0$: packet gets through. Therefore, the channel transition matrix at state i will be as follows

$$C^i = \begin{pmatrix} \epsilon & 0 & 1-\epsilon \\ 0 & \epsilon & 1-\epsilon \end{pmatrix}.$$

where $P(S_i = 0|s^{i-1}) = \epsilon$. In [10], it was shown that feedback does not increase capacity of this erasure channel using a different technique. We however note that the approach presented in the paper gives the same result. This can be verified as follows: first note that the channel is quasi-symmetric for all states and therefore, $H(Y_i|Y^{i-1})$ is maximized by uniform input policies $\forall i = 1, \cdots, n$. What we further need to show is that $P(S_i|y^{i-1})$ is independent of past input policies (see Lemma 4). It should be noted that since

$$P(S_i|y^{i-1}) = \sum_{s^{i-1}} P(S_i|s^{i-1})P(s^{i-1}|y^{i-1})$$

and for a given output the state is deterministic therefore $P(S_i|y^{i-1})$ independent of past policies.

VI. 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 found a class of symmetric finite-state Markovian channels for which feedback does not increase capacity.

Appendix A
Proof of Lemma 2

Proof: Let us define $\eta(x) = x\log(x)$.

$$\sum_{i=1}^n H(Y_i|W, Y^{i-1}) = \sum_{i=1}^n H(Y_i|W, X_i, X^{i-1}, Y^{i-1})$$

since $X_i = \psi_i(W, Y_1, \cdots, Y_{i-1})$, for $i = 2, \cdots, n$. Hence,

$$\begin{aligned} &H(Y_i|W, X_i, X^{i-1}, Y^{i-1}) \\ &\quad = -E_{W,X^i,Y^{i-1}} \sum_{y_i} \eta(P(y_i|w, x_i, x^{i-1}, y^{i-1})). \end{aligned}$$

We also observe that

$$\begin{aligned} &\sum_{y_i} \eta(P(y_i|w, x_i, x^{i-1}, y^{i-1})) \\ &\overset{(a)}{=} \sum_{y_i} \eta(\sum_{s_i} p_C(y_i|x_i, s_i)P(s_i|w, x_i, x^{i-1}, y^{i-1})) \\ &\overset{(b)}{=} \sum_{y_i} \eta(\sum_{s_i} p_C(y_i|x_i, s_i)P(s_i|w, x^{i-1}, y^{i-1})) \\ &\overset{(c)}{=} \sum_{y_i} \eta(\sum_{s_i} p_C(y_i|x_i, s_i)P(s_i|x^{i-1}, y^{i-1}, z^{i-1})) \\ &\overset{(d)}{=} \sum_{z_i} \eta(\sum_{s_i} p(z_i|s_i)P(s_i|x^{i-1}, y^{i-1}, z^{i-1})) \end{aligned} \quad (16)$$

where (a) is valid since given s_i and x_i the distribution of y_i is independent of the rest, (b) is valid since x_i is a function of w and y^{i-1} and (c) is valid since

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

where $(\acute{a})$ and $(\acute{c})$ is valid since state is Markovian and $\Phi(x^{i-1}, y^{i-1}) = z^{i-1}$ and $(\acute{b})$ can be shown inductively as follows:

$$\begin{aligned} P(s_1|w, x_1, y_1) &= \frac{P(s_1, w, x_1, y_1)}{\sum_{s_1} P(s_1, w, x_1, y_1)} \\ &= \frac{P(y_1|x_1, s_1)P(x_1, s_1, w)}{\sum_{s_1} P(y_1|x_1, s_1)P(x_1, s_1, w)} \\ &\overset{(i)}{=} \frac{P(y_1|x_1, s_1)P(s_1)P(x_1, w)}{\sum_{s_1} P(y_1|x_1, s_1)P(s_1)P(x_1, w)} \\ &= \frac{P(y_1|x_1, s_1)P(s_1)}{\sum_{s_1} P(y_1|x_1, s_1)P(s_1)} = P(s_1|x_1, y_1) \end{aligned} \quad (17)$$

where (i) is valid since s_1 is independent of the rest. Similarly,

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

where (ii) is valid since x_2 is a function of x_1, y_1 and w and (iii) is due to (17). Using these steps recursively verifies $(\acute{b})$. Finally, (d) follows from (7), (8) and I4. We next show that $P(s_i|x^{i-1}, y^{i-1}) = P(s_i|z^{i-1})$. Note that, $P(s_i|x^{i-1}, y^{i-1}) = P(s_i|x^{i-1}, y^{i-1}, z^{i-1}) = \sum_{s^{i-1}} P(s_i|s^{i-1})P(s^{i-1}|x^{i-1}, y^{i-1}, z^{i-1})$. Then,

$$\begin{aligned}
P(s_1|x_1, y_1) &= \frac{P(s_1, x_1, y_1)}{\sum_{s_1} P(s_1, x_1, y_1)} \\
&= \frac{P(y_1|x_1, s_1)p(x_1)p(s_1)}{\sum_{s_1} P(y_1|x_1, s_1)p(x_1)p(s_1)} \stackrel{(e)}{=} \frac{f_{s_1}(z_1)p(x_1)p(s_1)}{\sum_{s_1} f_{s_1}(z_1)p(x_1)p(s_1)} \\
&\stackrel{(f)}{=} \frac{p(z_1|x_1, s_1)p(x_1)p(s_1)}{\sum_{s_1} p(z_1|x_1, s_1)p(x_1)p(s_1)} = p(s_1|z_1) \qquad (18)
\end{aligned}$$

where (e) is due to (7), (8) and (f) is due to I4. Similarly,

$$\begin{aligned}
P(s_2|x_1, y_1) &= \sum_{s_1} P(s_2|s_1)P(s_1|x_1, y_1) \\
&\stackrel{(g)}{=} \sum_{s_1} P(s_2|s_1)P(s_1|z_1) = P(s_2|z_1) \qquad (19)
\end{aligned}$$

where (g) is due to (18). Now, let us consider $P(s_3|x^2, y^2) = \sum_{s_2} P(s_3|s_2)P(s_2|x^2, y^2)$. Then,

$$\begin{aligned}
&P(s_2|x^2, y^2) = P(s_2|x^2, z^2) \\
&= \frac{P(s_2, x_1, x_2, z_1, z_2)}{\sum_{s_2} P(s_2, x_1, x_2, z_1, z_2)} \\
&\stackrel{(h)}{=} \frac{P(z_2|s_2)P(x_2|x_1, z_1)P(s_2|x_1, z_1)P(x_1)P(z_1)}{\sum_{s_2} P(z_2|s_2)P(x_2|x_1, z_1)P(s_2|x_1, z_1)P(x_1)P(z_1)} \\
&\stackrel{(k)}{=} \frac{P(z_2|s_2)P(x_2|x_1, z_1)P(s_2|z_1)}{\sum_{s_2} P(z_2|s_2)P(x_2|x_1, z_1)P(s_2|z_1)} \\
&\stackrel{(l)}{=} \frac{P(z_2|s_2)P(x_2|x_1, z_1)P(s_2|z_1)}{P(x_2|x_1, z_1)\sum_{s_2} P(z_2|s_2)P(s_2|z_1)} \\
&= \frac{P(z_2|s_2)P(s_2|z_1)}{\sum_{s_2} P(z_2|s_2)P(s_2|z_1)} = P(s_2|z^2) \qquad (20)
\end{aligned}$$

where (h) is valid by (7), (k) is valid by (19) and (l) is valid since $P(x_2|x_1, z_1)$ is independent of s_2. Using (18), (19) and (20) recursively for $i = 2, \ldots, n$ we obtain

$$P(s_i|x^{i-1}, y^{i-1}, z^{i-1}) = P(s_i|x^{i-1}, y^{i-1}) = P(s_i|z^{i-1}). \quad (21)$$

Substituting (21) into (16) gives us $H(Z_i|Z^{i-1})$. ■

Appendix B
Proof of Lemma 3

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

$$H(Y_i|Y^{i-1}) = \sum_{y^{i-1}} P(y^{i-1})H(Y_i|Y^{i-1} = y^{i-1}) \qquad (22)$$

where

$$H(Y_i|Y^{i-1} = y^{i-1}) = -\sum_{y_i} P(y_i|y^{i-1}) \log P(y_i|y^{i-1}). \quad (23)$$

To show that $H(Y_i|Y^{i-1})$ in (22) is maximized by a uniform input policy, it is enough to show that such a uniform policy 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

$$\begin{aligned}
&\sum_{x_i}\sum_{x^{i-1}}\sum_{s_i}\sum_{s^{i-1}} P(y_i, x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1}) \\
&= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i, x^{i-1}, s^{i-1}, y^{i-1}) \\
&\qquad P(x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1}) \\
&\stackrel{(i)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i)P(x_i, s_i, x^{i-1}, s^{i-1}|y^{i-1}) \\
&= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i)P(x_i, x^{i-1}, s^{i-1}|y^{i-1}) \\
&\qquad P(s_i|x_i, x^{i-1}, s^{i-1}, y^{i-1}) \\
&\stackrel{(ii)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i)P(s_i|s^{i-1}) \\
&\qquad P(x_i|x^{i-1}, s^{i-1}, y^{i-1})P(x^{i-1}, s^{i-1}|y^{i-1}) \\
&\stackrel{(iii)}{=} \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} P(y_i|x_i, s_i)P(x_i|x^{i-1}, y^{i-1}) \\
&\qquad P(s_i|s^{i-1})P(x^{i-1}, s^{i-1}|y^{i-1}).
\end{aligned}$$

where (i) follows by (4), (ii) is valid due to the property I2 and finally (iii) is due to the fact that the feedback input depends only on (x^{i-1}, y^{i-1}). Thus

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

The key observation in equation (24) 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 (6),

$$\begin{aligned}
P(y_i|y^{i-1}) &= \sum_{x_i, x^{i-1}} \sum_{s_i, s^{i-1}} f_{s_i}(\Phi_{s_i}(x_i, y_i))P(s_i|s^{i-1}) \\
&\qquad P(x_i|x^{i-1}, y^{i-1})P(x^{i-1}, s^{i-1}|y^{i-1}). \qquad (25)
\end{aligned}$$

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

$$\begin{aligned}
P(y_{j_k}|y^{i-1}) = &\sum_{x_i, x^{i-1}, s_i, s^{i-1}} f_{s_i}(\Phi_{s_i}(x_i, y_{j_k}))P(s_i|s^{i-1}) \\
&\qquad P(x_i|x^{i-1}, y^{i-1})P(x^{i-1}, s^{i-1}|y^{i-1}). \quad (26)
\end{aligned}$$

We desire to maximize (22) over the feedback policies $P(X_i|X^{i-1}, Y^{i-1})$. Setting $\mathcal{X} = \{x_1, x_2, \ldots, x_k\}$ and denoting the feedback policies by

$$P(X_i = x_l|x^{i-1}, y^{i-1}) = \varphi_i(x_l), \text{ for } l = 1, \ldots, k, \quad (27)$$

we can write

$$P(y_{j_1}|y^{i-1}) = \sum_{s_i,s^{i-1},x^{i-1}} P(s_i|s^{i-1})P(x^{i-1},s^{i-1}|y^{i-1})$$
$$\{\varphi_i(x_1)f_{s_i}(\Phi_{s_i}(x_1,y_{j_1})) + \cdots + \varphi_i(x_k)f_{s_i}(\Phi_{s_i}(x_k,y_{j_1}))\},$$

$$P(y_{j_2}|y^{i-1}) = \sum_{s_i,s^{i-1},x^{i-1}} P(s_i|s^{i-1})P(x^{i-1},s^{i-1}|y^{i-1})$$
$$\{\varphi_i(x_1)f_{s_i}(\Phi_{s_i}(x_1,y_{j_2})) + \cdots + \varphi_i(x_k)f_{s_i}(\Phi_{s_i}(x_k,y_{j_2}))\},$$

and,

$$P(y_{j_{|\mathcal{Y}_j|}}|y^{i-1}) = \sum_{s_i,s^{i-1},x^{i-1}} P(s_i|s^{i-1})P(x^{i-1},s^{i-1}|y^{i-1})$$
$$\left\{\varphi_i(x_1)f_{s_i}(\Phi_{s_i}(x_1,y_{j_{|\mathcal{Y}_j|}})) + \cdots + \varphi_i(x_k)f_{s_i}(\Phi_{s_i}(x_k,y_{j_{|\mathcal{Y}_j|}}))\right\}.$$

It should be noted that, each $f_{s_i}(\Phi_{s_i}(x_i,y_{j_k}))$ corresponds to an entry in the channel transition matrix Q^{s_i} at state s_i. We also know that, the rows of the sub-array $Q_j^{s_i}$ are permutations of each other. In other words, each $f_{s_i}(\Phi_{s_i}(x_i,y_{j_k}))$ value appears exactly $|\mathcal{X}|$ times in the sub-array $Q_j^{s_i}$. Moreover, the feedback policy $\varphi_i(x_l)$ is multiplied by a different $f_{s_i}(\Phi_{s_i}(x_i,y_{j_k}))$ value in each of the equations of $P(y_{j_k}|y^{i-1})$ given above. Therefore, $\sum_{k=1}^{|\mathcal{Y}_j|} P(Y_i = y_{j_k}|y^{i-1})$ is equal to

$$\sum_{s_i,s^{i-1},x^{i-1}} P(s_i|s^{i-1})P(x^{i-1},s^{i-1}|y^{i-1}) \sum_{k=1}^{|\mathcal{Y}_j|} f_{s_i}(\Phi_{s_i}(x_i,y_{j_k})) \tag{28}$$

where $\sum_{k=1}^{|\mathcal{Y}_j|} f_{s_i}(\Phi_{s_i}(x_i,y_{j_k}))$ is the sum of any row in the sub-array $Q_j^{s_i}$ and is equal to constant (by the weak-symmetry of sub-array $Q_j^{s_i}$). Note that, (28) is independent of the feedback policies. Similarly for all other sub-arrays, assuming that there are m sub-arrays, their conditional output sums will be independent of the feedback policies. Let us denote these sums by $\Omega_1,\ldots,\Omega_m$. More specifically for sub-array i, let $\Omega_i = \sum_{k=1}^{|\mathcal{Y}_i|} P(Y_k = y_{i_k}|y^{i-1})$. Then the optimization problem in (23) now becomes,

$$\operatorname*{argmax}_{\Omega_{i,j}} - \sum_{i=1}^{m}\sum_{j=1}^{|\mathcal{Y}_i|} \Omega_{i,j}\log\Omega_{i,j} \tag{29}$$

where $\sum_{i=1}^{m}\sum_{j=1}^{|\mathcal{Y}_i|}\Omega_{i,j} = 1$ and $\Omega_{i,j},\ j = 1,\ldots,|\mathcal{Y}_i|$ denotes conditional output probabilities in sub-array i. For each sub-array i, we need to find the $\Omega_{i,j}$ values that maximize $\sum_{j=1}^{|\mathcal{Y}_i|}\Omega_{i,j}\log\Omega_{i,j}$. By the log-sum inequality, we have that

$$-\sum_{j=1}^{|\mathcal{Y}_i|}\Omega_{i,j}\log\Omega_{i,j} \le -\sum_{j=1}^{|\mathcal{Y}_i|}\Omega_{i,j}\log\frac{\sum_{j=1}^{|\mathcal{Y}_i|}\Omega_{i,j}}{|\mathcal{Y}_i|} \tag{30}$$

with equality iff

$$\Omega_{i,s} \;=\; \Omega_{i,t}\ \ \forall s,t \in \{1,\ldots,|\mathcal{Y}_i|\}. \tag{31}$$

In other words, for the sub-array i, the conditional entropy is maximized iff 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 feedback policies.

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 i. Then $P(Y_i = y_{j_s}|y^{i-1}) = P(Y_i = y_{j_t}|y^{i-1})$ which implies that

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

However, for a fixed output $\sum_{l=1}^{k} f_{s_i}(\Phi(x_l,y_{j_s}))$ returns the sum of the column corresponding to output y_{j_s} (similarly for y_{j_t}) and since sub-array i is weakly symmetric the column sums are equal. 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 policy will also satisfy the equivalency of conditional output probabilities. ∎

REFERENCES

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

[2] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inform. Theory,*, Mar. 1995.

[3] F. Alajaji and T. Fuja, "Effect of feedback on the capacity of discrete additive channels with memory," *Proc. IEEE Int. Symp. Information Theory* Trondheim, Norway, June 1994.

[4] M. Rezaeian, "Symmetric characterization of finite state markov channels," *Proc. IEEE Int. Symp. Information Theory* Seattle, USA, July 2006.

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

[6] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *submitted to IEEE Trans. Inform. Theory,* Nov. 2007.

[7] 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.

[8] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inform. Theory,* vol. 55 No.1, pp. 329-343 , Jan. 2009.

[9] H. Permuter, T. Weissman, and A.J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inform. Theory,* vol. 55, pp. 644-662 , Feb. 2009.

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

[11] N. Şen, F. Alajaji and S. Yüksel, "On the feedback capacity of a discrete non-binary noise channel with memory," in *Proc. 11th Canadian Workshop Inform. Theory,* Ottawa, Canada, May 2009.

[12] 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, June 2009.

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

[14] F. Alajaji, "Lecture notes in information theory," *Queen's University,* Fall. 2008.

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