

Zero-rate achievability of posterior matching schemes for channels with memory

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

Abstract—Shayevitz and Feder proposed a capacity-achieving sequential transmission scheme for memoryless channels called posterior matching (PM). The proof of capacity achievability of PM is involved and requires invertibility of the PM kernel (also referred to as one-step invertibility). Recent work by the same authors provided a simpler proof but still requires PM kernel invertibility.

An alternative technique for proving capacity achievability of PM for memoryless channels is due to Ma and Coleman which is based on connections between PM and non-linear filtering. Central to this technique is a different notion of “invertibility” which refers to the property of PM that allows recovering the original message W from knowledge of the observations and the value of the posterior message cdf evaluated at the true message. Unfortunately, this property is also proved by appealing to PM kernel invertibility. As a result, all these techniques cannot be readily generalized to channels with non-invertible PM kernels.

In this paper we analyze PM schemes for channels with memory and in particular unifilar channels with output feedback and intersymbol-interference (ISI) finite state channels with state and output feedback. We follow closely the alternative technique of Ma and Coleman and focus on zero-rate achievability which is the first step of the proof. Our main technical contribution is to show that “invertibility” can be obtained without requiring an invertible PM kernel. This is an indispensable property since in channels with memory the PM kernel is not invertible.

I. INTRODUCTION

Shayevitz and Feder [1] proposed a sequential transmission scheme for memoryless channels with feedback called posterior matching (PM). They showed that PM is capacity achieving following a two-step process. First, zero-rate achievability was proved, and this result was subsequently used to show capacity achievability. The—otherwise ingenious—analysis in [1] has two drawbacks. The first is that it is “quite involved and nontransparent”, as the authors themselves have indicated in a recent work [2], and therefore it is difficult to generalize. Indeed, [2] offers a simpler proof of capacity achievability of PM for memoryless channels. The second drawback is that the proof (both in [1] and [2]) relies on the invertibility of the PM kernel (this is the $F_{\Theta|\Phi}(\cdot|\phi)$ kernel in [1, Lemma 12-(iv)] and the $F_{\Theta|Y}(\cdot|y)$ kernel in [2, property P1]). This property is referred to as the “one-step invertibility”, so as to distinguish it from a more general “invertibility” property of PM to be discussed below. As it turns out this kernel is not invertible in the case of channels with memory and therefore the above methodologies cannot be readily generalized.

An alternative technique for proving that PM is capacity achieving for memoryless channels is due to Ma and Coleman [3] and Gorantla and Coleman [4]. The authors proved zero-rate achievability in [3] by establishing a connection between the PM scheme and non-linear filtering and utilizing the results of Van Handel [5] for maximal accuracy of non-linear filters. Capacity achievability was subsequently obtained in [4] by utilizing a reported equivalence between zero-rate achievability and capacity achievability for PM schemes. Interestingly enough, the methodology of [3] also depends on the invertibility of the PM kernel (one-step invertibility) mentioned above and therefore cannot be easily generalized for channels with non-invertible PM kernels either.

In this work we are motivated by the question of whether PM schemes are capacity-achieving for channels with memory. Such PM schemes have been proposed in [6] for intersymbol-interference (ISI) finite state channels (FSCs) with output and state feedback and in [7] for unifilar channels with output feedback. In view of the proposed methodologies for memoryless channels, in this paper, we focus on the problem of proving zero-rate achievability of PM schemes for channels with memory and in particular for unifilar and FSCs.

Our technique parallels that of [3], where zero-rate achievability is related to the notion of “invertibility”. In particular, invertibility in this context means that the transmitted message W can be recovered from the value of the posterior message cdf at time t evaluated at the message point and all previous observations up to time t . However, invertibility in [3] is trivially obtained from the invertibility of the PM kernel (one-step invertibility). Indeed, the recovery of the message W is done by subsequent applications of the inverse PM kernel. The main contribution of this work is to show how invertibility can be obtained without an invertible PM kernel. This is an indispensable condition since in the channels under consideration the PM kernel is not invertible. Our technique to circumvent the lack of one-step invertibility, may not only be applicable to the methodology of [3], but it can potentially be used to adapt the methodology in [2] as well.

This paper is organized as follows. In section II, we present the model for unifilar channels and summarize the PM scheme proposed in [7]. In section III, we establish the crucial invertibility, and prove zero-rate achievability for unifilar channels. Finally, in section IV we show how these results can be adapted to the case of ISI channels.

II. CHANNEL MODEL AND PM SCHEME

Consider a family of finite-state point-to-point channels with inputs $X_t \in \mathcal{X}$, output $Y_t \in \mathcal{Y}$ and state $S_t \in \mathcal{S}$ at time t , with all alphabets being finite. The channel conditional probability is

$$P(Y_t, S_{t+1} | X^t, Y^{t-1}, S^t) = Q(Y_t | X_t, S_t) \delta_{g(S_t, X_t, Y_t)}(S_{t+1}), \quad (1)$$

for a given stochastic kernel $Q \in \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{P}(\mathcal{Y})$ and deterministic function $g \in \mathcal{S} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{S}$. This family of channels is referred to as unifilar channels. At time t the receiver has access to the current channel output Y_t , while the output Y_t is fed back to the transmitter with unit delay through a noiseless channel.

In [8] the authors formulated the problem of finding the capacity of this channel 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

$$B_t(s) \triangleq P(S_{t+1} = s | Y^t) \quad \forall s \in \mathcal{S}. \quad (2)$$

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

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

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

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

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

$$P(B_t | B_{t-1}) = \sum_{y_t} \delta_{\theta(B_{t-1}, y_t)}(B_t) \left[\sum_{x_t, s_t} Q(y_t | x_t, s_t) \times P(x_t | s_t, B_{t-1}) B_{t-1}(s_t) \right], \quad (5)$$

and θ is given by the recursive updating formula $B_t = \theta(B_{t-1}, Y_t)$ in [7, eq. (6)].

We now summarize the PM scheme proposed in [7]. We assume that the capacity achieving distributions $(\tilde{P}(X_t | S_t, B_{t-1}))_{S_t, B_{t-1}}$ 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 $F_t \in \mathcal{S} \rightarrow \mathcal{F}$ as

$$F_t(w|s) \triangleq P(W \leq w | S_{t+1} = s, Y^t), \quad w \in [0, 1], s \in \mathcal{S}. \quad (6)$$

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 $F_t(w) \stackrel{\text{def}}{=} (F_t(w|s))_{s \in \mathcal{S}} \in [0, 1]^{|\mathcal{S}|}$.

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

$$X_t = F_{\tilde{P}(\cdot | S_t, B_{t-1})}^{-1}(F_{t-1}(W|S_t)) \quad (7a)$$

$$\triangleq e(F_{t-1}(W|S_t), S_t, B_{t-1}), \quad (7b)$$

where the inverse cdf is defined as $F_P^{-1}(y) \triangleq \inf\{x : F_P(x) \geq y\}$ 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 $w \in [0, 1], s \in \mathcal{S}$. The variable B_t can be recursively updated through the update equation $B_t = \theta(B_{t-1}, Y_t)$ as mentioned before, 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 quantity F_t can also be evaluated recursively at both the transmitter and the receiver. In particular, for any $a \in [0, 1]$

$$F_t(a|s') = \frac{\sum_{s_t} B_{t-1}(s_t) h(F_{t-1}(a|s_t), s_t, s', Y_t, B_{t-1})}{B_t(s') \sum_{x, s} Q(Y_t | x, s) \tilde{P}(x|s, B_{t-1}) B_{t-1}(s)}, \quad (8)$$

where $h : [0, 1] \times \mathcal{S} \times \mathcal{S} \times \mathcal{Y} \times \mathcal{P}(\mathcal{S}) \rightarrow [0, 1]$ is defined by

$$h(w, s, s', y, b) = \sum_{x'=0}^{e(w, s, b)-1} \delta_{g(s, x', y)}(s') Q(y | x', s) \tilde{P}(x' | s, b) + \delta_{g(s, e(w, s, b), y)}(s') Q(y | e(w, s, b), s) \left(w - \sum_{x'=0}^{e(w, s, b)-1} \tilde{P}(x' | s, b) \right). \quad (9)$$

We note that, for any given $s, s', y, b \in \mathcal{S} \times \mathcal{S} \times \mathcal{Y} \times \mathcal{P}(\mathcal{S})$, $h(w, s, s', y, b)$ is an increasing function w.r.t w . We will use the shorthand notation

$$F_t = \phi(F_{t-1}, B_{t-1}, Y_t) \quad (10)$$

for the above update. It should be clear from (8) that $F_t(a)$ is a function of F_{t-1} only through $F_{t-1}(a)$. This has important implications for the analysis of the PM scheme. In the following we will also use the notation $F_t(a) = \phi(F_{t-1}, B_{t-1}, Y_t)(a) = \phi(F_{t-1}(a), B_{t-1}, Y_t)$. It should also be clear that for the update of $F_t(a|s)$ the entire vector $F_{t-1}(a) = (F_{t-1}(a|s'))_{s' \in \mathcal{S}}$ is required and not just $F_{t-1}(a|s)$. Let $W_t \in \mathcal{S} \rightarrow [0, 1]$ be defined as

$$W_t \triangleq (W_t(s))_{s \in \mathcal{S}} \text{ with } W_t(s) \triangleq F_t(W|s), \quad \forall s \in \mathcal{S}. \quad (11)$$

With this definition, the encoder in (7) can be expressed using (10) in recursive form as

$$X_t = e(W_{t-1}(S_t), S_t, B_{t-1}) \quad (12a)$$

$$W_t = \phi(W_{t-1}, B_{t-1}, Y_t). \quad (12b)$$

At the receiver, minimum mean square error estimation (MMSE) 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

$$F'_t(w) = \sum_{s_{t+1}} F_t(w|s_{t+1}) B_t(s_{t+1}). \quad (13)$$

Finally, the message estimate is obtained as

$$\tilde{W}_t = E[W | Y^t]. \quad (14)$$

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

III. ZERO-RATE ACHIEVABILITY OF PM FOR UNIFILAR CHANNELS

The communication system achieves zero-rate if

$$P(|\tilde{W}_t - W| > \epsilon) \xrightarrow{t \rightarrow \infty} 0, \quad \forall \epsilon > 0. \quad (15)$$

To prove zero-rate achievability, we generalize the technique of [3], which is based on the result in van Handel [5]. The sketch of the proof is as follows. We will first define a hidden Markov chain which characterizes the PM scheme. We will then show that ergodicity of the hidden Markov chain implies maximal accuracy (defined below) which in turn implies zero-rate achievability. The point of departure from [3] and our main contribution is the proof of invertibility in Lemma 1.

Define $\{(U_t, Z_t)\}_{t \in \mathbb{Z}_-}$ as a hidden Markov chain by

$$U_{-t} = (W_t, S_{t+1}, B_t, Y_{t+1}), \quad (16a)$$

$$Z_{-t} = Y_{t+1} + \kappa N_t, \quad (16b)$$

where $\{N_t\}_t$ is a zero-mean and unit-variance Gaussian process and Markovianity of U_t can be obtained with the updating formulas $B_t = \theta(B_{t-1}, Y_t)$ and $F_t = \phi(F_{t-1}, B_{t-1}, Y_t)$. We now introduce the concept of filter accuracy [3]. Let π be the stationary measure of U_t . For any test function f with $\int f^2 d\pi < \infty$, the filter accuracy after n observations with an arbitrary initial measure $P_{U_{-n}} = \mu$ is

$$e_n^\mu(f, \kappa) \triangleq E[(f(U_0) - E[f(U_0)|\mathcal{F}_{-n:0}^{Z, \kappa}])^2], \quad (17)$$

where $\mathcal{F}_{-n:0}^{Z, \kappa}$ is the sigma field generated by the random variables $(Z_t)_{t=-n}^0$. We can further define

$$e_n(f, \kappa) \triangleq e_n^\pi(f, \kappa), \quad (18)$$

$$e(f, \kappa) \triangleq \lim_{n \rightarrow \infty} e_n(f, \kappa). \quad (19)$$

Definition 1. *The hidden Markov chain $\{(U_t, Z_t)\}_t$ achieves maximal accuracy if $e(f, \kappa) \rightarrow 0$ as $\kappa \rightarrow 0$ for all f with bounded second moment.*

The following proposition shows that under a mild assumption, the hidden Markov chain $\{(U_t, Z_t)\}_t$ achieves maximal accuracy, which further implies zero-rate achievability. This is the main result of this work.

Proposition 1. *If $(U_t)_t$ is ergodic, then the hidden Markov chain $\{(U_t, Z_t)\}_t$ achieves maximal accuracy. Furthermore, maximal accuracy implies zero-rate achievability.*

Proof. We will quote several results from [3] since the proofs are parallel. Fix a function f with finite second moment. By Lemma 2, which will be proved later, $e_n^\mu(f, 0) = 0$ for a specifically chosen measure μ . By an argument similar to that in [3, Lemma 5.2], we have

$$\lim_{\kappa \rightarrow 0} e_n^\mu(f, \kappa) = e_n^\mu(f, 0) = 0 \quad (20)$$

From [3, Lemma 5.3], $\forall \kappa > 0$ and any measure μ ,

$$e_n^\mu(f, \kappa) \xrightarrow{n \rightarrow \infty} e(f, \kappa) \quad (21)$$

By an argument similar to that in [3, Lemma 5.4], we have

$$\lim_{\kappa \rightarrow 0} e(f, \kappa) = e(f, 0). \quad (22)$$

Finally we have

$$\begin{aligned} \lim_{\kappa \rightarrow 0} e(f, \kappa) &\stackrel{(a)}{=} e(f, 0) \stackrel{(b)}{=} \lim_{n \rightarrow \infty} \lim_{\kappa \rightarrow 0} e_n^\mu(f, \kappa) \\ &\stackrel{(c)}{=} \lim_{n \rightarrow \infty} e_n^\mu(f, 0) \stackrel{(d)}{=} 0, \end{aligned} \quad (23)$$

where (a) is due to (22), (b) is due to (21), and (c), (d) are due to (20). Therefore the hidden Markov chain achieves maximal accuracy.

In the converse proof of [3, Lemma 3.5], maximal accuracy results in $W = E[W|Y^\infty]$. Therefore, we have

$$\tilde{W}_t = E[W|Y^t] \xrightarrow{t \rightarrow \infty} E[W|Y^\infty] = W, \quad (24)$$

where (a) is due to the martingale convergence theorem. (24) implies that \tilde{W}_t converges to W in probability, and thus the zero-rate achievability. \square

Before proving Lemma 2, we would like to mention that the essential idea of the proof is to justify the concept of ‘‘invertibility’’ of the PM scheme. The following lemma states this concept more rigorously.

Lemma 1. *Given the information $(W_t, S_{t+1}, B_t, Y^{t+1})$, $W_{t'}$ can be recovered for all $t' \leq t$.*

Note that if the ϕ kernel in (12b) were invertible, then the above property would be trivially satisfied by repeated applications of ϕ^{-1} (since Y^{t+1} is known and B^t can be recovered from Y^t through the recursive kernel θ). However, in our setting, the ϕ kernel in (12b) is not invertible. The proof of the above lemma hinges on the idea that the message W can be recovered from every W_t . Since $W_{t'}$ can be derived from W (with a forward recursion), $W_{t'}$ can also be recovered from W_t .

Proof. We will show that given the information $(W_t, S_{t+1}, B_t, Y^{t+1})$, W can be recovered and thus $W_{t'}$ can be recovered for all $t' \leq t$ by recursive application of the ϕ mapping in (12b) (together with the θ mapping).

We claim that the message W is a.s. equal to the random variable (measurable w.r.t. $(W_t, S_{t+1}, B_t, Y^{t+1})$)

$$\hat{W} \triangleq \max\{w | F_t(w|S_{t+1}) = W_t(S_{t+1})\}. \quad (25)$$

Indeed, w.p. 1 we have $\hat{W} \geq W$ by the definition of \hat{W} , and

$$\hat{W} > W \stackrel{(*)}{\Rightarrow} F_t(\hat{W}|S_{t+1}) > F_t(W|S_{t+1}) = W_t(S_{t+1}), \quad (26)$$

which contradicts the definition of \hat{W} , and therefore $\hat{W} = W$ a.s.. In the following we establish the validity of (*).

Define a partial order on $u, v \in \mathbb{R}^{|S|}$ by

$$u \succeq v \Leftrightarrow u(s) \geq v(s) \quad \forall s \in S. \quad (27)$$

We will show that given the information $(W_t, S^{t+1}, B_t, Y^{t+1})$, for all $i \leq t-1$,

$$\begin{aligned} \{F_i(\hat{W}|S_{i+1}) > W_i(S_{i+1}), F_i(\hat{W}|\cdot) \succeq W_i\} \Rightarrow \\ \{F_{i+1}(\hat{W}|S_{i+2}) > W_{i+1}(S_{i+2}), F_{i+1}(\hat{W}|\cdot) \succeq W_{i+1}\}. \end{aligned} \quad (28)$$

From the updating formula in (8), for any $s' \in \mathcal{S}$, we have

$$\begin{aligned} F_{i+1}(\hat{W}|s') &= \frac{\sum_s B_i(s) h(F_i(\hat{W}|s), s, s', Y_{i+1}, B_i)}{B_{i+1}(s') \sum_{x,s} Q(Y_{i+1}|x, s) \tilde{P}(x|s, B_i) B_i(s)} \\ &\stackrel{(a)}{\geq} \frac{\sum_s B_i(s) h(W_i(s), s, s', Y_{i+1}, B_i)}{B_{i+1}(s') \sum_{x,s} Q(Y_{i+1}|x, s) \tilde{P}(x|s, B_i) B_i(s)} \\ &= F_{i+1}(W|s') = W_{i+1}(s'), \end{aligned} \quad (29)$$

where (a) is due to that $F_i(\hat{W}|\cdot) \succeq W_i$. Since we established that $F_{i+1}(\hat{W}|\cdot) \succeq W_{i+1}$, to show $F_{i+1}(\hat{W}|S_{i+2}) > W_{i+1}(S_{i+2})$ it is sufficient to prove that $F_{i+1}(\hat{W}|S_{i+2}) = W_{i+1}(S_{i+2})$ results in a contradiction. Suppose $F_{i+1}(\hat{W}|S_{i+2}) = W_{i+1}(S_{i+2})$. Then,

$$\begin{aligned} P(B_i(S_{i+1}) = 0) &= \int_b \sum_s P_{B_i}(db) P(S_{i+1} = s | B_i = b) \delta_{b(s)}(0) \\ &= \int_b \sum_s P_{B_i}(db) b(s) \delta_{b(s)}(0) = 0, \end{aligned} \quad (30)$$

which means $B_i(S_{i+1}) > 0$ a.s.. Since $h(w, s, s', y, b)$ is increasing w.r.t. w for any s, s', y, b , from the above equation together with the fact that $F_{i+1}(\hat{W}|S_{i+2}) = W_{i+1}(S_{i+2})$ and $F_{i+1}(\hat{W}|\cdot) \succeq W_{i+1}$ we deduce that w.p. 1

$$\begin{aligned} h(F_i(\hat{W}|S_{i+1}), S_{i+1}, S_{i+2}, Y_{i+1}, B_i) &= \\ h(W_i(S_{i+1}), S_{i+1}, S_{i+2}, Y_{i+1}, B_i). \end{aligned} \quad (31)$$

Now, for any $i \geq 0$, $\delta_{g(S_{i+1}, X_{i+1}, Y_{i+1})}(S_{i+2}) = 1$ w.p. 1, due to the channel update. In addition,

$$\begin{aligned} P(Q(Y_{i+1}|S_{i+1}, X_{i+1}) = 0) &= \sum_{s,x,y} P(S_{i+1} = s, X_{i+1} = x) Q(y|s, x) \delta_{Q(y|s,x)}(0) \\ &= 0, \end{aligned} \quad (32)$$

which shows $Q(Y_{i+1}|S_{i+1}, X_{i+1}) > 0$ w.p. 1. Therefore we have w.p. 1

$$\delta_{g(S_{i+1}, X_{i+1}, Y_{i+1})}(S_{i+2}) Q(Y_{i+1}|S_{i+1}, X_{i+1}) > 0, \quad (33)$$

which implies $h(\cdot, S_{i+1}, S_{i+2}, Y_{i+1}, B_i)$ is strictly increasing. Thus (31) results in $F_i(\hat{W}|S_{i+1}) = W_i(S_{i+1})$, which draws a contradiction to the condition $F_i(\hat{W}|S_{i+1}) > W_i(S_{i+1})$. Hence the statement in (28) is justified. \square

We are now ready to state Lemma 2 required in the proof of Proposition 1.

Lemma 2. Define a measure $\mu(dw, ds, db, dy) \triangleq \delta_{w'}(w) \delta_{s'}(s) \delta_{b'}(b) Q(dy|e(w'(s'), s', b'), s')$ for some $w' \in [0, 1]^{|S|}$, $s' \in \mathcal{S}$, and $b' \in \mathcal{P}(\mathcal{S})$. Then we have $e^\mu(f, 0) = 0$.

Proof.

$$\begin{aligned} e_n^\mu(f, 0) &= E^\mu \left[\left(f(U_0) - E^\mu \left[f(U_0) | \mathcal{F}_{-n;0}^{Z,0} \right] \right)^2 \right] \\ &= E^\mu \left[\left(f(W_0, S_1, B_0, Y_1) - E^\mu \left[f(W_0, S_1, B_0, Y_1) | \mathcal{F}_{1;n+1}^Y \right] \right)^2 \right]. \end{aligned} \quad (34)$$

By Lemma 1, we have $f(W_0, S_1, B_0, Y_1) = E^\mu [f(W_0, S_1, B_0, Y_1) | \mathcal{F}_{1;n+1}^Y, W_n = w', S_{n+1} = s', B_n = b']$. Thus

$$e^\mu(f, 0) = \lim_{n \rightarrow \infty} e_n^\mu(f, 0) = 0. \quad (35)$$

\square

Looking at Proposition 1 and the two lemmas presented above, only Lemma 1 significantly depends on the channel model, while the others are transparent to the specific channel model. In the next section, we will prove the counterpart of Lemma 1 in a PM scheme over FSCs with ISI, which according to the previous analysis is sufficient for zero-rate achievability.

IV. ZERO-RATE ACHIEVABILITY OF FSCs WITH ISI

We now consider a family of finite-state point-to-point channels with inputs $X_t \in \mathcal{X}$, output $Y_t \in \mathcal{Y}$ and state $S_t \in \mathcal{S}$ at time t , with all alphabets being finite. The channel conditional probability is

$$P(Y_t, S_{t+1} | X^t, Y^{t-1}, S^t) = Q(Y_t | X_t, S_t) Q'(S_{t+1} | S_t, X_t), \quad (36)$$

for given stochastic kernels $Q \in \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{P}(\mathcal{Y})$ and $Q' \in \mathcal{S} \times \mathcal{X} \rightarrow \mathcal{P}(\mathcal{S})$. At time t the receiver has access to the current channel output Y_t and state S_t , while the Y_t and S_t are fed back to the transmitter with unit delay through a noiseless channel. Similar to the case of unifilar FSCs, the capacity can be derived by formulating an appropriate stochastic control problem [6][9]. In particular, a controlled Markov process is defined with state $S_t \in \mathcal{S}$ and $B_t \in \mathcal{P}(\mathcal{X})$, defined as

$$B_t(x) \triangleq P(X_t = x | S^t, Y^t). \quad (37)$$

The capacity is obtained under mild conditions in a single-letter expression of the form

$$C = \sup_{P(X_t | X_{t-1}, S_{t-1}, B_{t-1})} I(X_t, X_{t-1} \wedge S_t, Y_t | S_{t-1}, B_{t-1}). \quad (38)$$

Bae and Anastasopoulos [6] proposed the following PM scheme for this family of channels. We assume that the capacity achieving distributions $(\hat{P}(X_t | X_{t-1}, S_{t-1}, B_{t-1}))_{X_{t-1}, S_{t-1}, B_{t-1}}$ have been found for all values of $(X_{t-1}, B_{t-1}) \in \mathcal{X} \times \mathcal{P}(\mathcal{X})$. 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 $F_t \in \mathcal{X} \rightarrow \mathcal{F}$ as

$$F_t(w|x) \triangleq P(W \leq w | X_t = x, S^t, Y^t), \quad w \in [0, 1] \text{ and } x \in \mathcal{X}. \quad (39)$$

With the above definition, $F_t(\cdot|x)$ is the a-posteriori cdf of W conditioned on $X_t = x, S^t, Y^t$, and F_t is the collection of all these conditional cdfs for all $x \in \mathcal{X}$. Let $W_t \in \mathcal{X} \rightarrow [0, 1]$ be defined as

$$(34) \quad W_t \triangleq (W_t(x))_{x \in \mathcal{X}} \text{ with } W_t(x) \triangleq F_t(W|x) \quad \forall x \in \mathcal{X}. \quad (40)$$

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

$$X_t = F_{\hat{P}(\cdot|X_{t-1}, S_{t-1}, B_{t-1})}^{-1}(F_{t-1}(W|X_{t-1})) \quad (41a)$$

$$\stackrel{\text{def}}{=} \hat{e}(F_{t-1}(W|X_{t-1}), X_{t-1}, S_{t-1}, B_{t-1}). \quad (41b)$$

The quantities B_t, F_t can be updated by the formulas

$$B_t = \theta(B_{t-1}, Y_t, S_t, S_{t-1}) \quad (42a)$$

$$F_t = \phi(F_{t-1}, S_t, S_{t-1}, B_{t-1}) \quad (42b)$$

where θ, ϕ are defined in [6]. In particular, the mapping ϕ can be expressed explicitly as

$$F_{t+1}(a|x) = \frac{\sum_{x'} Q'(S_{t+1}|S_t, x') B_t(x') \hat{h}(F_t(a|x'), x, x', S_t, B_t)}{\sum_{\tilde{x}} Q'(S_{t+1}|S_t, \tilde{x}) B_t(\tilde{x}) \hat{P}(x|\tilde{x}, S_t, B_t)}, \quad (43)$$

where $\hat{h}: [0, 1] \times \mathcal{X} \times \mathcal{X} \times \mathcal{S} \times \mathcal{P}(\mathcal{X}) \rightarrow [0, 1]$ is defined by

$$\hat{h}(w, x, x', s', b) = \begin{cases} 0 & , w \leq \sum_{\tilde{x}=0}^{x-1} \hat{P}(\tilde{x}|x', s', b) \\ w - \sum_{x'=0}^{x-1} \hat{P}(\tilde{x}|x', s', b) & , \sum_{x'=0}^{x-1} \hat{P}(\tilde{x}|x', s', b) \leq w \\ \hat{P}(x|x', s', b) & , w \geq \sum_{\tilde{x}=0}^x \hat{P}(\tilde{x}|x', s', b) \end{cases}. \quad (44)$$

For any given $x, x', s', b \in \mathcal{X} \times \mathcal{X} \times \mathcal{S} \times \mathcal{P}(\mathcal{X})$, $\hat{h}(w, x, x', s', b)$ is an increasing function w.r.t w . The initial posterior cdfs are set to be uniform, i.e., $F_0(w|x) = w$ for $w \in [0, 1]$ and $x \in \mathcal{X}$. It is easy to see that the role of the channel state in unifilar channels is replaced by the input X_t in FSCs with ISI. We further define $\{(U_t, Z_t)\}_{t \in \mathbb{Z}_-}$ as a hidden Markov chain by

$$U_{-t} = (W_t, B_t, X_t, S_t, S_{t+1}, Y_{t+1}) \quad (45a)$$

$$Z_{-t} = (S_{t+1}, Y_{t+1}) + \kappa(N_t^1, N_t^2), \quad (45b)$$

where $\{N_t^1\}_t$ and $\{N_t^2\}_t$ are two independent zero-mean unit-variance Gaussian processes and Markovianity of U_t can be obtained with the updating formulas θ and ϕ in (42).

We now present the ‘‘invertibility’’ result which is the counterpart of Lemma 1.

Lemma 3. *Given the information $(W_t, B_t, X_t, Y^{t+1}, S^{t+1})$, $W_{t'}$ can be recovered for all $t' \leq t$.*

Proof. The proof parallels closely that of Lemma 1.

We claim that the message W is a.s. equal to the random variable (measurable w.r.t. $(W_t, B_t, X_t, Y^{t+1}, S^{t+1})$)

$$\hat{W} \triangleq \max\{w | F_t(w|X_t) = W_t(X_t)\}. \quad (46)$$

For this it is sufficient to prove a claim similar to that in (26)

$$\hat{W} > W \stackrel{(*)}{\Rightarrow} F_t(\hat{W}|X_t) > F_t(W|X_t) = W_t(X_t). \quad (47)$$

To do this we define a partial order on $u, v \in \mathbb{R}^{|\mathcal{X}|}$ in a way similar to (27). We will show that given the information $(W_t, B_t, X_t, Y^{t+1}, S^{t+1})$, for all $i \leq t-1$,

$$\{F_i(\hat{W}|X_i) > W_i(X_i), F_i(\hat{W}|\cdot) \succeq W_i\} \Rightarrow \{F_{i+1}(\hat{W}|X_{i+1}) > W_{i+1}(X_{i+1}), F_{i+1}(\hat{W}|\cdot) \succeq W_{i+1}\}. \quad (48)$$

Consequently, for any given $x \in \mathcal{X}$,

$$\begin{aligned} & F_{i+1}(\hat{W}|x) \\ &= \frac{\sum_{x'} Q'(S_{i+1}|S_i, x') B_i(x') \hat{h}(F_i(\hat{W}|x'), x, x', S_i, B_i)}{\sum_{\tilde{x}} Q'(S_{i+1}|S_i, \tilde{x}) B_i(\tilde{x}) \hat{P}(x|\tilde{x}, S_i, B_i)} \\ &\stackrel{(a)}{\geq} \frac{\sum_{x'} Q'(S_{i+1}|S_i, x') B_i(x') \hat{h}(W_i(x'), x, x', S_i, B_i)}{\sum_{\tilde{x}} Q'(S_{i+1}|S_i, \tilde{x}) B_i(\tilde{x}) \hat{P}(x|\tilde{x}, S_i, B_i)} \\ &= F_{i+1}(W|x) = W_{i+1}(x), \end{aligned} \quad (49)$$

where (a) is due to that $F_i(\hat{W}|\cdot) \succeq W_i$. By a similar argument used in Lemma 1, $F_{i+1}(\hat{W}|X_{i+1}) > W_{i+1}(X_{i+1})$ can be derived, and therefore the statement in (48) is justified. \square

With Lemma 3, together with the counterpart of Lemma 2 the following proposition can be shown to be true.

Proposition 2. *If the joint process $\{(W_t, B_t, X_t, S_t, S_{t+1}, Y_{t+1})\}_t$ is ergodic, then the PM scheme for the FSC with ISI achieves zero rate.*

V. CONCLUSION

In this paper, we proved zero-rate achievability of PM schemes for both unifilar FSCs and FSCs with ISI. We generalized the technique of Ma and Coleman [3] and established the crucial ‘‘invertibility’’ result for PM schemes with non-invertible PM kernels.

A subsequent future research problem is the proof of capacity achievability of PM schemes for channels with memory. There are two directions one may follow towards this goal. The first is to analyze the forward-iterated behavior of PM schemes (based on the result of zero-rate achievability) as in [1], or in [3], [4]. The second is to analyze the reverse-iterated behavior of PM schemes as in [2] based on the invertibility result of this paper.

REFERENCES

- [1] O. Shayevitz and M. Feder, ‘‘Optimal feedback communication via posterior matching,’’ *Information Theory, IEEE Transactions on*, vol. 57, no. 3, pp. 1186–1222, 2011.
- [2] —, ‘‘A simple proof for the optimality of randomized posterior matching,’’ *arXiv preprint arXiv:1507.08929*, 2015.
- [3] R. Ma and T. P. Coleman, ‘‘Necessary and sufficient conditions for reliability of posterior matching in arbitrary dimensions,’’ in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 2366–2370.
- [4] 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*. IEEE, 2011, pp. 164–168.
- [5] R. Van Handel, ‘‘When do nonlinear filters achieve maximal accuracy?’’ *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3151–3168, 2009.
- [6] J. H. Bae and A. Anastasopoulos, ‘‘A posterior matching scheme for finite-state channels with feedback,’’ in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 2338–2342.
- [7] A. Anastasopoulos, ‘‘A sequential transmission scheme for unifilar finite-state channels with feedback based on posterior matching,’’ in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 2914–2918.
- [8] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, ‘‘Capacity of the trapdoor channel with feedback,’’ *IEEE Transactions on Information Theory*, vol. 54, no. 7, 2008.
- [9] S. Tatikonda and S. Mitter, ‘‘The capacity of channels with feedback,’’ *Information Theory, IEEE Transactions on*, vol. 55, no. 1, pp. 323–349, 2009.