A sequential transmission scheme for unifilar finite-state channels with feedback based on posterior matching

Achilleas Anastasopoulos
Electrical Engineering and Computer Science Department
University of Michigan
Ann Arbor, MI, 48109-2122
email: anastas@umich.edu

***Abstract*—The capacity of unifilar finite-state channels with feedback has been recently derived in the form of a single-letter expression and it has been evaluated analytically for a number of channels of interest, such as the trapdoor channel and the Ising channel with feedback. In this paper, we investigate transmission schemes for this class of channels. These schemes are inspired by the posterior matching scheme (PMS) introduced for memoryless channels with feedback. The transmission scheme is proven to achieve zero rate and is conjectured to achieve channel capacity.**

I. Introduction

Communication in the presence of feedback has been a long studied problem which dates back to Shannon's early work [1], showing that feedback cannot increase the capacity of memoryless channels. For channels with memory, however, feedback may increase the capacity.

One of the first results on the capacity of channels with memory and feedback was by Viswanathan [2], who studied the capacity of a finite-state channel (FSC) with receiver channel-state information (CSI) and delayed feedback in the absence of inter-symbol interference (ISI). Later, Chen and Berger [3] derived the capacity of a FSC with ISI where current (i.e., not delayed) CSI is available at the transmitter and the receiver. Yang et. al. [4] used a stochastic control methodology to find the capacity of the ISI channel (where the evolution of the state is deterministic). In a related line of work, the authors of [5], [6] derived a single-letter expression for the class of unifilar channels and used dynamic programming to evaluate analytically the channel capacity for the trapdoor and the Ising channels with feedback. Recently, Tatikonda and Mitter [7] provided a general stochastic control framework for evaluating the capacity of the FSC with feedback. In that work, the capacity was characterized as the solution of a dynamic programming average reward optimality equation (AROE). Como et. al. [8] used an approach similar to that in [7] to find the capacity of the FSC when current CSI is available at the transmitter and the receiver. An upper bound on the capacity of the FSC without ISI and CSI was found using dynamic programming by Huang et. al. [9].

A significant amount of research has also been done in the area of code design for channels with memory. Horstein [10] proposed a simple sequential transmission scheme for the binary symmetric channel (BSC) which is capacity-achieving and provides a larger error exponent than traditional fixed-length block-coding. Similarly, Schalkwijk and Kailath [11] showed that capacity and a double exponentially decreasing error probability can be achieved by a simple sequential transmission scheme for the additive white Gaussian noise channel (AWGNC) with average power constraint. Recently, Shayevitz and Feder [12]–[14] identified an underlying principle shared by the aforementioned Horstein and Schalkwijk-Kailath schemes and introduced a simple encoding scheme, namely the posterior matching scheme (PMS) for general memoryless channels. Furthermore, they showed that the PMS achieves the capacity of general discrete memoryless channels (DMCs). Subsequently, Coleman [15] revisited the PMS and reformulated the problem in a stochastic control framework. In [16], [17] the authors proposed a PMS scheme for channels with memory and state and output feedback.

In this paper, we investigate potentially capacity-achieving transmission schemes for a special class of finite-state channels (FSC), namely unifilar FSCs, where the channel state at time $t+1$ is a deterministic function of the current state input and output. Based on the derived capacity expression, and the structure of the capacity-achieving input distribution derived in [5], we propose a PMS-like sequential transmission scheme. Intuitively, the transmitter and receiver both update a "state" which includes the a-posteriori distribution of the transmitted message conditioned on the observation output for all current states. At the same time, the transmitter is sending the most informative input symbol in such a way that the state-conditioned input distribution is the one implied by the capacity expression. Due to the presence of ISI, the proposed scheme has a number of unique characteristics that were not present in the PMS proposed for memoryless channels. Regarding the methodology followed for proving zero rate achievability, we generalize the tools utilized in [14], [16], [17].

The remainder of the paper is organized as follows. In Section II, the channel model and some preliminaries are introduced. The generalized PMS is presented in Section III and some important properties are developed in IV. In Section V, the proof of zero-rate achievability is presented. Section VI concludes the paper.

II. Channel model and preliminaries

We consider channels with input, output and state random processes denoted by $(X_t)_{t=1}^{\infty}$, $(Y_t)_{t=1}^{\infty}$, $(S_t)_{t=1}^{\infty}$, respectively. The corresponding input, output and state alphabets are finite and denoted by $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}$, respectively. For reasons that will become clear in the following we also assume that $\mathcal{X} = \{0, 1, \ldots, |\mathcal{X}| - 1\}$. At time t the receiver has access to the current channel output Y_t, which is also fed back to the transmitter with unit delay. The state transition and the channel output are defined as

$$P(Y_t|S^t, X^t, Y^{t-1}) = Q(Y_t|X_t, S_t) \tag{1a}$$

$$S_{t+1} = g(S_t, X_t, Y_t) \tag{1b}$$

for a given stochastic kernel $Q \in \mathcal{X} \times \mathcal{S} \to \mathcal{P}(\mathcal{Y})$ and a deterministic function $g \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \to \mathcal{S}$, where we denote by $\mathcal{P}(A)$ the set of distributions on A. These channels are referred to as unifilar finite-state channels (FSCs) [5], [6].

In [5] the authors formulated the problem of finding the capacity of unifilar FSCs with unit-delayed output feedback as a stochastic control problem. The basic idea was to define a controlled Markov process with state $B_t \in \mathcal{P}(\mathcal{S})$ defined as

$$B_t(s) \stackrel{\text{def}}{=} P(S_{t+1} = s|Y^t) \qquad \forall s \in \mathcal{S}. \tag{2}$$

Their results (although not presented in this form) imply that under mild conditions the capacity of unifilar FSCs can be put in a single-letter expression of the form

$$C = \sup_{P(X_t|S_t, B_{t-1})} I(X_t, S_t \wedge Y_t|B_{t-1}) \tag{3}$$

where the mutual information in (3) is evaluated with the joint distribution

$$P(X_t, Y_t, S_t, B_{t-1}) = \\ Q(Y_t|X_t, S_t)P(X_t|S_t, B_{t-1})B_{t-1}(S_t)P(B_{t-1}), \tag{4}$$

where $P(B_{t-1})$ is the steady-state distribution of the induced Markov process $(B_t)_t$ defined by the conditional distribution

$$P(B_t|B_{t-1}) = \sum_{y_t} \delta_{\theta(B_{t-1}, y_t)}(B_t) \\ \sum_{x_t, s_t} Q(y_t|x_t, s_t)P(x_t|s_t, B_{t-1})B_{t-1}(s_t), \tag{5}$$

and θ is given by the recursive updating of $B_t = \theta(B_{t-1}, Y_t)$ as shown in (6). We note that the above expressions can also be derived from the general framework developed in [7] as a special case. It was further shown in [5] that the capacity expression can, be evaluated as the solution of an appropriate average reward optimality equation (AROE) [18]. For the special case of the trapdoor and the Ising channels with feedback the authors provided analytical solutions to the AROE in [5], [6].

III. The posterior matching transmission scheme

In this section we describe a sequential transmission scheme for the unifilar FSC with unit-delay output feedback. We assume that the capacity achieving distributions $(P(X_t|S_t, B_{t-1}))_{S_t, B_{t-1}}$ have been found for all values of $(S_t, B_{t-1}) \in \mathcal{S} \times \mathcal{P}(\mathcal{S})$. We consider transmission of a message $W \in [0, 1)$. Let $\mathcal{F}$ be the set of all valid cdfs over $[0, 1)$ and define the random variable $F'_t \in \mathcal{S} \to \mathcal{F}$ as

$$F'_t(w|s) \stackrel{\text{def}}{=} P(W \leq w|S_{t+1} = s, Y^t), \; w \in [0, 1), s \in \mathcal{S}. \tag{7}$$

With the above definition, $F'_t(\cdot|s)$ is the a-posteriori cdf of W conditioned on $S_{t+1} = s, Y^t$, and F'_t is the collection of all these conditional cdfs for all $s \in \mathcal{S}$. In the following we will also use the notation $F'_t(w) \stackrel{\text{def}}{=} (F'_t(w|s))_{s \in \mathcal{S}}$, as well as the notation $F'_t(s) \stackrel{\text{def}}{=} F'_t(\cdot|s)$; the distinction should be obvious from the context.

The channel input X_t for $t = 1, 2, 3, \ldots$ is generated as

$$X_t = F^{-1}_{P(\cdot|S_t, B_{t-1})}(F'_{t-1}(W|S_t)) \tag{8a}$$

$$\stackrel{\text{def}}{=} e(F'_{t-1}(W), S_t, B_{t-1}). \tag{8b}$$

where the inverse cdf is defined as $F_P^{-1}(y) \stackrel{\text{def}}{=} \inf\{x : F_P(x) \geq y\}$ and F_P denotes the cdf corresponding to a random variable with distribution P. The initial posterior cdfs are set to be uniform, i.e., $F'_0(w|s) = w$ for $w \in [0, 1)$. This transmission scheme is similar to the PMS scheme proposed for DMCs in [14] with the following differences. First, the input signal X_t is matched to the capacity achieving distribution $P(X_t|S_t, B_{t-1})$ which is not constant, but changes depending on the current values of the quantities S_t, B_{t-1}. Second, the encoding scheme is a function of the message W through the value $F'_t(W|S_t)$ and thus an appropriate "state" for the transmitter involves the entire vector $F'_t(W) = (F'_t(W|s))_{s \in \mathcal{S}}$. Third, the additional "state" variable B_t can be recursively updated through the update equation $B_t = \theta(B_{t-1}, Y_t)$ as shown in (6), and the current state S_t is available at the transmitter at time t due to the unifilar property, i.e., $S_t = g(S_{t-1}, X_{t-1}, Y_{t-1}) = \tilde{g}(X^{t-1}, Y^{t-1})$[1]. The proposed transmission scheme is also similar to the one in [16] and [17, ch. 5] for ISI FSCs with the difference being that in the above references the channel is not necessarily unifilar, and CSI is available at the receiver and with unit delay at the transmitter. We now show that the quantity F'_t can also be evaluated recursively at the transmitter. Indeed, the corresponding pdf can be updated as shown in (9) with $P(Y_t|B_{t-1})$ derived from the joint distribution in (4). Integrating out (9) results in an explicit recursion for the cdfs F'_t. We will use the shorthand notation

$$F'_t = \phi(F'_{t-1}, B_{t-1}, Y_t) \tag{10}$$

for the above update. It can be shown from (9) that $F'_t(w)$ is a function of F'_{t-1} only through $F'_{t-1}(w)$. This has important implications for the analysis of the PMS scheme.

[1] Throughout the paper we suppress the dependency on the initial state S_1 which is assumed known to transmitter and receiver.

$$B_t(s_{t+1}) = \frac{\sum_{x_t,s_t} \delta_{g(s_t,x_t,Y_t)}(s_{t+1}) Q(Y_t|x_t,s_t) P(x_t|s_t,B_{t-1}) B_{t-1}(s_t)}{\sum_{x_t,s_t,s_{t+1}} \delta_{g(s_t,x_t,Y_t)}(s_{t+1}) Q(Y_t|x_t,s_t) P(x_t|s_t,B_{t-1}) B_{t-1}(s_t)} \tag{6}$$

$$dF'_t(w|s) = \frac{\sum_{s_t} \delta_{g(s_t,e(F'_{t-1}(w),s_t,B_{t-1}),Y_t)}(s) Q(Y_t|e(F'_{t-1}(w),s_t,B_{t-1}),s_t) B_{t-1}(s_t) dF'_{t-1}(w|s_t)}{B_t(s) P(Y_t|B_{t-1})}, \tag{9}$$

In the following we will also use the notation $F'_t(w) = \phi(F'_{t-1}, B_{t-1}, Y_t)(w) = \phi(F'_{t-1}(w), B_{t-1}, Y_t)$.

At the receiver, maximum likelihood (ML) estimation is performed based on the posterior distribution of the message W conditioned on the observations Y^t. Let the corresponding posterior cdf be $F_t \in \mathcal{F}$ with $F_t(w) = P(W \leq w|Y^t)$. It is straightforward to show that

$$F_t(w) = \sum_{s_{t+1}} F'_t(w|s_{t+1}) B_t(s_{t+1}), \tag{11}$$

which together with ϕ and θ results in $F_t = \psi(F'_{t-1}, B_{t-1}, Y_t)$ and moreover in $F_t(w) = \psi(F'_{t-1}(w), B_{t-1}, Y_t)$. In view of the above recursions, as well as the recursions (6) and (9) for B_t and F'_t, respectively, it is clear that the receiver can also evaluate the posterior cdf F_t recursively. Finally, the message estimate is obtained as

$$\hat{W}_t = d(F_t, 2^{-Rt}/2), \tag{12}$$

where the message estimate function $d(F, \epsilon)$ is defined as

$$d(F, \epsilon) \stackrel{\text{def}}{=} \arg\max_w \{F(w+\epsilon) - F(w-\epsilon)\}, \tag{13}$$

and R is the transmission rate.

IV. Properties of the transmission scheme

The following properties are now established for the described transmission scheme.

Lemma 1. *For the transmission scheme described above and for every $t \geq 1$ we have*

1) *The random variables $F'_t(W|S_{t+1})$ and $F_t(W)$ are uniformly distributed in $[0, 1)$.*
2) $$P(X_t|S_t, Y^{t-1}) = P(X_t|S_t, B_{t-1}) \tag{14}$$
3) $$P(Y_t|Y^{t-1}) = P(Y_t|B_{t-1}). \tag{15}$$

Proof: First note that repeated application of the recursions g, θ, and ϕ imply that $S_{t+1} = \tilde{g}(X^t, Y^t) = \hat{g}(W, Y^t)$, $B_t = \hat{\theta}(Y^t)$ and $F'_t = \hat{\phi}(Y^t)$.

For the first property we have for any $a \in [0, 1)$

$$\begin{aligned}
&P(F'_t(W|S_{t+1}) \leq a) && \text{(16a)}\\
&= \sum_{s_{t+1},y^t} P(F'_t(W|S_{t+1}) \leq a|s_{t+1}, y^t) P(s_{t+1}, y^t) && \text{(16b)}\\
&= \sum_{s_{t+1},y^t} P(f'_t(W|s_{t+1}) \leq a|s_{t+1}, y^t) P(s_{t+1}, y^t) && \text{(16c)}\\
&= \sum_{s_{t+1},y^t} P(W \leq f'^{-1}_t(a|s_{t+1})|s_{t+1}, y^t) P(s_{t+1}, y^t) && \text{(16d)}\\
&= \sum_{s_{t+1},y^t} f'_t(f'^{-1}_t(a|s_{t+1})|s_{t+1}) P(s_{t+1}, y^t) && \text{(16e)}\\
&= \sum_{s_{t+1},y^t} a P(s_{t+1}, y^t) && \text{(16f)}\\
&= a, && \text{(16g)}
\end{aligned}$$

where f'_t is the cdf corresponding to the realization y^t. Similarly for $F_t(W)$.

For the second property we can write

$$\begin{aligned}
&P(X_t|S_t, Y^{t-1})\\
&= P(X_t|S_t, Y^{t-1}, B_{t-1}, F'_{t-1}) && \text{(17a)}\\
&= \int_w P(X_t|S_t, Y^{t-1}, B_{t-1}, F'_{t-1}, W = w)\\
&\qquad P(dw|S_t, Y^{t-1}) && \text{(17b)}\\
&= \int_w \delta_{e(F'_{t-1}(w),S_t,B_{t-1})}(X_t) dF'_{t-1}(w|S_t) && \text{(17c)}\\
&= \sum_{x\in\mathcal{X}} \delta_x(X_t) \int_{w:e(F'_{t-1}(w),S_t,B_{t-1})=x} dF'_{t-1}(w|S_t) && \text{(17d)}\\
&= \sum_{x\in\mathcal{X}} \delta_x(X_t) P(x|S_t, B_{t-1}) && \text{(17e)}\\
&= P(X_t|S_t, B_{t-1}), && \text{(17f)}
\end{aligned}$$

where (17a) is due to the fact that B_{t-1} and F'_{t-1} can be recovered from Y^{t-1} using $\hat{\theta}$ and $\hat{\phi}$; (17c) is due to the transmission scheme in (8) and the definition of F'_t in (7); and (17e) is due to the posterior matching in (8). Similarly for the third property. ■

We now show that the proposed transmission scheme satisfies a necessary condition for capacity achievability. For transmission at rate R over a period of n channel uses we have

$$nR = H(W) = H(W|Y^n) + I(W \wedge Y^n), \tag{18}$$

and for the proposed scheme, the second term becomes

$$
\begin{aligned}
&I(W \wedge Y^n) \\
&= \sum_{t=1}^{n} I(W \wedge Y_t | Y^{t-1}) && (19a)\\
&= \sum_{t=1}^{n} H(Y_t|Y^{t-1}) - H(Y_t|W, Y^{t-1}) && (19b)\\
&= \sum_{t=1}^{n} H(Y_t|B_{t-1}) - H(Y_t|W, Y^{t-1}, X^t, S^t, B_{t-1}) && (19c)\\
&= \sum_{t=1}^{n} H(Y_t|B_{t-1}) - H(Y_t|X_t, S_t, B_{t-1}) && (19d)\\
&= \sum_{t=1}^{n} I(X_t, S_t \wedge Y_t|B_{t-1}) && (19e)\\
&= nC, && (19f)
\end{aligned}
$$

where (19c) is due to (15) and the fact that X^t is implicitly only a function of W and Y^{t-1}, S^t can be derived from X^{t-1} and Y^{t-1} through $\hat{g}$ and B_{t-1} is a function of Y^{t-1} through $\hat{\theta}$; and (19d) is due to the channel statistics and the fact that B_{t-1} is independent of Y_t conditioned on X_t and S_t.

V. Achievability

Let $\hat{W}_t$ be the message point estimate at the receiver at time t. Then, a transmission scheme achieves rate R if

$$\lim_{t\to\infty} P(|W - \hat{W}_t| > 2^{-tR}) = 0. \quad (20)$$

In particular, we say that a transmission schemes achieves zero rate if

$$\forall \epsilon > 0 \qquad \lim_{t\to\infty} P(|W - \hat{W}_t| > \epsilon) = 0. \quad (21)$$

The subsequent analysis is valid for a (broad) class of non-pathological channels, namely fixed-point free channels, that are defined in a way similar to [14], [17].

Definition 1. *A channel is called fixed-point free if for any* $a = (a(s))_{s\in\mathcal{S}} \in (0,1)^{|\mathcal{S}|}$, $b \in \mathcal{P}(\mathcal{S})$

$$P(\sum_s \phi(a, b, Y_t)(s)\theta(b, Y_t)(s) = \sum_s a(s)b(s)) < 1. \quad (22)$$

The above definition essentially implies that the recursions θ, ϕ do not have a fixed point with respect to the posterior cdf F_t that they imply. Finally, assuming that the probabilities $Q(y|x,s)$ and the capacity achieving distribution $P(x|s,b)$ are non-zero for all x, y, s, b, the recursions in (9) and (6) guarantee that for every realization of the random variables of interest, F'_t will always have a pdf; in addition the pdf will be non-zero everywhere in $(0,1]^2$.

[2] Channels that do not satisfy this assumption can be treated as limiting cases of this analysis.

A. Achieving $R = 0$

For a cdf $F \in \mathcal{F}$ define a Lyapunov function V_λ as follows.

$$V_\lambda(F) = \int_0^1 \lambda(F(w))dw, \quad (23)$$

where $\lambda : [0,1] \to [0,1]$ is onto, strictly concave and symmetric about 0.5. This definition implies that $\lambda(x)$ equals 0 at $x = 0,\ 1$ and equals 1 at $x = 1/2$. Furthermore, $V_\lambda(F)$ is small if F resembles a step function (it is exactly 0 for a step function). A function $\xi : [0,1] \to [0,1]$ is called a contraction if it is nonnegative, concave, and $\xi(x) < x$ for $x \in (0,1)$. The following lemma shows that the probability of having an F_t that does not resemble a step function is zero at the limit of large t.

Lemma 2. *If the channel is fixed-point free, then for* $\epsilon > 0$ *and for all* $f' \in \mathcal{F}^{|\mathcal{S}|}$,

$$\lim_{t\to\infty} P(V_\lambda(F_t) > \epsilon | F'_0 = f') = 0. \quad (24)$$

Proof: We would like to find a contraction mapping ξ such that for every w, b_{t-1}, f'_{t-1} we have

$$E[\lambda(F_t(w))|b_{t-1}, f'_{t-1}] \le \xi(\lambda(f_{t-1}(w))), \quad (25)$$

where $f_{t-1} = \sum_s f'_{t-1}(s)b_{t-1}(s)$. Let us assume for now that such a contraction mapping exists. Then, for any $f' \in \mathcal{F}^{|\mathcal{S}|}$

$$
\begin{aligned}
&P(V_\lambda(F_t) > \epsilon | F'_0 = f') \\
&\le E[V_\lambda(F_t)|F'_0 = f']/\epsilon && (26a)\\
&= E[E[V_\lambda(F_t)|Y^{t-1}, F'_0 = f']|F'_0 = f']/\epsilon && (26b)\\
&= E[E[V_\lambda(\psi(F'_{t-1}, B_{t-1}, Y_t))|B_{t-1}, F'_{t-1}]|F'_0 = f']/\epsilon && (26c)\\
&= E[E[V_\lambda(F_t)|B_{t-1}, F'_{t-1}]|F'_0 = f']/\epsilon && (26d)\\
&= E[E\Big[\int_0^1 \lambda(F_t(w))dw|B_{t-1}, F'_{t-1}\Big]|F'_0 = f']/\epsilon && (26e)\\
&\le E[\int_0^1 \xi(\lambda(F_{t-1}(w)))dw|F'_0 = f']/\epsilon && (26f)\\
&\le \xi(E[V_\lambda(F_{t-1})|F'_0 = f'])/\epsilon && (26g)\\
&\dots \\
&\le \xi^t(V_\lambda(f))/\epsilon \to 0, && (26h)
\end{aligned}
$$

where the first inequality is the Markov inequality, the second inequality is due to the assumption for the property of ξ, the third inequality is due to the concavity of ξ, the fourth inequality is due to repeated application of the above inequalities, and the (uniform) convergence to 0 is due to the property of the contraction [14, Lemma 8].

It remains to find the contraction ξ with the aforementioned property. To this end let $\lambda' : [0, 0.5] \to [0,1]$ be a restriction of λ on $[0, 0.5]$. Then, λ' becomes one-to-one and onto, hence it has inverse. Let $\tilde{\xi} : [0,1] \to [0,1]$ be defined as

$$\tilde{\xi}(x) = \sup E[\lambda(\psi(x', b, Y_t))|b_{t-1} = b, f'_{t-1}(w) = x'], \quad (27)$$

where the supremum is over $b, x' = (x'(s))_{s\in\mathcal{S}}$ with $x'(s) \ge 0, \sum_s x'(s)b(s) = x$. The above expression is a supremum of a continuous function over a compact space and thus it is achieved. Consider now the following function

$$\xi^*(x) = \max\left\{\tilde{\xi}(\lambda'^{-1}(x)), \tilde{\xi}(1-\lambda'^{-1}(x))\right\}. \tag{28}$$

Clearly, $\xi^*(x) \geq 0$. We will now show that ξ^* satisfies the aforementioned property. Indeed, let $a \stackrel{\text{def}}{=} f_{t-1}(w)$. If $a \in [0,1/2]$ then $\lambda'^{-1}(\lambda(a)) = a$ and the first term in the maximization on the r.h.s. of (28) (for $x = \lambda(a)$) equals $\tilde{\xi}(a)$. If $a \in [1/2,1]$ then $1-\lambda'^{-1}(\lambda(a)) = a$ and the second term in the maximization of the r.h.s. of (28) (for $x = \lambda(a)$) equals $\tilde{\xi}(a)$. Thus, $\xi^*(\lambda(a)) \geq \tilde{\xi}(a) \geq E[\lambda(\psi(f'_{t-1}(w), b_{t-1}, Y_t))|b_{t-1}, f'_{t-1}]$.

The remaining of the proof is similar to the one developed for DMC's in [14] and will not be repeated here.

■

The above result can be used to show that rate $R = 0$ is achievable.

Proposition 1. *If the channel is fixed-point free, then for $\epsilon, \delta > 0$*

$$\lim_{t\to\infty} P(F_t(W-\delta) > \epsilon) = 0, \tag{29a}$$

$$\lim_{t\to\infty} P(F_t(W+\delta) < 1-\epsilon) = 0. \tag{29b}$$

Proof: The proof is exactly the same as in the DMC case and can be found in [17, Appendix A]. ■

B. Achieving $R < C$

Following similar steps as the ones followed [17] to show capacity achievability for the PMS scheme for finite state channels with both output and state feedback, one may now prove achievability for any rate up to capacity. An alternative proof of achievability may be pursued based on the ideas of stochastic nonlinear filter stability described in [19]. In the following we state the achievability of $R < C$ as a conjecture.

Conjecture 1. *For any $\delta, \alpha > 0$*

$$\lim_{t\to\infty} P(F_{(1+\alpha)t}(W - 2^{-tR}) > \delta) = 0, \tag{30a}$$

$$\lim_{t\to\infty} P(F_{(1+\alpha)t}(W + 2^{-tR}) < 1-\delta) = 0. \tag{30b}$$

VI. Conclusions

Starting from a single-letter capacity expression for unifilar FSC with output feedback, we propose a sequential transmission scheme that is inspired by the PMS developed for memoryless channels and FSC with output and state feedback. The proposed scheme is shown to be have a recursive structure and it possesses the properties that every PMS scheme should possess. The proposed scheme is shown to achieve zero rate for the aforementioned channels by extending the achievability proof developed for memoryless channels. An interesting future research direction is the analysis of the error probability of this (and other) PMS schemes and in particular the error exponent analysis. In addition, an alternative unified proof of the achievability of PMS-like schemes may be pursued based on the ideas of stochastic nonlinear filter stability described in [19].

References

[1] C. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, September 1956.

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

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

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

[5] H. Permuter, P. Cuff, B. V. Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *Information Theory, IEEE Transactions on*, vol. 54, no. 7, pp. 3150 –3165, July 2008.

[6] O. Elishco and H. H. Permuter, "Capacity of the ising channel with feedback," in *Proc. International Symposium on Information Theory*, Aug. 2011, pp. 3004 –3008.

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

[8] G. Como, S. Yuksel, and S. Tatikonda, "The error exponent of variable-length codes over markov channels with feedback," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2139–2160, May 2009.

[9] X. Huang, A. Kavcic, X. Ma, and D. Mandic, "Upper bounds on the capacities of non-controllable finite-state channles using dynamic programming methods," in *Proc. International Symposium on Information Theory*, Seoul, South Korea, June 2009, pp. 2346 – 2350.

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

[11] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 172–182, Apr 1966.

[12] O. Shayevitz and M. Feder, "Communication with feedback via posterior matching," in *Proc. International Symposium on Information Theory*, 24–29 June 2007, pp. 391–395.

[13] ——, "The posterior matching feedback scheme: Capacity achieving and error analysis," in *Proc. International Symposium on Information Theory*, 6–11 July 2008, pp. 900–904.

[14] ——, "Optimal feedback communication via posterior matching," *IEEE Trans. Information Theory*, vol. 57, no. 3, pp. 1186–1222, Mar. 2011.

[15] T. Coleman, "A stochastic control viewpoint to Posterior Matching-style feedback communication schemes," in *Proc. International Symposium on Information Theory*, Seoul, Korea, June 2009, pp. 1520 – 1524.

[16] J. H. Bae and A. Anastasopoulos, "A posterior matching scheme for finite-state channels with feedback," in *Proc. International Symposium on Information Theory*, Austin, TX, June 2010, pp. 2338–2342.

[17] J. Bae, "Capacity-achieving schemes for finite state channels," Ph.D. dissertation, University of Michigan, Ann Arbor, MI, May 2011, can be downloaded from http://eecs.umich.edu/~anastas/docs/junghyun_thesis.pdf.

[18] A. Arapostathis, V. Borkar, E. Fernandez-Gaucherand, M. Ghost, and S. Marcels, "Discrete-time controlled Markov processes with average cost criterion: A survey," *Siam Journal on Control and Optimization*, vol. 31, no. 2, pp. 282–344, Mar. 1993.

[19] S. Gorantla and T. P. Coleman, "Equivalence between reliable feedback communication and nonlinear filter stability," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, 31 2011-aug. 5 2011, pp. 164 –168.