Capacity of Nearly Decomposable Markovian Fading Channels Under Asymmetric Receiver–Sender Side Information

Muriel Médard, *Senior Member, IEEE*, and R. Srikant, *Fellow, IEEE*

***Abstract*—We investigate the following issue: if fast fades are Markovian and known at the receiver, while the transmitter has only a coarse quantization of the fading process, what capacity penalty comes from having the transmitter act on the current coarse quantization alone? For time-varying channels which experience rapid time variations, sender and receiver typically have asymmetric channel side information. To avoid the expense of providing, through feedback, detailed channel side information to the sender, the receiver offers the sender only a coarse, generally time-averaged, representation of the state of the channel, which we term slow variations. Thus, the receiver tracks the fast variations of the channel (and the slow ones perforce) while the sender receives feedback only about the slow variations. While the fast variations (micro-states) remain Markovian, the slow variations (macro-states) are not. We compute an approximate channel capacity in the following sense: each rate smaller than the "approximate" capacity, computed using results by Caire and Shamai, can be achieved for sufficiently large separation between the time scales for the slow and fast fades. The difference between the true capacity and the approximate capacity is $O(\epsilon \log^2(\epsilon) \log(-\log(\epsilon)))$, where ϵ is the ratio between the speed of variation of the channel in the macro- and micro-states. The approximate capacity is computed by power allocation between the slowly varying states using appropriate water filling.**

***Index Terms*—Capacity, channel side information, Markovian models.**

I. Introduction

In order to achieve tractable results, channel models considered for capacity computation tend to be very simple. Yet there are few results concerning to what extent such simplification affects the validity or channel capacity computations and of the policies for achieving capacity. Our purpose in this paper is to investigate, for one example, the validity of a simplified channel model and of capacity-achieving policies based upon such a simplified model. We are interested in considering the effect of imperfect information decoupling when the sender channel side information (SCSI) is a coarse representation of the receiver channel side information (RCSI). We explore how the interplay between, on the one hand, the asymmetry in SCSI and RCSI and, on the other hand, their relationship, affects capacity.

The long-term state of the channel would generally be related to the time average, over some recent past, of the observed signal-to-noise ratios (SNR's). The receiver and sender do not know in general the cause of fading, to determine whether shadowing effects (usually the cause of slow variations) or multipath effects (usually the cause of fast variations) are at work. The sender and receiver will in general only have information concerning the channel state seen at the receiver. The slow variations may be interpreted as the long-term state of SNR behavior, while the fast variations are the short-term SNR behavior. A natural framework is one in which instantaneous good and bad SNR's are possible for different slow-fading states, but that transitions among slow-fading states tend to occur when the recent history of the channel follows certain patterns. For instance, a recent history of poor SNR's would generally indicate that we are transitioning to a slow fade.

In fading channels, channel side information (CSI) at the sender and the receiver can be obtained from many sources, for instance, sounding signals embedded in the communications signal, out-of-band pilot tones, or measurements from users in other frequency bands or time slots. Asymmetry in directions of channels in different bands or directions render perfect knowledge of CSI at the transmitter difficult to obtain without very detailed feedback from the receiver. In many circumstances, RCSI and SCSI are asymmetric, although related. In particular, when the channel is rapidly varying, providing full feedback from the receiver to the sender regarding the CSI may be onerous and inefficient. Consider the case of communications at typical carrier frequencies for commercial wireless systems, say in the 2-GHz range. For a user traveling at 60 mi/h, the Doppler spread is of the order of 0.5 kHz. If each update takes roughly 12 bits, then a 6-kHz channel would be required for feedback—a bandwidth which may be possibly more advantageously used for data transmission. For our motivating example, a more attractive option is to have the sender have partial SCSI, while the receiver may have full RCSI.

In this paper, we consider a discrete-time finite-state Markov channel (FSMC). The RCSI, which we term the micro-states, is a full description of the FSMC. The SCSI, which we term the macro-states, is a coarser representation of the RCSI states. The

Manuscript received November 20, 2001; revised January 18, 2006. This work was supported by the National Science Foundation under CAREER Grant CCR-009 3349. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000 and the IEEE International Symposium on Information Theory and Its Applications, Honolulu, HI, November 2000.

M. Médard is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: medard@mit.edu).

R. Srikant is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: rsrikant@uiuc.edu).

Communicated by P. Narayan, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2006.872843

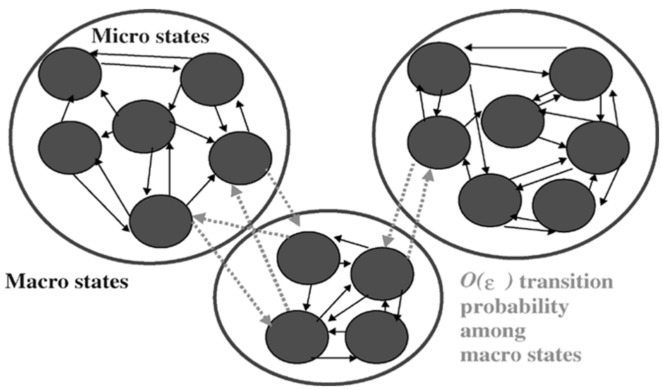

Fig. 1. Representation of our FSMC and its description of macro-states. The large circles are the macro-states and the solid circles are the micro-states. The solid lines represent transitions that occur with $O(1)$ probability and the dotted lines represent transitions that occur with $O(\epsilon)$ probability.

sender only knows that the current micro-state is within a certain macro-state. The macro-states represent the slow variations of the channel. Fig. 1 illustrates the micro-states and the macro-states we consider. The detailed states are the RCSI—the receiver knows the detailed micro-states, for instance, through the use of sounding signals embedded in the transmission. The SCSI information may be obtained from feedback from the receiver and is a coarse quantization of the RCSI. Of course, higher and lower values of instantaneous SNR may occur within a window of certain average or quantized values. Note that fades are possible while we are in a good macro-state and, conversely, SNR surges are possible while we are in a weak macro-state. Thus, fast variations which are atypical of the current macro-state do not necessarily indicate that we have left the macro-state. Our SCSI is a deterministic function of the RCSI. In our case, if the detailed states form a FSMC, then the macro states, which are a partition of the set of micro-states, are not Markov.

Our analysis relies on the fact that the variations of the macro-state are, on average, much slower than those of the micro-state. We thus consider a perturbation type of approach to our FSMC. Returning to our motivating example, let us consider that the macro-state changes, on average, every 2 s. Then, if we denote by ϵ the ratio between the speed of change of the macro-state and that of the micro-state, $\epsilon \approx 10^{-3}$. We make use of the fact that ϵ is very small to get a handle on the capacity of our channel. Perturbation approaches for modeling of Markov channels have been considered in [10]. Approximating channels using finite-state channels has been done in [9], using results from [8]. However, those results do not consider CSI and thus cannot be directly extended to our problem, although the insight from those results is useful.

The capacity of time-varying channels with perfect or imperfect SCSI at the sender and possibly imperfect RCSI has been considered in a very general framework in [2], which also provides a thorough overview of relevant work in the area of Markov channels. The results of [2] apply to the case we consider, and establish a relevant coding theorem. Our work is complementary to the results in [2]. While those results provide the relevant coding theorems, several important questions remain when we consider a specific channel model.

- What is the stochastic characterization of the behavior of a specific channel?
- How may we compute or approximate capacity, and how does capacity depend on relevant channel parameters?
- What are simple schemes, for instance power control policies, that will allow us to approach capacity?

This paper seeks to answer these questions for the type of channel we have discussed above and for which we provide a model in the next section. The first question is addressed in Section III. Sections II and V address the last two questions jointly, establishing a power control policy that approaches capacity. The main channel parameter we consider is ϵ.

Recent work in this area has established simple capacity results for certain types of SCSI and RCSI. In particular, the capacity is known for Markov channels when the SCSI is a delayed version of the RCSI [12]. The capacity for several general types of single-user channels where the SCSI is a deterministic function of the RCSI is given by [1], which also presents a comprehensive overview of capacity results for channels with CSI. In particular, [1] gives an exact, intuitive, and simple capacity result for the case when the RCSI is perfect; the SCSI state is a deterministic function of the RCSI state; and the probability mass function of the channel states conditioned on all the past SCSI is equal to that conditioned on the current SCSI state only. It is this last condition, which entails that the SCSI process is also Markov, that renders the results of [1] not applicable to our channel model, since the history of all past macro-states gives information about channel state beyond that contained in the present SCSI alone.

Although the model of [1] does not apply directly, it is natural to conjecture that, as ϵ decreases, a simplified model consistent with the results of [1] should yield an increasingly good approximation to the true capacity. Our results support this intuition and quantify the effect of the spread between the speed of the slow variations (SCSI) and that of the fast variations (RCSI). Our results also show, however, that the loss in accuracy due to assuming that the SCSI and the RCSI are both Markov decreases sublinearly in ϵ. Using the capacity results for Markovian SCSI, established by Caire and Shamai in [1], our work seeks to establish the sensitivity of capacity to particular model differences (Markovian SCSI versus near-Markovian).

In the next section, we present our model for channel variations and the main theorem of our paper, relating to the behavior of the channel given by our model to capacity. In Section III, we analyze the behavior of the channel by creating an order-ϵ perturbation model of the original channel. For this modified channel, which obeys the conditions of [1], a capacity can be found, as shown in Section IV. We call this capacity the approximate capacity of the channel. The difference between the true capacity and the approximate capacity is shown to be $O(\epsilon \log^2(\epsilon) \log(-\log(\epsilon)))$. In Section V, we discuss our results and present our conclusions.

II. Model and Coding Theorem

In this paper, which requires extensive notation, we have attempted to balance, on the one hand, precision and consistency, which may lead to a surfeit of notation, with, on the other hand, elegance and simplicity. As a result, we reserve certain indices for certain variables, except when we deem ambiguity to be unlikely, and its risk to be outweighed by the fatigue cumbersome notation causes in the reader. We indicate matrices and vectors through underlining. The lower and upper limits of the range of a vector are given by subscripts and superscripts, respectively, except when the lower limit is 1, in which case it is omitted. When confusion is unlikely, we omit the range of a vector altogether. For instance, $\underline{\pi}^{(m)}$, $\underline{\tilde{\pi}}^{(m)}$, and $\underline{q}^{(m)}$ are all probability vectors associated with the micro-states of the macro-state S_m. We omit their size, which corresponds to the number, N_m, of micro-states in S_m. Whenever practical, we reserve the indices m and i for the macro-state and micro-state being considered, respectively. Thus, we generally consider, in macro-state S_m, the ith micro-state, which we denote by $\mu_i^{(m)}$. The variables j, k, l, q, r, s we generally reserve for indices in a vector or matrix, although they may sometimes be used to index states.

We consider a bandlimited system and assume that bandwidth broadening due to time variations is negligible. The Nyquist input and output sampling rates are thus taken to be equal. After sampling, we use a discrete-time model to describe our channel, $\mathcal{C}$, and its input. The channel is described by an FSMC, whose states we refer to as micro-states. Let $T(n)$ denote the random variable corresponding to the micro-state at time unit n. To each micro-state corresponds a single gain value. The mapping from a state to its gain value we define to be the function γ. Thus, $\gamma(T(n))$ is the random variable corresponding to the signal attenuation at time n. The sampled received signal at time n is given by the random variable

$$Y(n) = \gamma(T(n))X(n) + W(n)$$

where $X(n)$ is the transmitted signal and $W(n)$ is the result of sampling bandlimited additive white Gaussian noise (AWGN). Thus, the $W(n)$'s are independent and identically distributed (i.i.d.), with mean 0 and variance σ^2. The sender has an average power constraint $\mathcal{P}$, i.e., codewords of length k are constrained by

$$\frac{\sum_{n=1}^{k} X^2(n)}{k} \le \mathcal{P}.$$

The sample values of $T(n), Y(n)$, and $X(n)$ are denoted by the same letters using lower case.

Let $S(n)$ be the random variable which identifies the macro-state that the Markov chain is in at the nth time instant. Thus, at time instant n, the SCSI is $\underline{S}^n$ and the RCSI is $\underline{T}^n$. Moreover, $S(n)$ is a deterministic function of $T(n)$. At time instant n, the micro-state is $\mu_i^{(m)}$ and the macro-state is S_m iff $T(n) = \mu_i^{(m)}$, which implies $S(n) = S_m$. The $X(n)$'s are chosen using the SCSI, which is causal. Hence,

$$\mathrm{Prob}(\underline{x}^n \,|\, \underline{s}^k) = \mathrm{Prob}(\underline{x}^n \,|\, \underline{s}^n), \quad \text{for all } n \le k.$$

The micro-states form a discrete-time Markovian fading process defined by the stochastic matrix $\underline{A} + \epsilon\underline{B}$, where $\underline{A}$ is block-diagonal with M blocks and the mth block (which is also a stochastic matrix) is denoted by $\underline{A}^{(m)}$. Each block $\underline{A}^{(m)}$ is a stochastic matrix. The (j,k) entry of $\underline{A}^{(m)}$ is $A^{(m)}(j,k)$, which denotes the probability of transition from micro-state $\mu_j^{(m)}$ to micro-state $\mu_k^{(m)}$ for the set of probabilities corresponding to $\underline{A}^{(m)}$.

Let $\underline{\pi}^{(m)}$ be the stationary probability vector associated with $\underline{A}^{(m)}$, i.e.,

$$\underline{\pi}^{(m)}\underline{A}^{(m)} = \underline{\pi}^{(m)}. \tag{1}$$

Define an $M \times M$ matrix $\underline{P}$ as follows: the (j,k) entry of $\underline{P}$ is given by

$$P(j,k) = \epsilon \sum_{\mu_l^{(j)} \in S_j} \sum_{\mu_q^{(k)} \in S_k} \pi^{(j)}(l) B\left(\mu_l^{(j)}, \mu_q^{(k)}\right), \quad j \neq k \tag{2}$$

and $P(j,j) = 1 - \sum_{k \neq j} P(j,k)$. Note that $\underline{P}$ is also a stochastic matrix and let $\underline{p}$ be its stationary probability vector, i.e., $\underline{p} = \underline{p}\underline{P}$. We can interpret the entries of $\underline{P}$ as being the long-term transition probabilities among macro-states and $p(m)$ as approximating the steady-state probability of being in S_m. Indeed, as we show in the next section, $p_\epsilon(m) = p(m) + O(\epsilon)$, where $p_\epsilon(m)$ is the actual steady-state probability of being in macro-state S_m. The channel model we consider for the state behavior is shown in Fig. 1.

Our main result is the following.

Theorem 1: Define

$$C(\epsilon) := \max_{\{\mathcal{P}(m)\}} \frac{1}{2} \sum_{m=1}^{M} p_\epsilon(m) \sum_{\mu_i^{(m)} \in S_m} \log\left(1 + \frac{\mathcal{P}(m)\gamma^2\left(\mu_i^{(m)}\right)}{\sigma^2}\right) \pi^{(m)}(i) \tag{3}$$

subject to

$$\sum_{m=1}^{M} p_\epsilon(m)\mathcal{P}(m) \le \mathcal{P}$$

where $\mathcal{P}$ is the power constraint at the sender. Also, define

$$C_{\mathrm{true}}(\epsilon) = \lim_{n \to \infty} \frac{1}{n} \max_{P(X(t)\,|\,S^t, X^{t-1}), 1 \le t \le n} I(X^n; Y^n \,|\, T^n).$$

Then, there exists ϵ^* such that for $\forall \epsilon \in (0, \epsilon^*)$

$$C_{\mathrm{true}}(\epsilon) = C(\epsilon) - O(\epsilon \log^2(\epsilon) \log(-\log(\epsilon))). \qquad \diamond$$

Before proving this theorem in the next sections, we give an interpretation of it. $C_{\mathrm{true}}(\epsilon)$ is the known characterization of the capacity, with an explicit parametrization in terms of ϵ. The theorem states that, for sufficiently large time-scale separation in the variations associated with the micro- and macro-states, the capacity $C_{\mathrm{true}}(\epsilon)$ can be approximated by $C(\epsilon)$ to within

$O(\epsilon \log^2(\epsilon)\log(-\log(\epsilon)))$. Thus, $C(\epsilon)$ is an approximate expression for capacity which exploits that two time-scale property of the Markov channel.

III. Channel Behavior in Terms of ϵ

In this section, we show that, if $\mathcal{C}$ stays in a macro-state S_m for at least τ time units, where $\tau = O(-\log(\epsilon))$, then, for any time instant $k > \tau$, the probability that $\mathcal{C}$ is in micro-state $\mu_i^{(m)} \in S_m$ conditioned on still being in S_m at time k, is $O(\epsilon)$ close to $\pi^{(m)}(i)$. Recall that $\underline{\pi}^{(m)}$ is the stationary probability vector associated with block $\underline{A}^{(m)}$. The motivation for showing this is as follows. Even if the sender had access to additional sender side information at the time of each macro-state transition, the results of this section will show that, after τ units, the sender's estimate of the channel state is nearly independent of this additional side information. We will make use of this fact to create an upper bound to the channel capacity in the next section. Throughout our discussion, we consider an arbitrary past $\underline{T}_{-L'}^0$, and future $\underline{T}_L^s$.

Define

$$\alpha_i^{(m)}(n) = \text{Prob}\left(T(n) = \mu_i^{(m)} \,\middle|\, S(k) = S_m \quad \forall k \in \{1,\ldots,n\}, \underline{T}_{-L'}^0 = \underline{t}_{-L'}^0\right).$$

For simplicity, we omit explicit indexing on $\underline{t}_{-L'}^0$ for α. We seek to show that, if n is sufficiently large, then $\alpha_i^{(m)}(n)$ is approximately equal to $\pi^{(m)}(i)$.

To this end, define a new transition matrix as follows:

$$\tilde{P}^{(m)}(j,k) = A^{(m)}(j,k) + \epsilon B\left(\mu_j^{(m)}, \mu_k^{(m)}\right), \quad \text{if } j \neq k \tag{4}$$

and

$$\tilde{P}^{(m)}(j,j) = A^{(m)}(j,j) + \epsilon B\left(\mu_j^{(m)}, \mu_j^{(m)}\right) + \epsilon \sum_{\mu_l^{(r)} \notin S_m} B\left(\mu_j^{(m)}, \mu_l^{(r)}\right). \tag{5}$$

Note that, for small ϵ, $\underline{\tilde{P}}^{(m)}$ can be interpreted as the transition matrix of a new Markov chain defined over the same set of states associated with the micro-states in S_m. We have in effect created a transition matrix for micro-states within macro-state S_m such that the weight of transitions out of macro-state S_m are reassigned to micro-state self-transitions. Let the vector $\underline{\tilde{q}}_n^{(m)}$ be defined as follows:

$$\underline{\tilde{q}}_n^{(m)} = \underline{q}_0^{(m)}\left(\underline{\tilde{P}}^{(m)}\right)^n \tag{6}$$

where $\underline{q}_0^{(m)}$ is the vector whose elements are the values of the initial probability mass function over the states of the Markov chain,

$$\begin{aligned}\alpha_i^{(m)}(0) &= \text{Prob}\left(T(0) = \mu_i^{(m)} \,\middle|\, S(0) \in S_m, \underline{T}_{-L'}^0 = \underline{t}_{-L'}^0\right)\\ &= q_0^{(m)}(i).\end{aligned}$$

To avoid cumbersome notation, we omit explicit indexing on $\underline{T}_{-L'}^0$ for the $q^{(m)}$'s. It follows from our definition of $\underline{q}_0^{(m)}$ that

$$\sum_{j \in S_m} q_0^{(m)}(j) = 1.$$

Lemma 1: For all $n > 0$

$$\left|\alpha_i^{(m)}(n) - \tilde{q}_n^{(m)}(i)\right| = O(\epsilon). \tag{7}$$

Proof: See the Appendix. ◊

It now remains for us to show that $\alpha_i^{(m)}(n)$ is also an $O(\epsilon)$ approximation of $\pi^{(m)}(i)$ for n large enough. This is the crux of the following, central lemma.

Lemma 2: Let $\lambda_{2,m}$ be the second largest eigenvalue of $\underline{\tilde{P}}^{(m)}$ and assume that there is only one eigenvalue with this magnitude. Then, there exists ϵ^* and K_m such that $\forall \epsilon \in (0,\epsilon^*)$ and for all

$$n > \frac{-\log(\epsilon) + \log K_m}{-\log(|\lambda_{2,m}|)}$$

$|\pi^{(m)}(i) - \alpha_i^{(m)}(n)| < \epsilon$, for all $i \in S_m$.

Proof: See the Appendix. ◊

We now show that a similar result holds when we consider both past and future macro-states. The probability of being in a particular micro-state, conditioned on being in the same macro-state for many more time samples in the past and future is virtually the same as the approximate steady-state probability obtained when the macro-state is an absorbing state. The following theorem states this fact in more detail. Note that we again omit explicit indexing on $\underline{t}_{-L'}^0, \underline{t}_L^s$. Note also that, in the following, we consider

$$L \geq \frac{-\log(\epsilon) + \log(K_m)}{-\log(|\lambda_{2,m}(\epsilon)|)}.$$

Theorem 2: Let $\mu_i^{(m)} \in S_m$ and define

$$\beta_i^{(m)}(n) = \text{Prob}\left(T(n) = \mu_i^{(m)} \,\middle|\, \underline{T}_1^{L-1} \in S_m, \underline{T}_{-L'}^0 = \underline{t}_{-L'}^0, \underline{T}_L^s = \underline{t}_L^s\right)$$

for all $n < L$. Then, there exists ϵ^* and K_m such that $\forall \epsilon \in (0,\epsilon^*)$ and all

$$\frac{-\log(\epsilon) + \log(K_m)}{-\log(|\lambda_{2,m}(\epsilon)|)} < n < L - \frac{-\log(\epsilon) + \log(K_m)}{-\log(|\lambda_{2,m}(\epsilon)|)} \tag{8}$$

$|\pi^{(m)}(i) - \beta_i^{(m)}(n)| = O(-\epsilon\log(\epsilon))$, for all $\mu_i^{(m)} \in S_m$.

Proof: See the Appendix 1. ◊

IV. Channel Capacity Using a Reduced-Order MODEL at the Transmitter

The difficulty in establishing the capacity of the channel lies in the fact that the macro-state transitions gives us partial information about the current micro-state. This information is difficult to quantify. However, as we have shown in the previous section, the longer we stay in a macro-state, the less relevant this information becomes over most of our stay in a macro-state.

In order to prove our main theorem, we construct channels whose capacity upper- and lower-bounds the capacity of our channel. We provide a lower bound by choosing a specific input distribution, and then show that the difference between the upper bound and the lower bound is $O(\epsilon \log^2(\epsilon) \log(-\log(\epsilon)))$.

A lower bound on capacity is given by

$$C = \max_{\{P(m)\}} \frac{1}{2} \sum_{m=1}^{M} p_\epsilon(m) \sum_{\mu_i^{(m)} \in S_m} \log\left(1 + \frac{P(m)\gamma^2\left(\mu_i^{(m)}\right)}{\sigma^2}\right) \pi^{(m)}(i) \quad (9)$$

subject to $\sum_{m=1}^{M} p_\epsilon(m)P(m) \le \mathcal{P}$, where $\pi^{(m)}(i)$ is the probability of being in micro-state $\mu_i^{(m)}$ and $p_\epsilon(m)$ is, as defined earlier, the probability of being in macro-state S_m. The lower bound holds by virtue of the fact that selecting a particular distribution for the input cannot yield a better result than maximization over all allowable input distributions. Specifically, we choose the input symbols to be independent zero-mean Gaussian random variables with variance $P(m)$ in macro-state S_m.

We now establish an upper bound to our channel capacity. Let $\mu_{\max}$ be the channel micro-state such that $|\gamma(\mu_{\max})|$ is the largest among all micro-states in $\bigcup_{m=1,\ldots,M} S_m$. The upper bound to the capacity of our channel $\mathcal{C}$ is obtained by constructing a new channel $\mathcal{C}_U$ as follows. Whenever there is a transition from one macro-state to another in $\mathcal{C}$, the channel $\mathcal{C}_U$ remains in the state $\mu_{\max}$ for time τ until no macro-state transitions have occurred for τ time units in $\mathcal{C}$. After that time, $\mathcal{C}_U$ reverts to the micro-state that $\mathcal{C}$ is in. Thus, under our definition, if a macro-state change occurs $\frac{\tau}{2}$ time units after the last macro-state change and the next macro-state change occurs more than τ time units after, then $\mathcal{C}_U$ will remain in state $\mu_{\max}$ for $\frac{3\tau}{2}$ time units. At the end of that time, $\mathcal{C}_U$ reverts to behaving like $\mathcal{C}$, except that the SCSI of $\mathcal{C}_U$ includes the micro-state at the time of the last macro-state change. Fig. 2 illustrates a sample behavior for $\mathcal{C}$ and the corresponding behavior for $\mathcal{C}_U$. The capacity of $\mathcal{C}_U$ is higher than that of $\mathcal{C}$. During the intervals shown as shaded areas in Fig. 2, the channel $\mathcal{C}_U$ has perfect RCSI and SCSI and, moreover, its SNR is the best possible SNR, which is given by the gain $\gamma(\mu_{\max})$ of the best micro-state $\mu_{\max}$ over all sets S_i. Outside those shaded areas, $\mathcal{C}_U$ behaves as $\mathcal{C}$. We call an interval where $\mathcal{C}_U$ behaves as $\mathcal{C}$ an active interval (shown in white in Fig. 2). We term L_k the beginning of the kth active interval and Λ_k its duration.

Since we do not know how to maximize the mutual information in the capacity expression for $\mathcal{C}_U$, we now seek to obtain an upper bound to C_U, the capacity of $\mathcal{C}_U$. In order to upper-bound C_U, we construct two channels, $\mathcal{C}_U^1$ and $\mathcal{C}_U^2$, with average power constraints $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. We shall later relate these channels, as well as $\mathcal{P}_1$ and $\mathcal{P}_2$, to our original channel, $\mathcal{C}$, and to

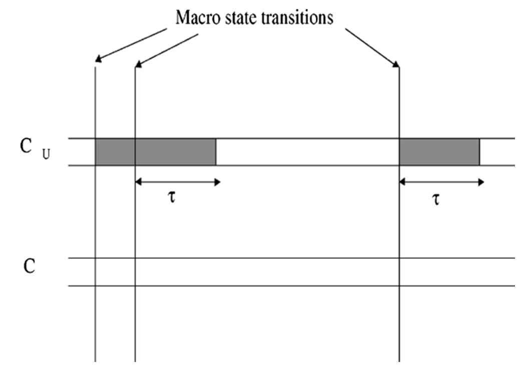

Fig. 2. Construction of $\mathcal{C}_U$ from $\mathcal{C}$.

$\mathcal{P}$. Channel $\mathcal{C}_U^1$ is created as follows. For every transition from one macro-state to another in $\mathcal{C}$, the channel $\mathcal{C}_U^1$ produces no output for time τ until no macro-state transitions have occurred for τ time in $\mathcal{C}$. At these times, we define the channel $\mathcal{C}_U^1$ to be in an inactive state. At all other times, which we term active intervals as for $\mathcal{C}_U^1, \mathcal{C}_U^1$ follows the behavior of $\mathcal{C}$.

Channel $\mathcal{C}_U^2$ is constructed as follows. Whenever channel $\mathcal{C}$ experiences a macro-state transition, $\mathcal{C}_U^2$ behaves as a simple bandlimited AWGN channel whose gain is given by $\mu_{\max}$. We term this behavior the active state of $\mathcal{C}_U^2$. At all other times, $\mathcal{C}_U^2$ provides no output. We term this behavior the inactive state of $\mathcal{C}_U^2$. Note that, because we may have macro-state transitions within a τ interval, the times when $\mathcal{C}_U^2$ may be active and the times when $\mathcal{C}_U^1$ may be active may overlap. Moreover, the total duration of the active intervals of $\mathcal{C}_U^1$ and $\mathcal{C}_U^2$ we consider may be greater than the time over which we observe $\mathcal{C}_U$.

Lemma 3: The capacity C_U^2 of $\mathcal{C}_U^2$ is

$$O\left(\frac{-\epsilon\log(\epsilon)}{2}\log\left(1+\frac{\gamma^2\left(\mu_{\max}\right)\mathcal{P}_2}{-\epsilon\log(\epsilon)\sigma^2}\right)\right) = O(\epsilon\log^2(\epsilon)\log(-\log(\epsilon))).$$

Proof: Let $\mathcal{M}(k)$ be the indicator function of macro-state transitions, i.e., it is 1 if there was a transition at time instant k and 0 otherwise. We define

$$\mathcal{S}(n) = \sum_{k=1}^{n} \mathcal{M}(k) \quad (10)$$

the number of transitions by time n. Thus, by time n, the number of time samples spent transmitting in $\mathcal{C}_U^2$ is $\tau\mathcal{S}(n)$. Over time n, the energy per transmitted symbol when we transmit over $\mathcal{C}_U^2$ is $\frac{\mathcal{P}_2 n}{\tau\mathcal{S}(n)}$. The maximum average mutual information per time unit between the input and the output of $\mathcal{C}_U^2$ over n time samples is given by

$$\boldsymbol{\mathcal{I}}(n) = \frac{\tau\mathcal{S}(n)}{2n}\log\left(1+\gamma^2\left(\mu_{\max}\right)\frac{\mathcal{P}_2 n}{\sigma^2\tau\mathcal{S}(n)}\right). \quad (11)$$

It now remains to be shown that $E[\boldsymbol{\mathcal{I}}(n)]$ reaches a limit as $n \to \infty$. First, given our Markov model, the $\mathcal{M}(k)$'s decorrelate and we may therefore apply the Law of Large Numbers for decorrelating random variables to obtain that $\forall \delta > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\mathcal{S}(n)}{n} - \xi\right| > \delta\right) = 0 \quad (12)$$

where $\xi = \lim_{n\to\infty} E[\frac{\mathcal{S}(n)}{n}]$.

Let us define the function v as

$$v(x) = \frac{\tau x}{2} \log\left(1 + \frac{\gamma^2(\mu_{\max})\mathcal{P}_2}{\sigma^2 \tau x}\right). \quad (13)$$

The function v is continuous, concave, bijective, and increasing over the positive reals $\mathbb{R}e^{*+}$. Note that $v(\frac{\mathcal{S}(n)}{n}) = \boldsymbol{\mathcal{I}}(n)$. Thus, from (12), $\forall \delta > 0$, $\forall \epsilon_0 > 0$, $\exists n_0 \in \mathbb{N}^+$ such that $\forall n > n_0$, where $\mathbb{N}^+$ is the set of positive integers.

$$P(|\boldsymbol{\mathcal{I}}(n) - v(\xi)| > \delta) < \epsilon_0. \quad (14)$$

Expression (14), implies that $\exists \delta_0 > 0, \exists \epsilon_0'$ such that $\forall \delta \in (0, \delta_0)$, $\forall \epsilon_0 \in (0, \epsilon_0')$, $\exists n_0 \in \mathbb{N}^+$ such that $\forall n > n_0$

$$E[\boldsymbol{\mathcal{I}}(n)] \geq (v(\xi) - \delta)(1 - \epsilon_0). \quad (15)$$

Moreover, from Jensen's inequality

$$\lim_{n\to\infty} E[\boldsymbol{\mathcal{I}}(n)] \leq v(\xi). \quad (16)$$

Thus, $\lim_{n\to\infty} E[\boldsymbol{\mathcal{I}}(n)] = v(\xi)$.

Finally, we may select τ to be $\frac{-\log(\epsilon)+\log(\kappa)}{-\log(\epsilon^*)} + 1$ where $\kappa = \max_m\{K_m\}$ (see Theorem 2). Then $\tau = O(-\log(\epsilon))$. Hence,

$$\lim_{n\to\infty} E[\boldsymbol{\mathcal{I}}(n)] = O\left(\frac{-\epsilon\log(\epsilon)}{2}\log\left(1 + \frac{\gamma^2(\mu_{\max})\mathcal{P}_2}{-\epsilon\log(\epsilon)\sigma^2}\right)\right). \quad (17)$$

Expression (17) yields the statement of our lemma. Moreover, note that for small ϵ, we may write that

$$O\left(\frac{-\epsilon\log(\epsilon)}{2}\log\left(1 + \frac{\gamma^2(\mu_{\max})\mathcal{P}_2}{-\epsilon\log(\epsilon)\sigma^2}\right)\right) = O(\epsilon\log^2(\epsilon)\log(-\log(\epsilon))). \quad \diamond$$

Let us now consider $\mathcal{C}_U^1$ and obtain an upper bound on its capacity, C_U^1.

Lemma 4: The capacity C_U^1 of $\mathcal{C}_U^1$ is upper-bounded by

$$C_U^1 \leq \left[\frac{1}{2}\sum_{m=1}^{M}\sum_{\mu_i^{(m)}\in S_m} p_m(\epsilon)\log\left(1 + \frac{P(m)\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_i^{(m)}\right] + O(-\epsilon\log(\epsilon)) \quad (18)$$

subject to $\sum_{m=1}^{M} P(m)p_m(\epsilon) \leq \mathcal{P}_1$.

Let us define $\Theta(k)$ to be the duration of a stay in a macro-state for the $(k-1)$th transition to a macro-state. From our model, since the micro-state transition matrix is of the form $\underline{A} + \epsilon\underline{B}$, then $\exists\varpi$ such that $\forall k$

$$E[\Theta(k)] > \frac{1}{\varpi\epsilon}. \quad (19)$$

Furthermore, let us define $\overline{\Theta}(n)$ to be the sample mean length of stay in a macro-state up until time n. Using the Law of Large Numbers for decorrelating random variables applied to $\Theta(k)$, we obtain that

$$\lim_{n\to\infty} P\left(\overline{\Theta}(n) > \frac{1}{\varpi\epsilon}\right) = 1. \quad (20)$$

Combining (20) and (19), we obtain that, as $n \to \infty$, the limit of the probability that the time we spend transmitting over $\mathcal{C}_U^1$ is at least $O(\frac{1}{\epsilon})$ is 1.

By definition, intervals for which $\mathcal{C}_U^1$ has a nonzero output satisfy condition (8). Let $\underline{\boldsymbol{P}}$ represent a vector of power assignments over n symbols. Since we know that to maximize mutual information, we may use symbols that have a Gaussian distribution and are mutually independent symbol to symbol, we can simply consider the problem of maximizing mutual information over the set of power assignments. If we knew, for all the n symbols, what the SCSI is for the past, present, *and* future, then we could only achieve a higher maximum. Conditioned on that information, we can select for a particular realization of the SCSI over the n symbols a vector $\underline{\boldsymbol{P}}_U$ which maximizes the average mutual information for the n symbols. Note that $\underline{\boldsymbol{P}}_U$ is a function of the SCSI realization, but we omit any indexing on that realization for the sake of simplicity. The components of the vector $\underline{\boldsymbol{P}}_U$ are $\boldsymbol{P}_U(n)$, which is the power used at time sample n. Let $\mathcal{D}_m(n)$ be the indicator function which takes the value 1 iff the macro-state at time n is S_m and 0 otherwise. Over n symbols, the maximum average per symbol mutual information between inputs and outputs, conditioned on the RCSI and subject to the average power constraint $\frac{\sum_{k=1}^{n}\boldsymbol{P}_U(k)}{n} \leq \mathcal{P}_1$, is $\Im(n)$ given by

$$\frac{1}{2n}\sum_{m=1}^{M}\sum_{\mu_i^{(m)}\in S_m}\sum_{k=1}^{n}\mathcal{D}_m(k) \times \log\left(1 + \frac{\boldsymbol{P}_U(k)\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\beta_i^{(m)}(k) \quad (21)$$

where $\beta_i^{(m)}(k)$ is, as defined previously, the probability of being in micro-state $\mu_i^{(m)}$ of macro-state S_m at time k, conditioned on our past, present, and future SCSI over all n symbols. Hence, from our results in Theorem 2, we may write that

$$\Im(n) = \frac{1}{2n}\sum_{m=1}^{M}\sum_{\mu_i^{(m)}\in S_m}\sum_{k=1}^{n}\mathcal{D}_m(k) \times \log\left(1 + \frac{\boldsymbol{P}_U(k)\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)} + O(-\epsilon\log(\epsilon)) \quad (22)$$

subject to the average power constraint $\frac{\sum_{k=1}^{n} \boldsymbol{P}_U(k)}{n} \leq \mathcal{P}_1$. For ease of notation, we may perform the following bijective mapping on $(1, 2, \ldots, n)$:

$$g : (1, 2, \ldots, n) \mapsto (1, 2, \ldots, n)$$

$m \to g(m)$ so that if $S(k) = S_m$, $S(j) = S_{m'}$, and $m < m'$, then $g(k) < g(j)$ and if $S(k) = S(j) = S_m$, then $k < j$ implies $g(k) < g(j)$.

Let g^{-1} be the inverse mapping of g. Let us define the random variable $\mathcal{N}_m(n) = \sum_{k=1}^{n} \mathcal{D}_m(k)$, which takes sample values n_m. Random variable $\mathcal{N}_m(n)$ represents the time spent in macro-state m in time 1 through n. We also define the random variable $\Upsilon_m(n) = \sum_{j=1}^{m} \mathcal{N}_j(n)$ and we denote its sample values by υ_m. Random variable $\Upsilon_m(n)$ represents the time spent in macro-state m or lower in time 1 through n. By definition, $\Upsilon_0(n) = 0$. Note that the realizations of $\Upsilon_m(n)$ and $\mathcal{N}_m(n)$ over all m are homomorphic. Using these definitions, we may rewrite (22) as

$$\begin{aligned}\Im(n) = \frac{1}{2n} \sum_{m=1}^{M} \sum_{\mu_{(i)}^{(m)} \in S_m} & \\ \times \sum_{k=\upsilon_{m-1}+1}^{\upsilon_m} & \log\left(1 + \frac{\boldsymbol{P}_U(g^{-1}(k))\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_i^{(m)} \\ & + O\left(-\epsilon \log(\epsilon)\right) \end{aligned} \tag{23}$$

subject to the average power constraint

$$\frac{\sum_{m=1}^{M} \sum_{k=\upsilon_{m-1}+1}^{\upsilon_m} \boldsymbol{P}_U(g^{-1}(k))}{n} \leq \mathcal{P}_\infty \tag{24}$$

for every realization. We may write that

$$C_U^1 \leq \lim_{n\to\infty} E_{\underline{S}_1^n}[\Im(n)] \tag{25}$$

where we are still subject to the average power constraint (24) for every realization of the $\underline{S}$'s. We have, from (23) and (25), (26) at the bottom of the page, subject to the average power constraint (24) for every realization of the S's. Using the concavity of $\ln(1 + x)$ and Jensen's inequality, we get (27) also at the bottom of the page, subject to (24) for every realization of the S's. We can rewrite (24) as

$$\sum_{m=1}^{M} \sum_{k=\upsilon_{m-1}+1}^{\upsilon_m} \frac{\boldsymbol{P}_U\left(g^{-1}(k)\right)}{n_m} \frac{n_m}{n} \leq \mathcal{P}_1. \tag{28}$$

The constraint can be weakened (thus giving a further upper bound) if we replace (28), which is a constraint for each realization, by its expected value over the S's, conditioned in the values of the υ_m's. The weakened constraint is

$$\sum_{i=1}^{M} \frac{E_{\underline{S}^n|\underline{\Upsilon(n)}_1^m}\left[\sum_{k=\upsilon_{m-1}+1}^{\upsilon_m} \boldsymbol{P}_U\left(g^{-1}(k)\right)\right]}{n_m} \frac{n_m}{n} \leq \mathcal{P}_1 \tag{29}$$

for every possible realization of $\underline{\Upsilon(n)}_1^m$. Thus, using Jensen's inequality, and the relaxed constraint in (29), we may write (30) at the bottom of the page.

Since power assignments only depend on the SCSI

$$E_{\underline{S}^n|\underline{\Upsilon(n)}_1^m}\left[\sum_{k=\upsilon_{m-1}+1}^{\upsilon_i m} \boldsymbol{P}_U\left(g^{-1}(k)\right)\right]$$

is proportional to n_m, the realization of $\mathcal{N}_m(n)$ over which we condition. Thus,

$$\frac{E_{\underline{S}^n|\underline{\Upsilon(n)}_1^m}\left[\sum_{k=\upsilon_{m-1}+1}^{\upsilon_i m} \boldsymbol{P}_U\left(g^{-1}(k)\right)\right]}{n_m}$$

is a constant, which we define to be $P(m)$. We may thus modify upper bound (30) as in (31) at the bottom of the next page, subject to

$$\sum_{m=1}^{M} P(m)\frac{n_m}{n} \leq \mathcal{P}_1 \tag{32}$$

$$C_U^1 \leq \lim_{n\to\infty} E_{\underline{S}^n}\left[\frac{1}{2}\sum_{m=1}^{M}\sum_{i\in S_m}\sum_{k=\upsilon_{m-1}+1}^{\upsilon_m}\frac{n_m}{nn_m}\log\left(1 + \frac{\boldsymbol{P}_U(g^{-1}(k))\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)}\right] + O\left(-\epsilon\log(\epsilon)\right) \tag{26}$$

$$C_U^1 \leq \lim_{n\to\infty} E_{\underline{S}^n}\left[\frac{1}{2}\sum_{m=1}^{M}\sum_{i\in S_m}\frac{n_m}{n}\log\left(1 + \frac{\frac{\sum_{k=\upsilon_{m-1}+1}^{\upsilon_m}\boldsymbol{P}_U(g^{-1}(k))}{n_m}\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)}\right] + O(-\epsilon\log(\epsilon)) \tag{27}$$

$$\begin{aligned}C_U^1 \leq \lim_{n\to\infty} E_{\mathcal{N}_m(n)}\left[\frac{1}{2}\sum_{m=1}^{M}\sum_{\mu_i^{(m)}\in S_m}\frac{n_m}{n}\log\left(1 + \frac{E_{\underline{S}_1^n|\mathcal{N}_m(n)}\left[\sum_{k=\upsilon_{m-1}+1}^{\upsilon_m}\boldsymbol{P}_U(g^{-1}(k))\right]}{n_m}\frac{\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)}\right] \\ + O(-\epsilon\log(\epsilon)). \end{aligned} \tag{30}$$

for every realization of $\underline{\mathcal{N}(n)}_1^m$. We may, yet again, weaken our constraint by replacing $\overline{(n_m)}/(n)$ by its expectation, yielding the constraint $\sum_{m=1}^M P(m)p_m(\epsilon) \leq \mathcal{P}_1$. This constraint, together with (31), is the expression given in Theorem 1 by (3), to within $O(-\epsilon\log(\epsilon))$ when $\mathcal{P}_1 = \mathcal{P}$. ◇

We may now use Lemmas 3 and 4 to find an upper bound to $C_{\rm true}(\epsilon)$. Moreover, we further upper-bound C_U^1 and C_U^2 by allowing $\mathcal{P}_1 = \mathcal{P}_2 = \mathcal{P}$. Combining these upper bounds with our lower bound (9), we obtain Theorem 1.

V. Conclusion

We have shown that the assumption that fast and slow fades are both Markovian is accurate within $O(\epsilon\log^2(\epsilon)\log(-\log(\epsilon)))$ in the case where the fast fades are Markovian and the slow fades are quasi-Markovian. Moreover, the power-allocating policy used in [1] for the case where both slow and fast fades are Markovian achieves this approximate capacity. Our results show that, as expected, the policies and capacity computations for the case when slow and fast fades are Markovian are approximately accurate when the slow fades are quasi-Markovian. However, the rate of convergence is much slower than ϵ.

Our results only consider one aspect of the question of the robustness of capacity computation and capacity-achieving policies to channel models. Many simplified models, such as block-fading channels, are attractive from the point of view of tractability but are clearly very simplified. The sensitivity of the results obtained by using such models is an interesting question to justify the applicability of results obtained through simple models.

Appendix

Proof of Lemma 1: Consider the elements of $\underline{\tilde{q}}_1^{(m)}$: see (33) at the bottom of the page, where we have used the fact that the denominator in the last line of that equation is equal to one. Next, note (34) at the bottom of the page. Comparing the expressions in (33) and (34), for $\alpha_i^{(m)}(1)$ and $\tilde{q}_1^{(m)}(i)$, we may readily see that

$$|\tilde{q}_1^{(m)}(i) - \alpha_i^{(m)}(1)| = O(\epsilon).$$

In general, we have the first equation at the top of the following page and we have (35) also at the top of the following page, where we have made use of the fact that the denominators in (35) are all equal to 1.

Now, by induction, the proof follows.

Proof of Lemma 2: Let $\underline{\tilde{\pi}}^{(m)}$ be the stationary probability vector associated with the stochastic matrix $\underline{\tilde{P}}^{(m)}$. The vectors $\underline{\tilde{\pi}}^{(m)}$ and $\underline{\pi}^{(m)}$ are the left eigenvectors of $\underline{\tilde{P}}^{(m)}$ and $\underline{A}^{(m)}$, respectively, corresponding to the eigenvalue of 1.

Note, from (4) and (5), that the elements of $\underline{\tilde{P}}^{(m)}$ and $\underline{A}^{(m)}$ differ only by $O(\epsilon)$. In general, eigenvectors are not continuous functions of the elements of the matrix ([7, p. 373]. However,

$$C_U^1 \leq \lim_{n\to\infty} E_{\underline{\mathcal{N}(n)}_1^m}\left[\frac{1}{2}\sum_{m=1}^M \sum_{\mu_i^{(m)}\in S_m} \frac{n_m}{n}\log\left(1+\frac{P(m)\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)}\right] + O(-\epsilon\log(\epsilon))$$
$$= \left[\frac{1}{2}\sum_{m=1}^M \sum_{\mu_i^{(m)}\in S_m} p_m(\epsilon)\log\left(1+\frac{P(m)\gamma^2(\mu_i^{(m)})}{\sigma^2}\right)\pi_{(i)}^{(m)}\right] + O(-\epsilon\log(\epsilon)) \quad (31)$$

$$\begin{aligned}\tilde{q}_1^{(m)}(i) &= \sum_{\mu_j^{(m)}\in S_m} \tilde{P}^{(m)}(j,i)q_0^{(m)}(j)\\ &= \sum_{\mu_j^{(m)}\in S_m}\left(A^{(m)}(j,i)+\epsilon B\left(\mu_j^{(m)},\mu_i^{(m)}\right)\right)q_0^{(m)}(j) + \epsilon\sum_{\mu_l^{(r)}\notin S_m} B\left(\mu_i^{(m)},\mu_l^{(r)}\right)q_0^{(m)}(i)\\ &= \frac{\sum_{\mu_j^{(m)}\in S_m}\left(A^{(m)}(j,i)+\epsilon B\left(\mu_j^{(m)},\mu_i^{(m)}\right)\right)q_0^{(m)}(j) + \epsilon\sum_{\mu_l^{(r)}\notin S_m} B\left(\mu_i^{(m)},\mu_l^{(r)}\right)q_0^{(m)}(i)}{\sum_{\mu_k^{(m)},\mu_j^{(m)}\in S_m}\left(A^{(m)}(j,k)+\epsilon B\left(\mu_j^{(m)},\mu_k^{(m)}\right)\right)q_0^{(m)}(j) + \epsilon\sum_{\mu_k^{(m)}\in S_m,\mu_l^{(r)}\notin S_m} B\left(\mu_k^{(m)},\mu_l^{(r)}\right)q_0^{(m)}(k)}\end{aligned} \quad (33)$$

$$\begin{aligned}\alpha_i^{(m)}(1) &= \frac{\mathrm{Prob}(T(1)=\mu_i^{(m)}|S(0)=S_m, \underline{T}_{-L'}^0 = \underline{t}_{-L'}^0)}{\mathrm{Prob}(S(1)=S_m|S(0)=S_m, \underline{T}_{-L'}^0 = \underline{t}_{-L'}^0)}\\ &= \frac{\sum_{\mu_j^{(m)}\in S_m}\left(A^{(m)}(j,i)+\epsilon B\left(\mu_j^{(m)},\mu_i^{(m)}\right)\right)q_0^{(m)}(j)}{\sum_{\mu_k^{(m)},\mu_j^{(m)}\in S_m} q_0^{(m)}(j)\left(A^{(m)}(j,k)+\epsilon B\left(\mu_j^{(m)},\mu_k^{(m)}\right)\right)}.\end{aligned} \quad (34)$$

$$\alpha_i^{(m)}(n) = \frac{\sum_{\mu_j^{(m)} \in S_m} \left(A^{(m)}(j,i) + \epsilon B\left(\mu_j^{(m)}, \mu_i^{(m)}\right)\right) \alpha_j^{(m)}(n-1)}{\sum_{\mu_k^{(m)}, \mu_j^{(m)} \in S_m} \left(A^{(m)}(j,k) + \epsilon B\left(\mu_j^{(m)}, \mu_k^{(m)}\right)\right) \alpha_j^{(m)}(n-1)}$$

$$\tilde{q}_n^{(m)}(i) = \frac{\sum_{\mu_j^{(m)} \in S_m} \tilde{q}_{n-1}^{(m)}(j) \tilde{P}(j,i)}{\sum_{\mu_k^{(m)}, \mu_j^{(m)} \in S_m} \tilde{q}_{n-1}^{(m)}(j) \tilde{P}(j,k)}$$
$$= \frac{\sum_{\mu_j^{(m)} \in S_m} \tilde{q}_{n-1}^{(m)}(j) \left(A^{(m)}(j,i) + \epsilon B\left(\mu_j^{(m)}, \mu_i^{(m)}\right)\right) + \epsilon \sum_{\mu_l^{(m)} \notin S_m} B\left(\mu_l^{(r)}, \mu_i^{(m)}\right) \tilde{q}_{n-1}^{(r)}(l)}{\sum_{\mu_j^{(m)}, \mu_k^{(m)} \in S_m} \tilde{q}_{n-1}^{(m)}(j) \left(A^{(m)}(j,k) + \epsilon B\left(\mu_j^{(m)}, \mu_k^{(m)}\right)\right) + \epsilon \sum_{\mu_k^{(m)} \in S_m, \mu_l^{(r)} \notin S_m} B\left(\mu_l^{(r)}, \mu_k^{(m)}\right) \tilde{q}_{n-1}^{(r)}(l)} \quad (35)$$

we have that, for stochastic matrices and the type of perturbation that we consider, the stationary probability is not only a continuous function of the elements of the matrix, but in fact the following relationship holds:

$$\tilde{\pi}^{(m)}(i) - \pi^{(m)}(i) = O(\epsilon), \qquad \forall i. \quad (36)$$

To show that (36) is true, we first note that $\underline{\tilde{\pi}}^{(m)}$ is the solution of $\underline{\mathcal{P}}^{(m)} \underline{\tilde{\pi}}^{(m)} = \underline{e}$, where $\underline{\mathcal{P}}^{(m)}$ is the matrix $(\underline{I} - \underline{\tilde{P}}^{(m)})$ whose N_mth row is replaced by a column vector of 1's and $\underline{e} = (0, 0, \dots, 0, 1)'$. Thus, $\tilde{\pi}^{(m)}(i)$ can be obtained from Cramer's rule as

$$\tilde{\pi}^{(m)}(i) = \frac{\det(\underline{\mathcal{P}}^{(m)} \leftarrow_i \underline{e})}{\det(\underline{\mathcal{P}}^{(m)})} \quad (37)$$

where, following [7, p. 21], $\underline{\mathcal{P}}^{(m)} \leftarrow_i \underline{e}$ denotes a matrix whose ith column is $\underline{e}$ and the rest of the columns coincide with those of $\underline{\mathcal{P}}^{(m)}$. From the structure of $\underline{\tilde{P}}^{(m)}$, we see that the numerator and denominator of (37) are $O(\epsilon)$ perturbations of the corresponding expressions for the elements of $\underline{\pi}^{(m)}$, thus establishing (36).

Next, we show that there exists a constant K such that

$$|\tilde{q}^{(m)}(n) - \tilde{\pi}^{(m)}| < K \lambda_{2,m}^n, \quad (38)$$

for any $n > N_m$. We first note that, by the Cayley-Hamilton theorem [7], for $n \geq N_m$

$$\left(\underline{\tilde{P}}^{(m)}\right)^n = \sum_{l=0}^{N_m - 1} \phi_l(n) (\underline{\tilde{P}}^{(m)})^l$$

for some constants $\phi_l(n)$. Furthermore, the above equation is also satisfied if $\underline{\tilde{P}}^{(m)}$ is replaced by any of its eigenvalues. Denoting the eigenvalues by $1, \lambda_{2,m}, \dots, \lambda_{N_m,m}$, we have

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_{2,m} & \lambda_{2,m}^2 & \dots & \lambda_{2,m}^{N_m-1} \\ \vdots & & & \vdots \\ \lambda_{N_m,m} & \lambda_{N_m,m}^2 & \dots & \lambda_{N_m,m}^{N_m-1} \end{pmatrix} \times \begin{pmatrix} \phi_0(n) \\ \phi_1(n) \\ \vdots \\ \phi_{N_m-1}(n) \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_{2,m}^n \\ \vdots \\ \lambda_{N_m,m}^n \end{pmatrix}. \quad (39)$$

Using Lagrange's interpolation formula [7], we can write

$$\sum_{l=0}^{N_m-1} \phi_l(n) \lambda^l = \sum_{l=1}^{N_m} f_i^n(\lambda) (\lambda_{i,m})^n$$

where $\lambda_{1,m} = 1$ and

$$f_i^n(\lambda) = (\lambda_{i,m})^n \frac{\prod_{j=1, j \neq i}^{N_m} (\lambda - \lambda_{j,m})}{\prod_{j=1, j \neq i}^{N_m} (\lambda_{i,m} - \lambda_{j,m})}.$$

Thus,

$$\left(\underline{\tilde{P}}^{(m)}\right)^n = \sum_{i=1}^{N_m} F_i^n \left(\underline{\tilde{P}}^{(m)}\right) (\lambda_{i,m})^n$$

where

$$F_i^n(\tilde{P}^{(m)}) = (\lambda_{i,m})^n \frac{\prod_{j=1, j \neq i}^{N_m} \left(\underline{\tilde{P}}^{(m)} - \lambda_{j,m} \underline{I}\right)}{\prod_{j=1, j \neq i}^{N_m} (\lambda_{i,m} - \lambda_{j,m})}.$$

Since $\underline{\tilde{P}}^{(m)}$ is a stochastic matrix, $|\lambda_{i,m}| < 1$ for $i > 1$. Since

$$\underline{\tilde{q}}^{(m)}(n) = \underline{p}_0^{(m)} (\underline{P}^{(m)})^n \rightarrow \underline{\tilde{\pi}}^{(m)}$$

as $n \rightarrow \infty$, we have that $\underline{p}^{(m)}(0) F_1(\underline{\tilde{P}}^{(m)}) = \underline{\tilde{\pi}}^{(m)}$. Thus,

$$|\underline{\tilde{q}}^{(m)}(n) - \underline{\tilde{\pi}}^{(m)}| < |\lambda_{2,m}|^n \sum_{i=2}^{N_m} |F_i(\underline{\tilde{P}}^{(m)})| \quad (40)$$

where $|F_i(\underline{\tilde{P}}^{(m)})|$ denotes a matrix whose elements are the magnitudes of the elements of $F_i(\underline{\tilde{P}}^{(m)})$. Since $\underline{A} + \epsilon \underline{B}$ must be a stochastic matrix, ϵ is constrained to lie in a compact set around zero. Further, owing to the fact that $F_i(\underline{\tilde{P}}^{(m)})$ is a continuous function of ϵ (which is a consequence of the fact that the eigenvalues of a matrix are continuous functions of the elements of the matrix), each element of $|F_i(\underline{\tilde{P}}^{(m)})|$ can be upper-bounded. Thus, from (5), (40) implies that there exists a K such that (38) holds. Now the statement of the lemma follows from (38), Lemma 1, and (36).

Proof of Theorem 2: We first note $\beta_i^{(m)}(n)$ defined in (41) at the bottom of the page. Let us define $\rho_{i,j}^{(m)}$ to be as shown in (42) and continue the proof with (43)–(45) at the bottom of the page.

REFERENCES

[1] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2020, Sep. 1999.

$$\beta_i^{(m)}(n) = \frac{\alpha_i^{(m)}(n)\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m, \underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_i^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}\right)}{\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m, \underline{T}_L^s = \underline{t}_L^s | \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L}\right)}. \tag{41}$$

$$\begin{aligned}\rho_{i,j}^{(m)} &= \frac{\beta_i^{(m)}(n)}{\beta_j^{(m)}(n)}\\ &= \frac{\alpha_i^{(m)}(n)\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m, \underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_i^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}\right)}{\alpha_j^{(m)}(n)\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m, \underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_j^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}\right)}\\ &\quad \text{using Bayes's rule}\\ &= \frac{\alpha_i^{(m)}(n)\text{Prob}\left(\underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_i^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L}, \underline{T}_{n+1}^{L-1} \in S_m\right)}{\alpha_j^{(m)}(n)\text{Prob}\left(\underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_j^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L}, \underline{T}_{n+1}^{L-1} \in S_m\right)}\\ &\quad \times \frac{\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m | T(n) = \mu_i^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}\right)}{\text{Prob}\left(\underline{T}_{n+1}^{L-1} \in S_m | T(n) = \mu_j^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}\right)}.\end{aligned} \tag{42}$$

Consider

$$\begin{aligned}&\text{Prob}(\underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_j^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}, \underline{T}_{n+1}^{L-1} \in S_m)\\ &\text{by Markovicity}\\ &= \text{Prob}(\underline{T}_L^s = \underline{t}_L^s | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-1} \in S_m)\\ &= \frac{\text{Prob}(\underline{T}_L^s = \underline{t}_L^s, T_{L-1} \in S_m | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}{\text{Prob}(T_{L-1} \in S_m | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}\\ &= \frac{\sum_{\mu_k^{(m)} \in S_m} \text{Prob}(\underline{T}_L^s = \underline{t}_L^s, T_{L-1} = \mu_k^{(m)} | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}{\sum_{\mu_k^{(m)} \in S_m} \text{Prob}(T_{L-1} = \mu_k^{(m)} | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}\\ &= \frac{\sum_{\mu_k^{(m)} \in S_m} \text{Prob}(\underline{T}_L^s = \underline{t}_L^s | T_{L-1} = \mu_k^{(m)})\text{Prob}(T_{L-1} = \mu_k^{(m)} | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}{\sum_{\mu_k^{(m)} \in S_m} \text{Prob}(T_{L-1} = \mu_k^{(m)} | T(n) = \mu_j^{(m)}, \underline{T}_{n+1}^{L-2} \in S_m)}\\ &\text{using Lemma 2}\\ &= \frac{\sum_{\mu_k^{(m)} \in S_m} \text{Prob}(\underline{T}_L^s = \underline{t}_L^s | T_{L-1} = \mu_k^{(m)})(\pi^{(m)}(k) + O(\epsilon))}{\sum_{\mu_k^{(m)} \in S_m} \pi^{(m)}(k) + O(\epsilon)}\\ &= \frac{\text{Prob}(\underline{T}_L^s = \underline{t}_L^s)}{1 + O(\epsilon)}\end{aligned} \tag{43}$$

by a similar argument

$$\text{Prob}(\underline{T}_{n+1}^{L-1} \in S_m | T(n) = \mu_j^{(m)}, \underline{T}_1^n \in S_m, \underline{T}^0_{-L'} = \underline{t}^0_{-L'}) = \frac{\text{Prob}(\underline{T}_{n+1}^{L-1} \in S_m)}{1 + O(\epsilon)} \tag{44}$$

hence

$$\rho_{i,j}^{(m)} = \frac{\pi^{(m)}(i)}{\pi^{(m)}(j)} + O(\epsilon). \tag{45}$$

[2] A. Das and P. Narayan, “Capacities of time-varying multiple-access channels with side information,” *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, Jan. 2002.

[3] W. Feller, *An Introduction to Probability Theory and its Applications*. New York: Wiley, vol. I.

[4] R. G. Gallager, *Information Theory and Reliable Communications*. New York: Wiley, 1968.

[5] ——, *Discrete Stochastic Processes*. Norwell, MA: Kluwer Academic, 1996.

[6] A. J. Goldsmith and M. Médard, “Capacity of time-varying channels with causal side information,” IEEE Trans. Inf. Theory 1999, submitted for publication.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[8] D. L. Neuhoff and P. L. Shields, “Channels with almost finite memory,” *IEEE Trans. Inf. Theory*, vol. IT-25, no. 4, pp. 440–447, July 1979.

[9] D. L. Neuhoff and P. L. Shields, “Indecomposable finite state channels and primitive approximation,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 11–18, Jan. 1982.

[10] R. G. Phillips and P. V. Kokotovic, “A singular perturbation approach to modeling and control of Markov chains,” *IEEE Trans. Autom. Control*, vol. AC-26, no. 5, pp. 1087–1094, Oct. 1981.

[11] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. San Francisco, CA: Holden-Day, 1964, Translated by A. Feinstein.

[12] H. Viswanathan, “Capacity of time-varying channels with delayed feedback,” *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.

[13] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. New York: Springer-Verlag, 1978.