On the Sufficiency of Power Control for a Class of Channels with Feedback

Desmond S. Lun and Muriel Médard
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
E-mail: {dslun, medard}@mit.edu

***Abstract*— We show that, for a particular class of channels that we believe applies to many physical problems of interest, the utility of feedback, insofar as channel capacity is concerned, is simply for allowing the transmitter to perform power control. This class of channels, which assumes noiseless feedback but allows for the feedback to be of arbitrary rate, includes channels that model slow, flat fading channels with variable input power. Thus our result gives some guidance on the design of effective transmission schemes for slow, flat fading channels with feedback. Some technical results relating to the capacity of channels with noiseless feedback are also demonstrated and a dynamic programming formulation for the optimal power control policy problem is proposed.**

I. Introduction

Many of the information-theoretic results that we currently have regarding channels with memory assume the presence of some channel state information (CSI) at the transmitter, the receiver, or both (see, for example, [1], [2], [3]); but while the presence of CSI often greatly increases tractability, it also renders solutions somewhat unsatisfactory since we still need to consider the problem of obtaining CSI in the first place. In most situations, CSI at the receiver is obtained by sending symbols over the channel, so for full generality, such symbols should be incorporated as part of a codebook designed for a channel with no CSI; and CSI at the transmitter is obtained by means of a separate feedback channel. We have several results that address the scenario where there is no CSI and no feedback (see, for example, [4], [5], [6]); but without feedback, we cannot have adaptation at the transmitter, which we expect to be beneficial. In the scenario where there is feedback and no CSI, however, results are scant. We have little idea of what the beneficial effect of feedback can be when there is no CSI (the case with receiver CSI is quite well understood, at least for a finite-state Markov channel [3]) and how it is provided. In short, we have difficulty answering the question, ‘What is feedback good for?’

In this paper, we show that, for a particular class of channels that we believe applies to many physical problems of interest, the utility of feedback, insofar as channel capacity is concerned, is simply for allowing the transmitter to perform power control. More precisely, we show that if noiseless feedback of an arbitrary rate is available, then capacity is achieved by adopting an optimal policy for controlling the power based on the feedback coupled with appropriate channel coding. Note, however, that determining the optimal power control policy is not a straightforward problem in general: In choosing the amount of energy to use for a particular symbol, we need to consider not only the amount of information that it conveys, but also how it will affect the knowledge of the channel state at both the receiver and the transmitter.

II. Channel model

A block diagram of the channel model is shown in Figure 1. We assume that the feedforward channel has finite input and output alphabets, which we denote by $\mathcal{X}$ and $\mathcal{Y}$ respectively. The feedback link is assumed to be noiseless with finite input and output alphabet $\mathcal{U}$.

The channel is used for a total time of N. A message W is mapped to a sequence of random *code-functions* F^N, each element of which maps past symbols from the feedback link to input symbols for the feedforward channel, i.e. $F_n : \mathcal{U}^{n-1} \to \mathcal{X}$, so $X_n = F_n(U^{n-1})$. At the receiver end, there is a sequence of deterministic functions g^N, each element of which maps output symbols from the feedforward channel to symbols for the feedback link, i.e. $g_n : \mathcal{Y}^n \to \mathcal{U}$, so $U_n = g_n(Y^n)$. The symbols that are fed back are received at the transmitter after a delay of one time unit. When the full output sequence from the feedforward channel Y^N is received, it is decoded to obtain a message estimate $\hat{W}$. Note that the assumptions placed on the feedback link are not necessarily unrealistic: first, it is discrete and not continuous (in which case noiselessness is much more difficult to justify); and secondly, we place no restriction on $|\mathcal{U}|$ nor do we require g_n to be surjective, hence the feedback link, though it must be noiseless, can be of arbitrary rate.

The state process takes one of a number states at each time n from a finite set $\mathcal{S}$ and the channel output at time n depends only on the channel input and state at time n; more precisely, we assume that

$$P_{Y_n|X^n,S^n}(y_n|x^n, s^n) = P_{Y_n|X_n,S_n}(y_n|x_n, s_n). \qquad (1)$$

Moreover, we assume that the state at time n is independent of previous input/output pairs when conditioned on previous states, i.e. $(X^{n-1}, Y^{n-1}) \to S^{n-1} \to S_n$ forms a Markov chain. We refer to a channel satisfying the assumptions made so far on the feedforward channel as a *finite-state channel* (FSC). If, moreover, the state process of an FSC is Markov, we say that it is a *finite-state Markov channel* (FSMC).

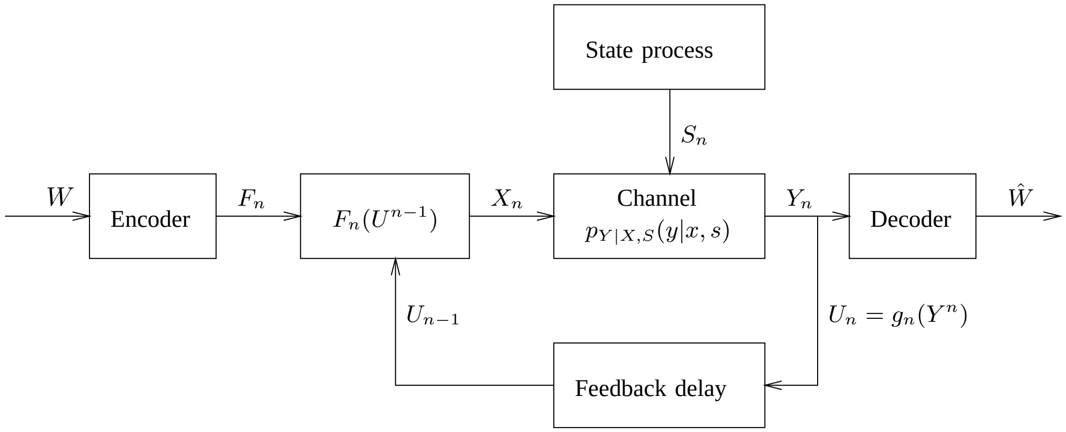

Fig. 1. Block diagram of the channel model.

III. The capacity of channels with noiseless feedback

The capacity of channels with full noiseless feedback has been extensively studied by Tatikonda and Mitter [7], [8]. In this section, we extend their results to channels with noiseless feedback of arbitrary rate (i.e. channels of the type described in Section II, but without any assumption on the behavior or even existence of the state process; in particular, the channel need neither be a FSC or a FSMC) and discuss some additional issues.

Theorem 1: The capacity of channels with noiseless feedback of arbitrary rate is given by[1]

$$C = \sup_{\mathbf{P}} \text{p-liminf} \frac{1}{N} \sum_{n=1}^{N} i_n(Y_n; X^n | Y^{n-1}), \tag{2}$$

where

$$i_n(Y_n; X^n|Y^{n-1}) := \frac{P_{Y_n|Y^{n-1},X^n}(Y_n|Y^{n-1}, X^n)}{P_{Y_n|Y^{n-1}}(Y_n|Y^{n-1})} \tag{3}$$

is the *sample mutual information* or *mutual information density* between Y_n and X^n conditioned on Y^{n-1}, and $\mathbf{P}$ is the sequence of conditional probability mass functions $P_{X_1}, P_{X_2|X_1,U_1}, P_{X_3|X^2,U^2}, \ldots$.

Proof: Since the channel between $\{F_n\}$ and $\{Y_n\}$ is feedback-free, we obtain, by applying the results of Verdu and Han[9],

$$C = \sup_{\mathbf{F}} \text{p-liminf} \frac{1}{N} i_N(F^N; Y^N), \tag{4}$$

where

$$i_N(F^N; Y^N) = \log \frac{P_{Y^N|F^N}(Y^N|F^N)}{P_{Y^N}(Y^N)} \tag{5}$$

and $\mathbf{F}$ denotes the space of probability distributions on $\{F_n\}$.

[1]We use p-liminfA_n to denote the liminf in probability of a sequence of random variables $\{A_n\}$, i.e. the supremum of all real numbers α such that $\Pr(A_n \leq \alpha) \to 0$ as $n \to \infty$.

Now, for all (f^N, x^N, y^N, u^N) with $x_n = f_n(u^{n-1})$ and $u_n = g_n(y^n)$ for $n = 1, \ldots, N$, we have

$$\begin{aligned} \log \frac{P_{Y^N|F^N}(Y^N|F^N)}{P_{Y^N}(Y^N)} &= \log \prod_{n=1}^{N} \frac{P_{Y_n|Y^{n-1},F^N}(y_n|y^{n-1}, f^N)}{P_{Y_n|Y^{n-1}}(y_n|y^{n-1})} \\ &= \sum_{n=1}^{N} \log \frac{P_{Y_n|Y^{n-1},X^n}(y_n|y^{n-1}, x^n)}{P_{Y_n|Y^{n-1}}(y_n|y^{n-1})}. \end{aligned} \tag{6}$$

Therefore

$$\frac{1}{N} i_N(F^N; Y^N) = \frac{1}{N} \sum_{n=1}^{N} i_n(Y_n; X^n|Y^{n-1}) \quad \text{a.s.} \tag{7}$$

The fact that taking the supremum over $\mathbf{P}$ is equivalent to taking the supremum over $\mathbf{F}$ is discussed in [8, Section 4.4.1] and in [7, Section 5.1] for the special case where $u_n = y_n$. The generalization to $u_n = g_n(y^n)$ is straightforward. ■

Remark. The above result generalizes easily for delays greater than one. We simply incorporate such additional delays into the sequence of functions $\{g_n\}$.

Remark. The *directed mutual information* $I(X^N \to Y^N)$ is frequently used in work related to channels with feedback (see, for example, [7], [8], [10], [11]). We avoid such notation in our work since we believe that it would be more of a hindrance than an aid. The reader should note, however, that our work is consistent with the works cited above that use directed mutual information.

The problem with the capacity expression (2) is that the liminf in probability of a general random sequence is very difficult to evaluate. Hence we are interested establishing conditions under which the expression simplifies.

Proposition 1: If the sequence of pairs $\{(X_n, Y_n)\}$ is ergodic, then

$$\text{p-liminf}\frac{1}{N}\sum_{n=1}^{N} i_n(Y_n;X^n|Y^{n-1}) = \lim_{n\to\infty}\frac{1}{n}I(X^n;Y_n|Y^{n-1}). \quad (8)$$

Proof: The sum of mutual information densities on the left-hand side of (8) is equal to that that would result if $\{X_n\}$ and $\{Y_n\}$ were related by a causal channel without feedback (i.e. a channel satisfying $P_{Y^n|X^\infty} = P_{Y^n|X^n}$ for all n). Hence we obtain the desired result upon application of Theorem 5 of [12]. ∎

For the case of feedforward channels, we know that there is a large class of channels called ergodic channels (that includes memoryless channels, block memoryless channels, output mixing channels [13], and some Markov channels [14]) with the property that $\{(X_n,Y_n)\}$ is ergodic whenever $\{X_n\}$ is. Moreover, for stationary channels, it suffices to consider only ergodic input sequences $\{X_n\}$ for the computation of channel capacity [15]. Thus, for stationary and ergodic channels, capacity can be computed without recourse to limits of mutual information densities. Unfortunately, the same characterization does not apply when feedback is present; at least not straightforwardly, since whether or not $\{X_n,Y_n\}$ is ergodic depends both on the channel and the choice of code-functions. Nevertheless, we know, from Theorem 8 of [9], that

$$\text{p-liminf}\frac{1}{N}\sum_{n=1}^{N} i_n(Y_n;X^n|Y^{n-1}) \le \lim_{n\to\infty}\frac{1}{n}I(X^n;Y_n|Y^{n-1}) \quad (9)$$

for all sequences $\{(X_n,Y_n)\}$. Hence if we find $P \in \mathbf{P}$ that achieves the supremum of $\lim_{n\to\infty}(1/n)I(X^n;Y_n|Y^{n-1})$ and such that $\{(X^n,Y^n)\}$ is ergodic, then it is clear that P is a capacity-achieving input distribution.

IV. Uniformly-symmetric variable-noise variable-power FSCs

We now proceed to specify the class of feedforward channels whose capacities depend solely on the input energy, and thus for which the utility of feedback is simply for power control. Our definitions are based on those given in [4] but are expanded to allow for the concept of power control at the transmitter.

Definition 1: We say that an FSC is *variable-power* with L energy levels if it is associated with L disjoint, non-empty sets $\mathcal{X}_1,\ldots,\mathcal{X}_L$ such that $\mathcal{X} = \bigcup_{l=1}^{L}\mathcal{X}_l$; and a sequence of *input energies* $\{E_n\}$ such that $E_n = l$ if the input symbol at time n is from alphabet $\mathcal{X}_l$. The FSC that results when $E_n = l$ for all n is called the FSC with input energy l.

We assume that $\mathcal{X}_1,\ldots,\mathcal{X}_L$ is indexed in such a way that the capacity of the FSC with input energy l and perfect CSI is greater than or equal to that with input energy l' if $l > l'$.

Definition 2: For a discrete memoryless channel, let M denote the matrix of input/output probabilities

$$M_{ij} = p_{Y|X}(j|i), \quad j \in \mathcal{Y},\, i \in \mathcal{X}. \quad (10)$$

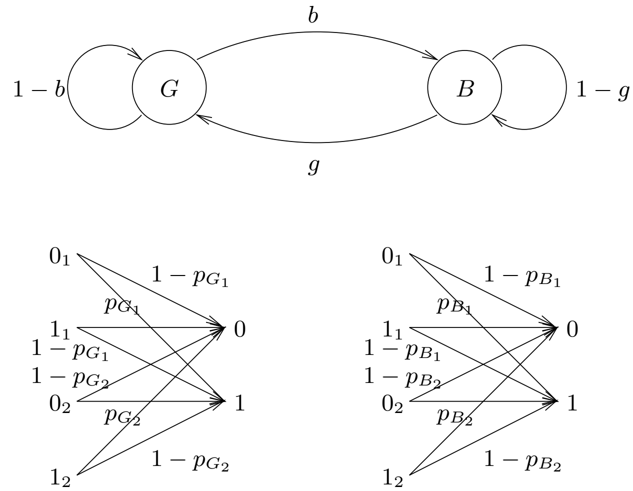

Fig. 2. Gilbert-Elliot channel with two input energy levels.

We call the channel *output-symmetric* if the rows of M are permutations of each other and the columns of M are permutations of each other.

Definition 3: A FSC is *uniformly-symmetric* if, for every state $s \in \mathcal{S}$, the discrete memoryless channel with input/output probabilities given by $p_{Y|X,S}(\cdot|\cdot,s)$ is output-symmetric.

A variable-power FSC with L energy levels is uniformly-symmetric if the FSC with input energy l is uniformly-symmetric for all $l = 1,\ldots,L$.

Definition 4: Let X_n and Y_n denote the input and output, respectively, of an FSC. We call an FSC *variable-noise* if there exists a function φ such that $Z_n = \varphi(X_n,Y_n)$, Z^n is independent of X^n, and Z^n is a sufficient statistic for S^n (i.e. $(X^n,Y^n) \to Z^n \to S^n$ forms a Markov chain).

A variable-power FSC is variable-noise if Z^n is independent of X^n when conditioned on E^n and (Z^n,E^n) is a sufficient statistic for S^n.

We consider the class of uniformly-symmetric variable-noise variable-power FSCs. Note that, in defining FSCs, we have assumed that $(X^{n-1},Y^{n-1}) \to S^{n-1} \to S_n$ forms a Markov chain, so the channel cannot explicitly exhibit intersymbol interference (ISI); and indeed, if the channel did exhibit ISI, we would not expect the capacity to depend solely on the input energy. We would expect, rather, the choice of individual input symbols to be very important and not simply their energies. Some results regarding a particular ISI channel, the finite-state machine channel, have been obtained by Yang, Kavčić, and Tatikonda [16], [17].

Example. The channel illustrated in Figure 2, which we call the Gilbert-Elliot channel with two input energy levels, is an example of a uniformly-symmetric variable-noise variable-power FSMC. It has two states, a "good" state designated by G and a "bad" state designated by B. The state process evolves as a first-order Markov chain with transition probabilities between states of g and b. In each state, we can divide the input alphabet into two disjoint sets, $\{0_1,1_1\}$ and $\{0_2,1_2\}$, which represent two differing input energy levels. Assuming that $\{0_2, 1_2\}$ represents a higher input energy than $\{0_1, 1_1\}$, we would have $p_{G_1} > p_{G_2}$ and $p_{B_1} > p_{B_2}$, as well as $p_{B_1} > p_{G_1}$ and $p_{B_2} > p_{G_2}$. This channel could constitute a rather crude model for a slow, flat fading channel with variable input power. By adding additional states, energy levels, and alphabet symbols and perhaps increasing the order of the Markov chain, a more accurate model can be formed [18], [19], [20], [21].

Example. A slow fading channel with independent frequency sub-bands can be crudely modeled by several Gilbert-Elliot channels with two input energy levels in parallel. In this case, it is not hard to see that we still have a uniformly-symmetric variable-noise variable-power FSMC. Now the state is described by a vector as are the possible input energies.

The following theorem is our main result.

Theorem 2: Suppose we have a uniformly-symmetric variable-noise variable-power FSC with noiseless feedback of arbitrary rate. Let

$$\begin{aligned}\mathbf{P}_e := \{&(P_{X_1}, P_{X_2|X_1,U_1}, P_{X_3|X^2,U^2}, \ldots) \mid \\ &\exists (P_{E_1}, P_{E_2|X_1,U_1}, P_{E_3|X^2,U^2}, \ldots) \text{ s.t. for all } n \\ &P_{X_n|X^{n-1},U^{n-1}}(x_n|x^{n-1}, u^{n-1}) \\ &= \frac{P_{E_n|X^{n-1},U^{n-1}}(\lambda(x_n)|x^{n-1}, u^{n-1})}{|\mathcal{X}_{\lambda(x_n)}|} \\ &\text{for all } x^n \in \mathcal{X}^n, u^{n-1} \in \mathcal{U}^{n-1}\},\end{aligned} \tag{11}$$

where $\lambda : \mathcal{X} \to \{1, \ldots, L\}$, $\lambda(x) = l$ if $x \in \mathcal{X}_l$. Suppose P^* achieves the supremum over $\mathbf{P}_e$ of $\lim_{n\to\infty}(1/n)I(X^n; Y_n|Y^{n-1})$ and that $\{(X_n, Y_n)\}$ is ergodic under the input distribution P^*. Then the capacity of the channel is given by

$$C = \sup_{P\in\mathbf{P}_e} \lim_{n\to\infty} \frac{1}{n} I(X^n; Y_n|Y^{n-1}). \tag{12}$$

Proof: Consider an input distribution $P \in \mathbf{P}_e$. We have

$$I(X^n; Y_n|Y^{n-1}) = H(Y_n|Y^{n-1}) - H(Y_n|X^n; Y^{n-1}). \tag{13}$$

Let us first consider $H(Y_n|Y^{n-1})$. For any $y^n \in \mathcal{Y}^n$,

$$\begin{aligned}&P_{Y_n|Y^{n-1}}(y_n|y^{n-1}) \\ &= \sum_{s_n} \left(\sum_{x_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) P_{X_n|Y^{n-1}}(x_n|y^{n-1}) \right) \\ &\quad \cdot P_{S_n|Y^{n-1}}(s_n|y^{n-1}).\end{aligned} \tag{14}$$

Now

$$\begin{aligned}&\sum_{x_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) P_{X_n|Y^{n-1}}(x_n|y^{n-1}) \\ &= \sum_{e_n} \left(\sum_{x_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) P_{X_n|E_n}(x_n|e_n) \right) \\ &\quad \cdot P_{E_n|Y^{n-1}}(e_n|y^{n-1}).\end{aligned} \tag{15}$$

But $P_{X_n|E_n}$ is uniform over $\mathcal{X}_{E_n}$, so

$$\sum_{x_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) P_{X_n|E_n}(x_n|e_n) = \frac{1}{|\mathcal{Y}|} \tag{16}$$

since the channel is uniformly-symmetric and variable-power. Therefore

$$P_{Y_n|Y^{n-1}}(y_n|y^{n-1}) = \frac{1}{|\mathcal{Y}|} \tag{17}$$

for any $y^n \in \mathcal{Y}^n$, so

$$\begin{aligned}&H(Y_n|Y^{n-1}) \\ &= -\sum_{y^{n-1}} P_{Y^{n-1}}(y^{n-1}) \\ &\qquad \cdot \sum_{y_n} P_{Y_n|Y^{n-1}}(y_n|y^{n-1}) \log P_{Y_n|Y^{n-1}}(y_n|y^{n-1}) \\ &= -\sum_{y^{n-1}} P_{Y^{n-1}}(y^{n-1}) \log |\mathcal{Y}| = \log |\mathcal{Y}|.\end{aligned} \tag{18}$$

Hence we see that any $P \in \mathbf{P}_e$ maximizes $H(Y_n|Y^{n-1})$.

Let us turn our attention to $H(Y_n|X^n, Y^{n-1})$. For all $x^n \in \mathcal{X}^n$ and $y^{n-1} \in \mathcal{Y}^{n-1}$,

$$\begin{aligned}&H(Y_n|X^n = x^n, Y^{n-1} = y^{n-1}) \\ &= -\sum_{y_n} \left(\sum_{s_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) \right. \\ &\qquad \left. \cdot P_{S_n|X^{n-1},Y^{n-1}}(s_n|x^{n-1}, y^{n-1}) \right) \\ &\quad \cdot \log \left(\sum_{s_n} P_{Y_n|X_n,S_n}(y_n|x_n, s_n) \right. \\ &\qquad \left. \cdot P_{S_n|X^{n-1},Y^{n-1}}(s_n|x^{n-1}, y^{n-1}) \right).\end{aligned} \tag{19}$$

But because the channel is uniformly-symmetric and variable-power, (19) is equal for all $x_n \in \mathcal{X}_l$ for any $l = 1, \ldots, L$. Moreover, S_n is independent of (X^{n-1}, Y^{n-1}) given (E^{n-1}, Z^{n-1}), so $H(Y_n|X^n, Y^{n-1})$ is a function only of the distribution of (E^n, Z^{n-1}). Now, Z^{n-1} is independent of X^n given E^n, hence we conclude that the only aspect of the input distribution that is important to $H(Y_n|X^n, Y^{n-1})$ is the distribution of the input energies.

We have shown that

$$\begin{aligned}&\sup_{P\in\mathbf{P}_e} \lim_{n\to\infty} \frac{1}{n} I(X^n; Y_n|Y^{n-1}) \\ &\quad = \log |\mathcal{Y}| - \sup_{P\in\mathbf{P}_e} \lim_{n\to\infty} \frac{1}{n} H(Y_n|X^n, Y^{n-1}) \\ &\quad = \log |\mathcal{Y}| - \sup_{P\in\mathbf{P}} \lim_{n\to\infty} \frac{1}{n} H(Y_n|X^n, Y^{n-1}) \\ &\quad \geq \sup_{P\in\mathbf{P}} \lim_{n\to\infty} \frac{1}{n} I(X^n; Y_n|Y^{n-1}).\end{aligned} \tag{20}$$

The reverse inequality is straightforward since $\mathbf{P}_e \subset \mathbf{P}$. Therefore if $P^* \in \mathbf{P}_e$ is such that $\{(X_n, Y_n)\}$ is ergodic, then it is clear that capacity is given by (12). ∎

Remark. Our result implies that feedback does not improve the capacity of uniformly-symmetric variable-noise FSMCs [4, Section V] since no choice of input energy can be made.

V. Dynamic programming formulation

In this section, we outline a dynamic programming approach for finding near-optimal power control policies for uniformly-symmetric variable-noise variable-power FSMCs.

Subject to a constraint on the average energy, we wish to minimize

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} H(Y_n|X^n, Y^{n-1}) &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_n|X^n, Y^{n-1}) \\ &= \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{n=1}^{N} h_n(Y_n|X^n, Y^{n-1})\right] \\ &= \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{n=1}^{N} \mathbb{E}\left[h_n(Y_n|X^n, Y^{n-1})\right| X^n = x^n, U^{n-1} = u^{n-1}\right]\right],\end{aligned} \tag{21}$$

where

$$h_n(Y_n|X^n, Y^{n-1}) = -P_{Y_n|X^n,Y^{n-1}}(Y_n|X^n, Y^{n-1}) \cdot \log P_{Y_n|X^n,Y^{n-1}}(Y_n|X^n, Y^{n-1}). \tag{22}$$

We see that we almost have a problem that can be cast as an average cost per stage problem. Unfortunately, the state space of the problem, $\mathcal{X}^{n-1} \times \mathcal{U}^{n-1}$, increases without bound. Hence we make the simplifying assumption that

$$h_n(Y_n|X^n, Y^{n-1}) \simeq h_n(Y_n|X^n_{n-k}, Y^{n-1}_{n-k}) \tag{23}$$

for some sufficiently large k. For the assumption to be reasonable, we would at the very least need to assume that the order of the state process is less than k.

So we wish to minimize

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{n=1}^{N} \mathbb{E}\left[h_n(Y_n|X^n_{n-k}, Y^{n-1}_{n-k})\right| X^n_{n-k} = x^n_{n-k}, U^{n-1}_{n-k} = u^{n-1}_{n-k}\right]\right], \tag{24}$$

which we can pose as an average cost per stage problem with state $(x^{n-1}_{n-k}, u^{n-1}_{n-k})$, control $P_{E_n|X^{n-1}_{n-k},U^{n-1}_{n-k}}$, and random disturbance (X_n, Y_n). We impose an average power constraint of ε by restricting the controls to those that satisfy $1/(k+1)\mathbb{E}[\sum_{i=n-k}^{n} E_i] \le \varepsilon$. An optimal policy and average cost can be found by methods such as value iteration (see, for example, [22, Chapter 4]). Suppose $P^*_{E_n|X^{n-1}_{n-k},U^{n-1}_{n-k}}$ is the optimal stationary policy thus obtained. It remains only to check whether or not $\{(X_n, Y_n)\}$ is ergodic under $P^*_{E_n|X^{n-1}_{n-k},U^{n-1}_{n-k}}$.

Given $P^*_{E_n|X^{n-1}_{n-k},U^{n-1}_{n-k}}$, we can derive the unique stationary distribution $P_{X_n|X^{n-1}_{n-k},Y^{n-1}_{n-k}}$. Then, for all $(x^n_{n-k}, s^n_{n-k}, y^n_{n-k})$, we have

$$\begin{aligned}&P_{X_n,S_n,Y_n|X^{n-1}_{n-k},S^{n-1}_{n-k},Y^{n-1}_{n-k}}(x_n, s_n, y_n|x^{n-1}_{n-k}, s^{n-1}_{n-k}, y^{n-1}_{n-k}) \\ &= P_{Y_n|X_n,S_n}(y_n|x_n, s_n) P_{X_n|X^{n-1}_{n-k},Y^{n-1}_{n-k}}(x_n|x^{n-1}_{n-k}, y^{n-1}_{n-k}) \\ &\quad \cdot P_{S_n|S^{n-1}_{n-k}}(s_n|s^{n-1}_{n-k}).\end{aligned} \tag{25}$$

This equation defines transition probabilities for a Markov chain with states $\sigma_1, \sigma_2, \ldots$, where $\sigma_n := (x^{n+k-1}_n, s^{n+k-1}_n, y^{n+k-1}_n)$. Therefore, ergodicity of the triple $\{(X_n, S_n, Y_n)\}$ can be studied in terms of the ergodicity of the Markov chain $\{\sigma_n\}$. Indeed, for any irreducible, aperiodic set of the Markov chain, there exists a stationary distribution and the state at a particular time is asymptotically independent of the state at another time. By starting the chain with the stationary distribution, we get ergodic behavior of the triple $\{(X_n, S_n, Y_n)\}$ and hence of the pair $\{(X_n, Y_n)\}$.

The above method does not appear to be very practical, however: Even for moderate k, the state space is already quite sizeable; and we would not expect small k to yield good solutions. The method may nevertheless inspire alternative approximate methods that are more practical.

VI. Conclusion

The main result that we have shown is that if noiseless feedback of an arbitrary rate is available, then capacity is achieved for a particular class of channels by adopting an optimal policy for controlling the power based on the feedback coupled with appropriate channel coding. This class of channels includes channels that model slow, flat fading channels with variable input power. Thus our result gives some guidance on the design of effective transmission schemes for slow, flat fading channels with feedback. But it still by no means clear how such a scheme might be designed. A logical starting point would be to incorporate power control into coding schemes for FSMCs, for example, [23], [24], [25].

Acknowledgments

The authors would like to thank Ibrahim Abou-Faycal for helpful discussions and suggestions.

References

[1] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept. 1999.

[2] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[3] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, pp. 761–771, Mar. 1999.

[4] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 868–886, May 1996.

[5] T. Holliday, P. Glynn, and A. Goldsmith, "On entropy and Lyapunov exponents for finite-state channels." Submitted to *IEEE Trans. Inform. Theory*.

[6] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.

[7] S. Tatikonda and S. Mitter, "The capacity of channels with feedback — Part I: The general case." Submitted to *IEEE Trans. Inform. Theory*, Oct. 2001.

[8] S. C. Tatikonda, *Control Under Communication Constraints*. PhD thesis, Massachusetts Institute of Technology, Sept. 2000.

[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[10] G. Kramer, *Directed Information for Channels with Feedback*, vol. 11 of *ETH Series in Information Processing*. Konstanz: Hartung-Gorre Verlag, 1998.

[11] J. L. Massey, "Causality, feedback, and directed information," in *Proc. 1990 International Symposium on Information Theory and its Applications (ISITA 1990)*, pp. 303–305, Nov. 1990.

[12] A. R. Barron, "The strong ergodic theorem for densities: Generalized Shannon-McMillan-Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, Nov. 1985.

[13] R. L. Adler, "Ergodic and mixing properties of infinite memory channels," *Proc. Am. Math. Soc.*, vol. 12, pp. 924–930, Dec. 1961.

[14] R. M. Gray, M. O. Dunham, and R. L. Gobbi, "Ergodicity of Markov channels," *IEEE Trans. Inform. Theory*, vol. 33, pp. 656–664, Sept. 1987.

[15] J. C. Kieffer, "A general formula for the capacity of stationary nonanticipatory channels," *Inf. Control*, vol. 26, pp. 381–391, 1974.

[16] S. Yang, A. Kavčić, and S. Tatikonda, "Delayed feedback capacity of finite-state machine channels: Upper bounds on the feedforward capacity," in *Proc. 2003 IEEE International Symposium on Information Theory (ISIT 2003)*, 2003.

[17] S. Yang and A. Kavčić, "Markov sources achieve the feedback capacity of finite-state machine channels," in *Proc. 2002 IEEE International Symposium on Information Theory (ISIT 2002)*, p. 361, 2002.

[18] C. C. Tan and N. C. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, pp. 2032–2040, Dec. 2000.

[19] F. Babich and G. Lombardi, "A Markov model for the mobile propagation channel," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 63–73, Jan. 2000.

[20] W. Turin and R. van Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1809–1817, Dec. 1998.

[21] H. S. Wang and N. Moayeri, "Finite-state Markov channel — a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 163–171, Feb. 1995.

[22] D. P. Berksekas, *Dynamic Programming and Optimal Control*, vol. 2. Belmont, MA: Athena Scientific, second ed., 2001.

[23] L. Li and A. J. Goldsmith, "Low-complexity maximum-likelihood detection of coded signals sent over finite-state Markov channels," *IEEE Trans. Commun.*, vol. 50, pp. 524–530, Apr. 2002.

[24] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1706–1716, Sept. 2001.

[25] J. Garcia-Frias and J. D. Villasenor, "Turbo decoders for Markov channels," *IEEE Commun. Lett.*, vol. 2, pp. 257–259, Sept. 1998.