Proceeding of the 2004 American Control Conference
Boston, Massachusetts June 30 - July 2, 2004

Capacity-achieving Feedback Scheme for Flat Fading Channels with Channel State Information

Jialing Liu *
liujl@iastate.edu

Nicola Elia *
nelia@iastate.edu

Sekhar Tatikonda †
sekhar.tatikonda@yale.edu

* Dept. of Electrical and Computer Engineering, Iowa State University
† Dept. of Electrical Engineering, Yale University

***Abstract*— In this paper we propose a capacity-achieving feedback scheme for an AWGN i.i.d. flat fading channel with channel state information. This scheme considerably reduces the encoding and decoding complexity and delay, and it extends the Schalkwijk and Kailath scheme to i.i.d. time-varying channels.**

I. Introduction

In this paper, we consider the problem of finding a capacity-achieving feedback scheme for an additive white Gaussian noise (AWGN) i.i.d. flat fading channel, with channel state information (CSI) available instantaneously at the receiver and with one delay at the transmitter. This channel with various CSI assumptions has been studied in [6], [2], [16], etc. [6] obtains the capacity of a fading channel with instantaneous CSI at both the transmitter and receiver (TRCSI), or at the receiver only (RCSI). [2] investigates capacity of channels with memory and with imprecise or delayed CSI. [16] provides Markov channel capacity with instantaneous receiver-side CSI and delayed transmitter-side CSI (DTRCSI). See also [10], [17].

In this paper, we present a capacity-achieving feedback scheme for an AWGN i.i.d. flat fading channel (AIFFC) under DTRCSI assumption. Although feedback cannot improve capacity, we show that it leads to simpler encoders and decoders, less coding delay, and doubly exponential decay of the probability of error.

The feedback communication scheme in this paper is an extension of that in [3] to i.i.d. time-varying channels, and hence it is a further extension of the Schalkwijk and Kailath scheme [13], [12]. Our scheme reveals the equivalence between feedback stabilization over an AIFFC and communication with access to noiseless feedback over the same channel, and thus this work also fits into the framework of investigating the interaction between information and control; see [11], [5], [15], [4], [8] and references therein.

II. Preliminaries

In this section we present briefly the channel model, channel side information and capacities, and some relevant properties of random processes.

Notations: We use upper case letters to designate random variables and lower case letters to indicate their realizations. In case where no confusion may arise, we may use lower case letters for random variables. We represent time indices by subscripts. We denote by A_n^m the sequence of random variables $\{A_n, A_{n+1}, \cdots, A_m\}$, and a_n^m the sequence of realizations $\{a_n, a_{n+1}, \cdots, a_m\}$. Note that A_n^m is a random sequence, and $(A_n)^m$ is the m-th power of the random variable A_n. We denote by $\{A_k\}$, $\{a_k\}$ the generic sequences. We use $a[1], a[2], \cdots$ for a collection of fixed numbers.

This research was supported by the NSF under Grant ECS-0093950. The first author wishes to thank Zhengdao Wang for useful discussions.

A. Channel model and channel state information

Figure 1 (a) illustrates the channel model that we consider in this paper, which we refer to as AWGN i.i.d. flat fading channel (AIFFC). At time k, this discrete-time memoryless channel $\mathcal{F}$ is described as

$$\mathcal{F}: \quad Y_k = \xi_k X_k + N_k, \tag{1}$$

where X_k is the channel input, ξ_k is the channel state, N_k is the channel noise, and Y_k is the channel output. Here all variables are assumed to be real. The channel state $\{\xi_k\}$ forms an i.i.d. random process given by

$$\Pr(\xi_k = \xi[i]) = \alpha[i], \text{ for } k = 0, 1, \cdots, \tag{2}$$

where for $i = 1, 2, \cdots, m$, $\alpha[i], \xi[i]$ are fixed numbers, $\alpha[i] > 0$, and $\sum_{i=1}^m \alpha[i] = 1$. We assume noise $\{N_k\}$ is an AWGN with unit variance.

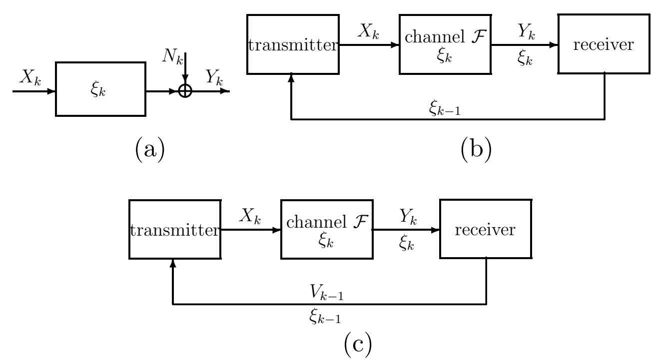

Fig. 1. (a) An AIFFC $\mathcal{F}$. (b) Channel with DTRCSI. The receiver has CSI instantaneously, and transmitter has CSI with one delay through feedback from the receiver. (c) Channel with feedback. The transmitter has CSI and information about the received signals with one delay through feedback from the receiver.

We call ξ_k the *channel state* at time k. Knowing the channel state information (CSI) at the transmitter side and/or the receiver side significantly affects the transmission strategy and the capacity [1]. In this paper we focus on the case of delayed transmitter-side and instantaneous receiver-side CSI (DTRCSI) [1], [2], since we are interested in channels with noiseless feedback, which has the ability to feed back CSI at the receiver to the transmitter with one delay. Hence, the channel state ξ_k is available for the receiver at time k and available for the transmitter at time $k+1$, see Figure 1 (b).

In addition, in this paper we allow the transmitter to have access to the noiseless feedback from the receiver. We assume that, during the transmission at time k, the transmitter has information about both ξ_{k-1} and V_{k-1}, where V_{k-1} depends on the received signals Y_0^{k-1}. See Figure 1 (c). This feedback communication scheme is consistent with that in [16], which investigates a finite-state Markov channel (FSMC) whose receiver feeds back both the precise CSI and the channel output to the transmitter with delay.

B. Channel capacity under different CSI assumptions

In this subsection, we briefly present some capacity formulas of fading channels under different CSI assumptions, see [6], [1], [2], [16] for detailed results.

First, for an FSMC with AWGN, if CSI is available to both the transmitter and receiver before the symbol transmission, i.e., if under TRCSI assumption, the capacity subject to the average input power constraint

$$\mathbf{E}X^2 \leq \mathcal{P} \tag{3}$$

is given by

$$C_{TR} = \frac{1}{2}\max_{\gamma} \mathbf{E}_\xi \ln(1+\xi^2\gamma(\xi)), \tag{4}$$

where $\gamma(\cdot)$ satisfying

$$\mathbf{E}_\xi \gamma(\xi) \leq \mathcal{P} \tag{5}$$

is the power allocation function that maps the channel state $\xi[i]$ to the channel input power $\gamma(\xi[i])$ for each i. The optimal power allocation is obtained analytically and has a "water-filling in time" interpretation.

Second, for an FSMC with AWGN, if under DTRCSI assumption, the capacity subject to (3) is given by

$$C_{DTR} = \frac{1}{2}\max_{\gamma} \mathbf{E}\ln(1+(\xi_{k+1})^2\gamma(\xi_k)), \tag{6}$$

where expectation is w.r.t. the joint distribution of ξ_k and ξ_{k+1}, and $\gamma(\cdot)$ is the power allocation function that maps the channel state ξ to the channel input power $\gamma(\xi)$ to be used at the *next* instant of time and such that (5) holds.

Third, for an FSMC with AWGN, if CSI is available only to the receiver before the symbol transmission, i.e., if under RCSI assumption, the capacity subject to (3) is given by

$$C_R = \frac{1}{2}\mathbf{E}_\xi \ln(1+\xi^2\mathcal{P}). \tag{7}$$

Note that the capacities given in (4), (6), and (7) satisfy

$$C_R \leq C_{DTR} \leq C_{TR} \tag{8}$$

since the transmitter may preform better power adaptation when it can access more information about the channel state.

In the case of AIFFC, i.e., the case where the channel state forms an i.i.d. process, the capacities under TRCSI and RCSI assumptions remain unchanged. However, the capacity under DTRCSI assumption becomes smaller and in fact equal to C_R.

Proposition 1 (Capacity for DTRCSI). *Given $\mathcal{P}>0$, the feedforward capacity of an AIFFC with DTRCSI subject to the average input power constraint (3) is given by*

$$C_{DTR} = \frac{1}{2}\sum_{i=1}^{m}\alpha[i]\ln(1+\xi[i]^2\mathcal{P}) = \frac{1}{2}\mathbf{E}_\xi \ln(1+\xi^2\mathcal{P}). \tag{9}$$

Proof: In i.i.d. case, the capacity formula (6) reduces to

$$\begin{aligned} C_{DTR} &= \frac{1}{2}\max_{\gamma}\mathbf{E}_{\xi_{k+1}}\mathbf{E}_{\xi_k}\ln(1+(\xi_{k+1})^2\gamma(\xi_k)) \\ &\leq \frac{1}{2}\max_{\gamma}\mathbf{E}_{\xi_{k+1}}\ln(1+(\xi_{k+1})^2\mathbf{E}_{\xi_k}\gamma(\xi_k)) \\ &\leq \frac{1}{2}\mathbf{E}_\xi\ln(1+\xi^2\mathcal{P}) = C_R. \end{aligned}$$

The first inequality is according to the concavity of capacity functions and Jensen's inequality. By (8), we obtain $C_{DTR}=C_R$ for i.i.d. fading. ∎

Remark 1 (Constant power allocation strategy). Note that for i.i.d. fading channels with DTRCSI, capacity is achievable without power adaptation, as is the case without transmitter-side CSI. This is because power adaptation can improve capacity only when the transmitter has prior (deterministic or stochastic) knowledge of the channel state before the symbol transmission, e.g., channels with TRCSI, channels with memory; see also [2].

The capacity given by (9) is also known as the *ergodic capacity*, since it is achieved when the channel becomes ergodic [6], [1], [2]. We can intuitively interpret it as follows. When ξ_k takes value $\xi[i]$, the channel behaves like an AWGN channel with capacity $C[i] = \ln(1+\xi[i]^2\mathcal{P})/2$. Suppose among K channel uses, there are $K[i]$ times in which ξ takes value $\xi[i]$, hence the channel can transmit approximately $\sum K[i]C[i]/K$ nats per channel use. By the Strong Law of Large Numbers [14], the empirical frequency $K[i]/K$ converges to its expectation $\alpha[i]$ almost surely. Then (9) follows.

Now we address the problem of feedback capacity. For an AIFFC with DTRCSI, the capacities for the channels in Figure 1 (b) and 1 (c) are equal; see [16] for a rigorous proof. However, the transmitter-side information about the channel state and channel output can considerably simplify the encoding and decoding and lead to better performance in terms of probability of error, as we will show in this paper (see also [13], [12]).

C. Some useful properties about random processes

Let W_0^{K-1}, $K \leq +\infty$ be an i.i.d. discrete real random process with $\Pr(W=w[i])=\alpha[i]$, where $\sum_{i=1}^m \alpha[i]=1$ and $\alpha[i]>0$ for $i=1,2,\cdots,m$. For K trials, $K<\infty$, denote $\beta_K^{[i]}$ the empirical frequency, i.e., the fraction of trials in which $w[i]$ is obtained. From the Strong Law of Large Numbers, we have for any $\epsilon>0$, it holds

$$\Pr(\lim_{K\to\infty}|\beta_K^{[i]}-\alpha[i]|<\epsilon)=1. \tag{10}$$

In other words, the empirical frequency $\beta_K^{[i]}$ converges almost surely to $\alpha[i]$, the expected value of $\beta_K^{[i]}$. Thus we define the *typical set* and *typical sequence* in the following way (see also [14]).

Definition 1 (Typical set and typical sequence). *For any $\epsilon > 0$, define the* typical set $E_K^{(\epsilon)}$ *w.r.t. the sequence W_0^{K-1} to be the set of all realizations w_0^{K-1} satisfying*

$$|\frac{\sum_{k=0}^{K-1} \mathbf{1}_{w[i]}(w_k)}{K} - \alpha[i]| \leq \epsilon, \;\; i = 1, 2, \cdots, m. \tag{11}$$

Each sequence in the typical set is a typical sequence.

Here $\mathbf{1}_a(\cdot)$ is the indicator function. From (10), it follows

$$\Pr(\lim_{K\to\infty} E_K^{(\epsilon)}) = 1, \tag{12}$$

namely, for large enough K, almost every sequence is typical, and it contains $w[i]$ roughly $K\alpha[i]$ times.

When considering stochastic dynamical systems, we often need to deal with a sequence of cumulative products U_1^K of the random sequence W_0^{K-1}, where $U_k := \prod_{i=0}^{k-1} W_i$, $K \leq \infty$. In this paper we are interested in the case where the sequence is almost surely convergent to zero only. Suppose

$$\bar{w} := \prod_{i=1}^{m} w[i]^{\alpha[i]} \tag{13}$$

is such that $|\bar{w}| < 1$. Then from (10), we have $\bar{w}^k \to 0$ and $U_k \overset{\text{a.s.}}{\to} 0$ as $k \to \infty$. Furthermore,

$$\frac{U_k}{\bar{w}^k} \overset{\text{a.s.}}{\to} 1 \;\text{ as } k \to \infty. \tag{14}$$

This property allows us to define the *typical set* and *typical sequence* for the cumulative products as follows.

Definition 2 (Typical set and typical sequence for cumulative products). *For any $\epsilon > 0$, define the* typical set $E_K^{(\epsilon)}$ *w.r.t. the cumulative product sequence U_1^K to be the set of all realizations u_1^K satisfying*

$$|\frac{u_k}{\bar{w}^k} - 1| \leq \epsilon \quad \text{for all } k = 1, 2, \cdots, K. \tag{15}$$

Each sequence in the typical set is a typical sequence.

Thus by definition it holds

$$\Pr(\lim_{K\to\infty} E_K^{(\epsilon)}) = 1, \tag{16}$$

in other words,

$$\Pr(\lim_{K\to\infty} \{u_K | (1-\epsilon)\bar{w}^K \leq u_K \leq (1+\epsilon)\bar{w}^K\}) = 1. \tag{17}$$

Therefore, the sequence $\{U_K\}$ can be approximated by sequence $\{\bar{w}^K\}$ with arbitrarily small error except for a set of sequences with zero probability measure, provided $|\bar{w}| < 1$.

III. Capacity-achieving Feedback Scheme for AIFFC with DTRCSI

In this section, we present a feedback scheme that achieves the capacity of the AIFFC with DTRCSI. We first describe the proposed scheme, and then prove that it achieves the capacity and the doubly exponential decay of error probability.

A. The setup and strategy

Communication setup

The communication setup to be used for information transmission analysis is shown in Figure 2 (a), where x_0 is the message to be encoded at the transmitter and recovered at the receiver, and $-\hat{x}_{0,k}$ is the decoded message, i.e., the estimate of x_0 at time k. From this figure we can identify the transmitter, the channel $\mathcal{F}$, and the receiver. Parameters A and B are scalars depending on the channel state ξ_k. Note that the channel state ξ_k is available and used at the receiver side at time k, and used at the transmitter side at time $k+1$, which is consistent with the assumption of DTRCSI. Note also that the communication setup allows us to adapt the transmission strategy to the channel variation.

We describe the operation of the communication setup as follows. At time 0, the channel input is $x_0 = \tilde{x}_0$ (noticing $\hat{x}_0 = 0$). Then the receiver receives the channel output

$$y_0 = \xi_0 x_0 + N_0 \tag{18}$$

as well as the channel state ξ_0, and it sends ξ_0 to the transmitter. After this the transmitter and receiver do the following updates

$$\begin{aligned} \tilde{x}_1 &= A(\xi_0)\tilde{x}_0 \\ \hat{x}_1 &= B(\xi_0)y_0 \\ \hat{x}_{0,0} &= \hat{x}_0 = 0. \end{aligned} \tag{19}$$

Then the system reaches the next discrete instant of time, i.e., time 2, and the channel input becomes

$$x_1 = \tilde{x}_1 + \hat{x}_1, \tag{20}$$

and the channel output becomes

$$y_1 = \xi_1 x_1 + N_1. \tag{21}$$

Like before, the receiver sends ξ_1 to the transmitter, and then the transmitter and receiver do the updates

$$\begin{aligned} \tilde{x}_2 &= A(\xi_1)\tilde{x}_1 \\ \hat{x}_2 &= A(\xi_1)\hat{x}_1 + B(\xi_1)y_1 \\ \hat{x}_{0,1} &= \tfrac{1}{A(\xi_0)}\hat{x}_1. \end{aligned} \tag{22}$$

Then the system reaches time 2 and repeat the above procedures. In general, at time k, the system performs

$$\begin{aligned} x_k &= \tilde{x}_k + \hat{x}_k \\ y_k &= \xi_k x_k + N_k \\ \tilde{x}_{k+1} &= A(\xi_k)\tilde{x}_k \\ \hat{x}_{k+1} &= A(\xi_k)\hat{x}_k + B(\xi_k)y_k \\ \hat{x}_{0,k} &= \tfrac{1}{\prod_{j=0}^{k-1} A(\xi_j)}\hat{x}_k. \end{aligned} \tag{23}$$

These recursions will generate a sequence of estimates $\{-\hat{x}_{0,k}\}$ that approaches x_0, as we will show later.

Without affecting the transmitted message, estimates of the message, and the channel inputs and outputs, we can make the power of the feedback signal bounded by using the time-varying scaling factor α in Figure 1 of [3], which implies that this scheme is an extension of the Schalkwijk and Kailath scheme, and furthermore, it leads to simpler encoders and decoders and less coding delay.

Control setup

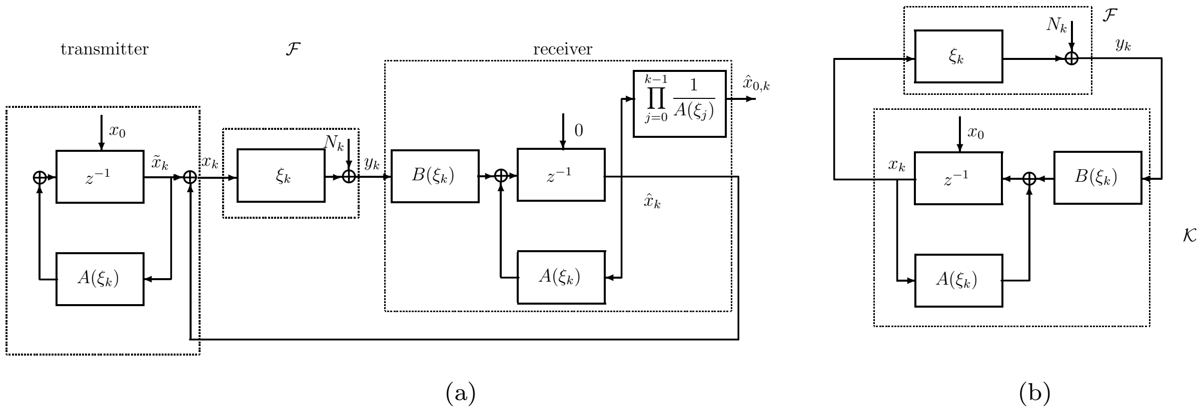

Fig. 2. (a) The communication setup. (b) The control setup.

The recursion of x is

$$\begin{array}{lll} x_{k+1} & = A(\xi_k)x_k + B(\xi_k)y_k & \text{initial state: } x_0 \\ y_k & = \xi_k x_k + N_k, & \end{array} \quad (24)$$

see Figure 2 (b) for the corresponding block diagram. Note that (24) is indeed a simple control system, which we refer to as the control setup, as opposed to the communication setup shown in Figure 2 (a). The zero-input response and zero-state responses of (24) correspond to $\tilde{x}_k$ and $\hat{x}_k$ respectively, and the channel input power is bounded only if (24) is stable.

Choice of parameters

Given any $\mathcal{P} > 0$, we choose the parameters in the communication setup as

$$\begin{array}{ll} a[i] := A(\xi[i]) : & = \sqrt{\mathcal{P}\xi[i]^2 + 1} \\ b[i] := B(\xi[i]) : & = -\frac{\mathcal{P}\xi[i]}{\sqrt{\mathcal{P}\xi[i]^2+1}}. \end{array} \quad (25)$$

We will show the communication setup along with this choice of parameters is indeed capacity-achieving in the next subsection.

B. Main result

Our main result states:

Theorem 1 (Capacity-achieving scheme). *Suppose $\mathcal{F}$ is any given finite-state AIFFC, where the channel state ξ_k has an i.i.d. distribution given by (2) and is available instantaneously to the receiver and with one delay to the transmitter. Given any $\mathcal{P} > 0$, the setup described in Section III-A, with the parameters given by (25), achieves the feedback capacity*

$$C = \frac{1}{2}\sum_{i=1}^{m} \alpha[i] \ln(1 + \xi[i]^2\mathcal{P}) \quad (26)$$

subject to the average input power constraint

$$\mathbf{E}X^2 \le \mathcal{P}. \quad (27)$$

Moreover, the probability of error decays to zero doubly exponentially.

Remark 2 (Significance of "almost sure" notion). Theorem 1 relies on the Strong Law of Large Numbers, and capacity (26) is achieved when the channel states produce typical sequences. When the channel states produce nontypical sequences, any non-zero capacity cannot be guaranteed. Therefore, we focus our attention on typical sequences only, and state our results in an almost sure fashion.

To prove this theorem, we first show that the scheme described in the previous subsection can transmit reliably at rate

$$\sum_{i=1}^{m} \alpha[i] \ln(|A(\xi[i])|) = \frac{1}{2}\mathbf{E}_\xi \ln(1 + \xi^2\mathcal{P}). \quad (28)$$

Then we show the power constraint (27) is satisfied.

Achievable rate

In this part we present a proposition about the achievable rate of the communication setup with any given parameters $a[i] := A(\xi[i])$ and $b[i] := B(\xi[i])$. It essentially says that, if the choice of parameters is such that in the control setup of Figure 2 (b), the plant $\mathcal{K}$ is unstable but the closed-loop with $\mathcal{F}$ is stabilized, then the communication setup reliably transmits x_0 at an average rate determined by the average growing rate of the unstable plant. We thus establish the equivalence between communication and control.

Suppose $a[i]$ is any real number with $|a[i]| \ge 1$ and

$$\prod_{i=1}^{m} |a[i]| > 1. \quad (29)$$

Define

$$\bar{a} := \prod_{i=1}^{m} a[i]^{\alpha[i]}. \quad (30)$$

Then it holds $|\bar{a}| > 1$. Let

$$A_{cl}(\xi_k) := A(\xi_k) + B(\xi_k)\xi_k, \quad (31)$$

i.e., $A_{cl}(\xi_k)$ is the closed-loop matrix at time k. We have the following proposition:

Proposition 2 (Achievable rate). *If the interconnection between $\mathcal{F}$ and $\mathcal{K}$ of Figure 2 (b) is almost surely stable, then the associated communication setup of Figure 2 (a) reliably (in the sense that the probability of error converges to zero almost surely) transmits the initial state, x_0, from the transmitter to the receiver at rate*

$$R = (1 - \epsilon)\ln(|\bar{a}|), \quad (32)$$

for any given $\epsilon > 0$. Moreover, the probability of error decays to zero doubly exponentially.

Proof: To prove the achievable rate, we first show the estimate $\hat{x}_{0,k}$ has mean $-x_0$ and a variance decaying to zero, and then we choose the data rate accordingly to ensure that the error probability goes to zero.

Clearly we have

$$\hat{x}_{0,k} = (\prod_{j=0}^{k-1} A(\xi_j))^{-1}\hat{x}_k = (\prod_{j=0}^{k-1} A(\xi_j))^{-1}x_k - x_0.$$

Because of (29), the sequence $\{(\prod_{j=0}^{k-1} A(\xi_j))^{-1}\}$ almost surely converges to zero. By (14), we know for any given $\epsilon_1 > 0$, for large k it holds that $(\prod_{j=0}^{k-1} A(\xi_j))^{-1}$ lies within $(1-\epsilon_1)\bar{a}^{-k}$ and $(1+\epsilon_1)\bar{a}^{-k}$ almost surely. So we have for k large enough $(\prod_{j=0}^{k-1} A(\xi_j))^{-1} \overset{\text{a.s.}}{=} \bar{a}^{-k}$, and hence

$$\hat{x}_{0,k} \overset{\text{a.s.}}{=} \bar{a}^{-k}x_k - x_0. \tag{33}$$

As the closed-loop is almost surely stable and the noise is zero mean, x_k tends to zero and $\mathbf{E}(x_k)^2$ converges to a steady-state value σ^2 almost surely as k tends to infinity. Therefore $\hat{x}_{0,k}$ almost surely has a Gaussian distribution $\mathcal{N}(-x_0, \bar{a}^{-2k}\sigma^2)$ for large enough k. (The reader can verify that the approximations in the mean and variance do not affect our analysis.)

Recall that x_0 is the message to be recovered at the receiver. We now show that we can increase the number of messages to be sent exponentially according to the decay rate of the variance of the estimate, and the probability of error still decays to zero.

Consider k channel uses. Denote by σ_j the square root of the j-th variance of the estimate, namely $\sigma_j := |\bar{a}|^{-j}\sigma$. Now we equally partition the interval $\mathcal{I} = (-\sqrt{\mathcal{P}}, \sqrt{\mathcal{P}})$ into $M_{k-1} = (\sigma_{k-1})^{-(1-\epsilon)}$ segments, and let the center of each segment represents a message to be transmitted. Suppose at time $k-1$ at the receiver side, the estimate $\hat{x}_{0,k-1}$ is in the i-th segment. Then the i-th message, the center of that segment, is decided. Then the probability of error is

$$PE_{k-1} = 2Q(\sqrt{\mathcal{P}}(\sigma_{k-1})^{-\epsilon}), \tag{34}$$

which goes to zero as $k \to \infty$ if $\epsilon > 0$. Here $Q(\cdot)$ is the complementary cdf of normalized Gaussian distribution (Q-function) [7].

Noticing that the number of possible messages is M_{k-1}, we have the transmission rate to be

$$\begin{aligned} R &= \lim_{k\to\infty} \frac{\ln M_{k-1}}{k} = \lim_{k\to\infty} \frac{-(1-\epsilon)}{k} \ln \sigma_{k-1} \\ &= \lim_{k\to\infty} \frac{-(1-\epsilon)}{k}(-k\ln|\bar{a}| + \ln\sigma) = (1-\epsilon)\ln(|\bar{a}|). \end{aligned}$$

Hence the channel can transmit reliably at rate $R = (1-\epsilon)\ln(|\bar{a}|)$ for any $\epsilon > 0$, in the sense that the probability of error converges to zero, if the sequence $\{\xi_k\}$ is a typical sequence. Since the typical sequences occur almost surely, we conclude that the system transmits reliably (in the sense that the probability of error converges to zero almost surely) the initial state x_0 at rate R.

Now we show the error probability decays doubly exponentially. Using the Chernoff bound [7], we have

$$\begin{aligned} PE_{k-1} &\leq \sqrt{\frac{2}{\pi\mathcal{P}}}(\sigma_{k-1})^{\epsilon}\exp(-\frac{1}{2}(\sigma_{k-1})^{-2\epsilon}\mathcal{P}) \\ &= \sqrt{\frac{2}{\pi\mathcal{P}}}\frac{\sigma^{\epsilon}}{|\bar{a}|^{(k-1)\epsilon}}\exp(-\frac{1}{2}\sigma^{-2\epsilon}|\bar{a}|^{2(k-1)\epsilon}\mathcal{P}), \end{aligned}$$

namely, the error probability decays doubly exponentially. ∎

We apply the proposition to the specially chosen parameters in (25), and we get

Corollary 1. *Let $a[i]$ and $b[i]$ be as in (25). The interconnection between $\mathcal{F}$ and $\mathcal{K}$ of Figure 2 (b) is almost surely stable, and the associated communication setup of Figure 2 (a) reliably (in the sense that the probability of error converges to zero almost surely) transmits the initial state, x_0, from the transmitter to the receiver at rate*

$$R = \frac{1-\epsilon}{2}\mathbf{E}_{\xi}\ln(1+\xi^2\mathcal{P}), \tag{35}$$

for any given $\epsilon > 0$.

Proof: It suffices to show the control setup of Figure 2 (b) is almost surely stable. Consider the dynamics of the control setup:

$$\begin{aligned} x_{k+1} &= (A(\xi_k) + B(\xi_k)\xi_k)x_k + B(\xi_k)N_k \\ &= \tfrac{1}{A(\xi_k)}x_k + B(\xi_k)N_k. \end{aligned} \tag{36}$$

To establish the almost sure stability of (36), we only need to show the unforced dynamics

$$x_{k+1} = \frac{1}{A(\xi_k)}x_k \tag{37}$$

is almost surely stable; namely, we need to show (37) leads to $x_k \overset{\text{a.s.}}{\to} 0$ as $k \to 0$, which is clearly true by $|\bar{a}| > 1$ and (14). Thus (36) is almost surely stable. ∎

Power computation

Now we compute the channel input power for the specially chosen parameters given in (25).

Proposition 3 (Power computation). *Let $a[i]$ and $b[i]$ be as in (25). The channel input power is given by*

$$\mathbf{E}x^2 = \frac{a[i]^2 - 1}{\xi[i]^2} = \mathcal{P} \quad \text{for } i = 1, 2, \cdots, m, \tag{38}$$

and hence it satisfies the average input power constraint (27).

Proof: Note first that $\mathcal{P} = (a[i]^2 - 1)/\xi[i]^2$ for $i = 1, 2, \cdots, m$ by (25). To show (38), it is sufficient to show that, $\mathbf{E}_{x,N}(x_k)^2$ exponentially converges to $\mathcal{P}$ almost surely, since it then implies the time average of $\mathbf{E}_{x,N}(x_k)^2$ is $\mathcal{P}$. Note the expectation in $\mathbf{E}_{x,N}(x_k)^2$ is w.r.t. the system state x and channel noise N but not the channel state ξ.

Now we study the recursion for $\mathbf{E}_{x,N}(x_k)^2$. By

$$x_{k+1} = a[i]^{-1}x_k + b[i]N_k \quad \text{with Pr } \alpha[i], \tag{39}$$

we obtain

$$\mathbf{E}_{x,N}(x_{k+1})^2 = a[i]^{-2}\mathbf{E}_{x,N}(x_k)^2 + b[i]^2 \quad \text{with Pr } \alpha[i].$$

Using $b[i] = -\mathcal{P}\xi[i]/a[i]$, we get

$$\mathbf{E}_{x,N}(x_{k+1})^2 = a[i]^{-2}(\mathbf{E}_{x,N}(x_k)^2 + \mathcal{P}^2\xi[i]^2) \quad \text{with Pr } \alpha[i],$$

which yields

$$\mathbf{E}_{x,N}(x_{k+1})^2 - \mathcal{P} = a[i]^{-2}(\mathbf{E}_{x,N}(x_k)^2 - \mathcal{P}) \quad \text{with Pr } \alpha_i.$$

Hence, as k goes to infinity, $\mathbf{E}_{x,N}(x_k)^2 - \mathcal{P}$ exponentially converges to zero almost surely, i.e., $\mathbf{E}_{x,N}(x_k)^2$ exponentially converges to $\mathcal{P}$ almost surely. Thus the result follows. ∎

An example

An analog AWGN erasure channel is a discrete-time memoryless channel with an AWGN followed by an erasure, in other words, the receiver obtains a noisy version of the transmitted symbol with probability $1 - e$ or nothing (i.e., the symbol being lost) with probability e at each time step. This channel with delayed noiseless feedback satisfies the DTRCSI assumption since the receiver and transmitter can determine the ideal CSI from the channel output: they determine $S_k = 0$ if and only if they receive $y_k = 0$.

This channel has the same capacity as the channel with an erasure followed by an AWGN, and our result applies to this channel. It yields that, when the channel is open, i.e., $S = 0$, we use $a(0) = 1$ and $b(0) = 0$; when the channel is closed, i.e., $S = 1$, we use $a(1) = \sqrt{\mathcal{P}+1}$ and $b(1) = \frac{1}{a(1)} - a(1)$. This strategy has a clear practical meaning: when the receiver finds a signal is lost due to $S = 0$, it notifies the transmitter, then the transmitter uses $a(0) = 1$, and hence, the lost signal is to be transmitted and received again in the next step; otherwise the transmitter will generate another signal and send it through the channel. In this process, the transmitter and receiver adapt their parameters according to the channel variation, and therefore the uncertainty interval (see [11]) about the transmitted symbol shrinks if $S_k = 1$, remains if $S_k = 0$, and converges to zero eventually. We can show if the adaptive transmission is not used, i.e., if constant a and b are used instead all the time, the closed-loop system cannot transmit at the capacity rate. Thus, this idea of adaptive transmission helps us to achieve the feedback capacity, and it coincides with the ideas in [11], [15].

IV. Conclusions and Future Work

In this paper, we showed that an extension of the feedback scheme in [12], [3] achieves the capacity of an AWGN i.i.d. flat fading channel with precise CSI available to the receiver immediately and to the transmitter with delay. The error probability decreases to zero doubly exponentially. We established the equivalence between feedback stabilization over the channel and communication with access to noiseless feedback over the same channel. In our scheme, we designed the transmitter and receiver to adapt their strategies according to the status of the channel, and we explored the notion of almost sure convergence that is intrinsic to the Shannon capacity of a time-varying channel.

Up to this point, many important problems remain unsolved. First, how to construct feedback schemes to achieve capacities without CSI is unclear. Second, how to construct capacity-achieving feedback schemes for more complicated fading channels, for example, frequency-selective fading channels, remains to be seen; in fact, we are not even clear about the feedback capacity of some fading channels. These are subjects to future work. We report the study of a capacity-achieving feedback scheme for a Markov channel in [9].

V. References

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. Inform. Theory*, 44(6):2619–2691, Oct. 1998.

[2] G. Caire and S. Shamai (Shitz). On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6):2007–2019, Sept. 1999.

[3] N. Elia. Control-oriented feedback communication schemes. *Proc. 42nd IEEE Conf. Decision and Control*, pages 3161–3166, Dec. 2003. See also: When Bode meets Shannon: Control-oriented feedback communication schemes, accepted by *IEEE Trans. Autom. Contr.*

[4] N. Elia. Feedback stabilization in the presence of fading channels. *Proc. of the 2003 American Control Conference*, pages 4438–4443, June 2003.

[5] N. Elia and S.K. Mitter. Stabilization of linear systems with limited information. *IEEE Trans. Automat. Contr.*, 46(9):1384–1400, 2001.

[6] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, Nov. 1997.

[7] S. M. Kay. *Fundamentals of Statistical Signal Processing II: Detection Theory*. Prentice-Hall PTR, Englewood Cliffs, N.J., 1998.

[8] J. Liu and N. Elia. Quantized feedback stabilization of non-linear affine systems. *International Journal of Control*, 77(3):239–249, Feb. 2004.

[9] J. Liu, N. Elia, and S. Tatikonda. Capacity-achieving feedback scheme for Markov channels with channel state information. *submitted to IEEE Trans. Inform. Theory*, Feb. 2004.

[10] R. Negi and J. M. Cioffi. Delay-constrained capacity with causal feedback. *IEEE Trans. Inform. Theory*, 48(9):2478–2494, Sept. 2002.

[11] A. Sahai. *Anytime Information Theory*. PhD thesis, MIT, Cambridge, MA, 2001.

[12] J.P.M. Schalkwijk. A coding scheme for additive noise channels with feedback Part II:Bandlimited signals,. *IEEE Trans. Inform. Theory*, IT-112(2):183–189, Apr. 1966.

[13] J.P.M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback I: No bandwidth constraint. *IEEE Trans. Inform. Theory*, IT-12:172–182, Apr. 1966.

[14] A. N. Shiryaev. *Probability*. Springer-Verlag, New York, 2nd edition, 1996.

[15] S. Tatikonda and S. Mitter. Control under communication constraints: Part I and Part II. *submitted to IEEE Trans. Autom. Contr.*, March 2002.

[16] H. Viswanathan. Capacity of markov channels with receiver csi and delayed feedback. *IEEE Trans. Inform. Theory*, 45(2):761–771, March 1999.

[17] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of finite-state machine channels. *submitted to IEEE Trans. Inform. Theory*, Dec. 2002.