

Capacities of Time-Varying Multiple-Access Channels With Side Information

Arnab Das and Prakash Narayan, *Fellow, IEEE*

Abstract—We determine the capacity regions for a class of time-varying multiple-access channels (TVMACs), when the underlying channel state evolves in time according to a probability law which is known to the transmitters and the receiver. Additionally, the transmitters and the receiver have access to varying degrees of channel state information (CSI) concerning the condition of the channel. Discrete-time channels with finite input, output, and state alphabets are considered first. The special case of a TVMAC, with the channel state process being a time-invariant, indecomposable, aperiodic Markov chain, shows a surprising anomaly in that imperfect transmitter CSI can cause the capacity under some distributions for the initial state to be strictly larger than that under a stationary distribution for the initial state. We also study a time-varying multiple-access fading channel with additive Gaussian noise, when various amounts of CSI are provided to the transmitters and perfect CSI is available to the receiver, and the fades are assumed to be stationary and ergodic. Implications for transmitter power control are discussed.

Index Terms—Capacity, channel state information (CSI), fading channel, multiple access, power allocation, time-varying channel.

I. INTRODUCTION

THIS paper constitutes a systematic study of the capacity problem for a class of time-varying multiple-access channels (TVMACs), when varying degrees of side information concerning the condition of the channel are provided to the transmitters and the receiver. In these models, the probability law characterizing the channel between the transmitters and the receiver can change with time. This time-varying behavior of the channel probability law is typically described in terms of the evolution of an underlying channel “state,” which is a description of the salient parameters of the TVMAC; the corresponding side information concerning the channel state is referred to as *channel state information* (CSI). An example of a TVMAC which has received much attention recently arises

in mobile wireless communication; here, owing to the mobility of the users, the degradation of the transmitted signals due to multipath, shadowing, and propagation losses, is time-varying.

Different statistical assumptions concerning the evolution of the state of a TVMAC lead to different channel models. For instance, the channel state may be assumed to remain constant, albeit unknown, throughout the duration of a transmission, giving rise to a compound channel model (cf., e.g., [1], [2]). If, on the other hand, the channel state is assumed to evolve in an arbitrary manner, thereby requiring an analysis under a “worst” concatenation of state realizations, then the appropriate model is referred to as an arbitrarily varying channel (cf., e.g., [3], [4]). Yet another approach, which we adopt, assumes that the channel state evolves autonomously in accordance with a (known) probability law which allows the state at each time to depend on all the past states, but not on the previous inputs or outputs of the channel. Reliable rates of communication are then characterized with the probability of decoding error being assessed as an average with respect to the probability law of the state sequence.

The relevant literature concerning time-varying channels, in general, can be broadly classified into three categories, in terms of the channel models considered or the nature of the problems investigated. For a survey of such channels, and, in particular, mobile wireless channels, the reader is referred to [5] (see also [4]).

The first category, comprising the (single-sender) *finite-state channel* (FSC) model and its variants, assumes finite channel input and output alphabets, and a finite state space (cf., e.g., [6]). The capacity of an FSC, under various conditions, has been determined in [1], [2], [6], [7]. Of particular importance is a special class of FSCs consisting of *indecomposable* FSCs (cf., e.g., [6]). The capacity of an indecomposable FSC when no CSI is available to the transmitter and receiver, has been determined by Blackwell, Breiman, and Thomasian [7] and Gallager [6]; in particular, it admits a “product-space” characterization which does not depend on the initial state of the channel [6], [7]. When the underlying state process is a time-invariant, indecomposable,¹ aperiodic Markov chain, and the state at each time does not depend on the past channel inputs or outputs, the resulting indecomposable FSC, which is known as a finite-state Markov channel (FSMC) (cf., e.g., Jelinek [9]), serves as a useful model for mobile wireless channels [10], [11]. The capacity of an FSMC when perfect CSI is available at the transmitter or receiver is determined in [6], [9], [12]; here, “single-letter” capacity formulas are available only in the situation when

¹A Markov chain is indecomposable if it contains at most one closed communicating set of states. This definition differs slightly from that used traditionally (cf. e.g., [8]).

Manuscript received June 2, 2000; revised July 3, 2001. This work was supported by the Institute for Systems Research at the University of Maryland, College Park, under NSF Grant OIR-85-00108; the Maryland Procurement Office under Grant MDA90497C3015; the Center for Satellite and Hybrid Communication Networks, a NASA Commercial Space Center, under NASA Cooperative Agreement NCC3-528; by the Army Research Laboratory under Cooperative Agreement DAALO1-96-2-0002; and by DARPA under Grant 528938.

A. Das was with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA. He is now with Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733 USA (e-mail: arnab@lucent.com).

P. Narayan is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: prakash@eng.umd.edu).

Communicated by I. Csiszár, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(02)00058-5.

perfect CSI is provided to the receiver. From a practical standpoint, perfect or instantaneous CSI is seldom available to either the transmitter or the receiver. Recent works which deal with associated issues include [13], where the capacity of an FSMC is determined when perfect but delayed CSI is available to the transmitter and receiver; and [14], where the capacity for a class of single-sender time-varying channels is evaluated when the state process is stationary and ergodic, and perfect CSI is available to the receiver; the transmitter CSI, on the other hand, is noisy, and is subject to certain technical assumptions. Some of the models in our paper constitute generalizations to the multiple-access situation of single-sender time-varying channel models studied by Caire and Shamai [14].

The second category deals with the *time-varying multiple-access fading channel* (TVFC), which serves as a (discrete-time) model for the uplink of a mobile wireless channel with multiple senders transmitting to a single base station [15], and subject to frequency nonselective fading (see Section II for a precise definition of the TVFC). Here, the positive real-valued fades experienced by the senders at each time can be regarded as the state of the channel. Two simplifying assumptions, used extensively in studying the TVFC, are: the state process is stationary and ergodic, and the state at each time is fully known to the receiver. Given these assumptions, the capacity region of the TVFC when no CSI or perfect CSI is available to the transmitters has a single-letter characterization (cf., e.g., [11], [15]–[18], and references therein). For the single-sender TVFC, situations where delayed but perfect CSI is available to the transmitter, or when the transmitter CSI is not perfect but is known to satisfy certain technical conditions, have been studied in [13] and [14], respectively.

In the final category, multiple-access and single-sender channels have been studied under very general statistical conditions by Han [19] and Verdú and Han [20], and capacity formulas obtained in terms of normalized (product–space) information densities. Although the channels considered in [19], [20] do not have the notion of an underlying channel state, some of our results rely on the approaches therein.

In this paper, we begin by considering a (discrete-time) two-sender multiple-access channel, with finite input, output, and state alphabets, where the CSI is specified in terms of deterministic mappings of the channel state. Following the results of Han [19], the capacity region of this channel can be characterized in terms of information densities involving map-valued random variables (rvs), also referred to as “strategies” (cf., e.g., [14]). When the encoder CSI is contained in that available to the decoder, we indicate that the capacity region admits a simpler and more tractable characterization involving familiar channel input-valued rvs rather than abstract “strategies.”

We then consider the special case of a “memoryless” TVMAC, with the state process being a time-invariant, indecomposable, aperiodic Markov chain. This is a natural multiple-access counterpart of the (single-sender) FSMC [6], [9]. A surprising result is that unlike existing results concerning the FSMC in [6], [9], the capacity region—when the encoder CSI is not perfect—will depend, in general, on the probability mass function (pmf) of the initial state; indeed, it can be *strictly larger than the capacity region under a stationary pmf for*

the initial state. In fact, this is shown by considering the case of a single-sender FSMC. *This result does not rely on the general and elaborate formulation in Section II which is used throughout the paper, and can be understood in isolation*. Sufficient conditions are also presented under which this capacity region does not depend on the pmf of the initial state. Next, in order to reduce transmitter complexity, we restrict the encoders at each time instant to depend only on a limited “time window” of CSI. The corresponding capacity regions are determined for two specific cases: a TVMAC with a Markov state process as above, and a TVMAC with the state process being stationary and ergodic (but not necessarily Markov), with perfect decoder CSI. Also, under certain technical conditions, the capacity regions of these TVMACs without any such restriction are shown to result from those with encoder restrictions.

Our final concern is with a TVFC. The capacity region of the TVFC, when various degrees of CSI are available to the transmitters and *perfect* CSI is available to the receiver, is obtained under the assumption that the state process is stationary and ergodic. A simple encoding strategy for achieving rate pairs in the capacity region is also presented.

It is not surprising that some of our results for the capacity regions of TVMACs with varying extents of CSI at the transmitters and receiver, are characterized in terms of product–space formulas. This is already true for certain known special cases involving a single sender. Furthermore, our results enable us to draw new conclusions even for the single-sender TVC, in addition to affording alternative proofs for known results in [13], [14] (cf. Corollary 2 and Corollary 3).

The remainder of this paper is organized as follows. The various channel models are described in Section II, which also contains the definitions relevant to our analysis. In Section III, we state our capacity results, the proofs of which are presented in Section IV. We conclude with a discussion of our results, including transmitter power control for the TVFC, in Section V.

II. PRELIMINARIES

Our focus will be on the multiple-access channel consisting of two senders transmitting information to a single receiver. As usual, an extension to situations with more than two senders is conceptually straightforward although cumbersome. Particularizations to situations involving a single sender will also be treated when they are of special interest.

Consider first a discrete-time two-sender *time-varying multiple-access channel* (TVMAC) with (finite) input alphabets $\mathcal{X}_1, \mathcal{X}_2$, (finite) output alphabet \mathcal{Y} , and (finite) “state space” \mathcal{S} . Each use of the channel is governed by a state $s \in \mathcal{S}$. The \mathcal{S} -valued state process $\{S_t\}_{t=1}^{\infty}$ evolves in time according to a (known) probability law $P_{\mathcal{S}}$ which allows the state at any time to depend on the previous states, but not on the previous inputs and outputs of the channel. The state process $\{S_t\}_{t=1}^{\infty}$ is not restricted to be stationary or ergodic. The probability law of the channel is specified by a sequence of (known) pmfs

$$\mathbf{W} = \{W_n(y^n | x_1^n, x_2^n, s^n): x_1^n \in \mathcal{X}_1^n, x_2^n \in \mathcal{X}_2^n, s^n \in \mathcal{S}^n, y^n \in \mathcal{Y}^n\}_{n=1}^{\infty}, \quad (1)$$

where $W_n(\cdot | \cdot, \cdot, \cdot)$ denotes the joint pmf of the channel output sequence for n uses of the channel corresponding to a given

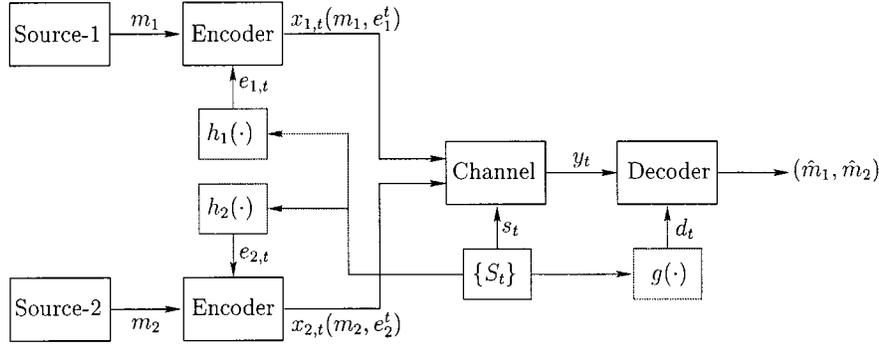


Fig. 1. The TVMAC.

channel input sequence and a prevailing sequence of states. *Our main results concern a restricted situation in which the pmfs in (1) satisfy a “memorylessness” property* (cf. (23)). For notational convenience, we shall hereafter suppress the subscript n and use $W(y^n|x_1^n, x_2^n, s^n)$ instead of $W_n(y^n|x_1^n, x_2^n, s^n)$. Random variables will be denoted by upper case letters, and their realizations by the corresponding lower case letters.

As a special case, the probability law of a single-sender *time-varying channel* (TVC) is specified by a sequence of (known) pmfs

$$\mathbf{W} = \{W(y^n|x^n, s^n): x^n \in \mathcal{X}^n, s^n \in \mathcal{S}^n, y^n \in \mathcal{Y}^n\}_{n=1}^{\infty}. \quad (2)$$

Let \mathcal{E}_1 , \mathcal{E}_2 , and \mathcal{D} be finite sets and

$$h_1: \mathcal{S} \rightarrow \mathcal{E}_1 \quad h_2: \mathcal{S} \rightarrow \mathcal{E}_2 \quad \text{and} \quad g: \mathcal{S} \rightarrow \mathcal{D} \quad (3)$$

be mappings which are used to describe the CSI available to the two senders and the decoder, respectively. Thus, at each time instant t , the encoder for sender-1 (resp., sender-2) is provided with the *instantaneous* CSI $e_{1,t} = h_1(s_t)$ (resp., $e_{2,t} = h_2(s_t)$) while the decoder is provided with CSI $d_t = g(s_t)$, all in a causal manner (see Fig. 1). In general, the CSI available to the senders and the decoder can be all different.² Here, and in the sequel, when referring to channel inputs or encoder CSI, the first subscript will denote the sender and the second subscript, if any, will denote the beginning time instant; the superscript will denote the ending time instant. Also, in this shorthand notation for a sequence, a beginning time instant $t = 1$ will not be explicitly displayed, and will correspond to the absence of a second subscript. For example, x_j^n denotes the length- n channel input sequence $(x_{j,1}, \dots, x_{j,n}) \in \mathcal{X}_j^n$ of sender- j , while $x_{j,k}^n$ denotes the sequence $(x_{j,k}, x_{j,k+1}, \dots, x_{j,n}) \in \mathcal{X}_j^{n-k+1}$. Further, with an abuse of notation, we shall use $h_j(s^n)$ and $h_j(s_k^n)$ to denote the encoder CSI $(h_j(s_1), \dots, h_j(s_n))$ and $(h_j(s_k), \dots, h_j(s_n))$, respectively, for sender- j , $j = 1, 2$. Finally, for any sequence of rvs A^n, B^n , $n \geq 1$, we shall use the following notation when the usage is clear from the context:

$$P_{AB}(a_i^j, b_k^l) = \Pr(A_i^j = a_i^j, B_k^l = b_k^l)$$

and

$$P_{A|B}(a_i^j|b_k^l) = \Pr(A_i^j = a_i^j | B_k^l = b_k^l) \quad (4)$$

where i, j, k, l are positive integers.

²Although at first blush, this manner of description of the CSI might seem restrictive, it can, in fact, accommodate CSI with delays or noisy CSI (cf. Section V for the latter).

For the channel in (1), a length- n block code with message sets $\mathcal{M}_j = \{1, \dots, M_j\}$, $j = 1, 2$, is a triple $(\theta_1, \theta_2, \phi)$; the encoder θ_j for sender- j is given by a sequence of mappings $\{\theta_{j,t}\}_{t=1}^n$, with

$$\theta_{j,t}: \mathcal{M}_j \times \mathcal{E}_j^t \rightarrow \mathcal{X}_j \quad (5)$$

while the decoder is given by

$$\phi: \mathcal{Y}^n \times \mathcal{D}^n \rightarrow \mathcal{M}_1 \times \mathcal{M}_2. \quad (6)$$

Note that while each code symbol is allowed to depend only on the past and current CSI, the decoder is allowed to use the entire length- n CSI in producing its output. The rate of the code is the pair $(\frac{1}{n} \log M_1, \frac{1}{n} \log M_2)$, and the average probability of error $P_e(\theta_1, \theta_2, \phi)$ of the code $(\theta_1, \theta_2, \phi)$ is

$$\begin{aligned} P_e(\theta_1, \theta_2, \phi) &= \frac{1}{M_1 M_2} \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \sum_{s^n \in \mathcal{S}^n} P_{\mathcal{S}}(s^n) \\ &\times \sum_{y^n: \phi(y^n, g(s^n)) \neq (m_1, m_2)} \\ &W(y^n | \theta_1(m_1, h_1(s^n)), \theta_2(m_2, h_2(s^n)), s^n) \end{aligned} \quad (7)$$

where

$$\theta_j(m_j, h_j(s^n)) = (\theta_{j,1}(m_j, h_j(s_1)), \theta_{j,2}(m_j, h_j(s_2)), \dots, \theta_{j,n}(m_j, h_j(s_n))), \quad j = 1, 2.$$

Given $0 < \epsilon < 1$, a pair of nonnegative numbers (R_1, R_2) is an ϵ -achievable rate pair if for every $\delta > 0$ and for all n sufficiently large there exists a length- n block code $(\theta_1, \theta_2, \phi)$ with message sets $\mathcal{M}_j = \{1, \dots, M_j\}$, $j = 1, 2$, such that

$$\begin{aligned} \frac{1}{n} \log M_1 &\geq R_1 - \delta \\ \frac{1}{n} \log M_2 &\geq R_2 - \delta \end{aligned}$$

and $P_e(\theta_1, \theta_2, \phi) < \epsilon$. A pair (R_1, R_2) constitutes an achievable rate pair if it is ϵ -achievable for every $\epsilon > 0$. The set of all achievable rate pairs is the capacity region of the TVMAC in (1), and will be denoted by \mathcal{C} . Note that the capacity region is a closed set by definition.

Next, in order to reduce the complexity of the encoder it may often be desirable to restrict the encoder for sender- j , $j = 1, 2$, to depend only on the *limited* CSI $e_{j, \max\{1, t-k+1\}}^t$ at time t , for some fixed integer $k \geq 1$. Precisely, the encoder θ_j , $j = 1, 2$, of a length- n block code $(\theta_1, \theta_2, \phi)$ for the

TVMAC in (1), is now given by a sequence of mappings $\{\theta_{j,t}: \mathcal{M}_j \times \mathcal{E}_{j, \max\{1, t-k+1\}}^t \rightarrow \mathcal{X}_j\}_{t=1}^n$, while the decoder ϕ is given by (6). We shall call this restriction an ‘‘encoder restriction of window- k .’’ The capacity region under an encoder restriction of window- k will be denoted by $\mathcal{C}(k)$. The situations in which the two encoders are restricted by different window sizes (including when one of the encoders is not restricted at all) can be analyzed as special cases.

In addition to the previous communication situations in which the channel input, output, and state alphabets are all finite, we shall also study the *multiple-access time-varying fading channel* (TVFC) which serves as a (discrete-time) model for the uplink of a wireless channel with two senders transmitting to a single base station [15]. Specifically, the received \mathbb{R} -valued signal is given by³

$$Y_t = \sum_{j=1}^2 \tilde{S}_{j,t} x_{j,t} + N_t, \quad t \geq 1 \quad (8)$$

where $\{x_{j,t}\}_{t=1}^\infty$ and $\{\tilde{S}_{j,t}\}_{t=1}^\infty$ are the \mathbb{R} -valued transmitted signal and \mathbb{R}^+ -valued fade of sender- j , $j = 1, 2$, respectively, and $\{N_t\}_{t=1}^\infty$ is independent and identically distributed (i.i.d.) Gaussian noise with mean zero and variance σ_N^2 . The fading processes $\{\tilde{S}_{j,t}\}_{t=1}^\infty$, $j = 1, 2$, are assumed to be jointly stationary and ergodic, though not necessarily independent of each other. We further assume that the fading processes are independent of the additive noise process $\{N_t\}_{t=1}^\infty$. Using the notation $S_t \triangleq (\tilde{S}_{1,t}, \tilde{S}_{2,t})$ for the fade rvs at time t , a realization pair $s_t \triangleq (\tilde{s}_{1,t}, \tilde{s}_{2,t}) \in (\mathbb{R}^+)^2$, representing the fades experienced by the two senders at time t , will be called the state of the channel at time t , $t \geq 1$. The CSI available to sender- j is given by a mapping $h_j: (\mathbb{R}^+)^2 \rightarrow \mathcal{E}_j$, where \mathcal{E}_j can be an arbitrary subset of $(\mathbb{R}^+)^2$ which is not necessarily finite. The decoder is assumed to possess perfect CSI, i.e., $d_t = s_t$, $t \geq 1$. Sender- j is typically constrained in terms of its average power. Precisely, we shall say that sender- j is subject to an input power constraint \mathcal{P}_j if the codewords of sender- j satisfy

$$\frac{1}{n} \mathbb{E}_S [|\theta_j(m, h_j(S^n))|^2] \leq \mathcal{P}_j, \quad m \in \mathcal{M}_j \quad (9)$$

where \mathbb{E}_S denotes expectation with respect to the distribution of the state process $\{S_t\}_{t=1}^\infty$, and $\|\cdot\|$ denotes the Euclidean norm.

As a special case, the single-sender TVFC is specified by

$$Y_t = S_t x_t + N_t, \quad t \geq 1. \quad (10)$$

III. STATEMENT OF RESULTS

We begin this section with some additional definitions which will be needed to state our results. First, we note that for the TVMAC in (1), given any length- n block code $(\theta_1, \theta_2, \phi)$, the encoder θ_j , $j = 1, 2$, at time t , maps message m_j and the CSI e_j^t available to it into a channel input symbol $x_j(m_j, e_j^t) \in \mathcal{X}_j$. Therefore, using $\mathcal{F}_j(t)$ to denote the space of all mappings⁴ $\{f_{j,t}: \mathcal{E}_j^t \rightarrow \mathcal{X}_j\}$ and \mathcal{F}_j^n to denote the product space

$$\mathcal{F}_j(1) \times \cdots \times \mathcal{F}_j(n), \quad j = 1, 2$$

each encoder θ_j , $j = 1, 2$, can be represented in terms of mappings in \mathcal{F}_j^n ; namely, there exist equivalent length- n block encoders (θ_1', θ_2') , where

$$\theta_1': \mathcal{M}_1 \rightarrow \mathcal{F}_1^n \quad \text{and} \quad \theta_2': \mathcal{M}_2 \rightarrow \mathcal{F}_2^n. \quad (11)$$

Consequently, it will be convenient to deal with a derived TVMAC $\mathcal{F}_1^n \times \mathcal{F}_2^n \times \mathcal{S}^n \rightarrow \mathcal{Y}^n$, whose probability law is well-defined and is given by

$$P_{Y|F_1 F_2 S}(y^n | f_1^n, f_2^n, s^n) = W(y^n | x_1^n, x_2^n, s^n), \quad \text{where} \\ x_{j,t} = f_{j,t}(h_j(s^t)), \quad j = 1, 2, \quad \text{and} \quad t = 1, \dots, n. \quad (12)$$

For the special case of the single-sender TVC in (2), the analogous derived TVC $\mathcal{F}^n \times \mathcal{S}^n \rightarrow \mathcal{Y}^n$ conforms to the probability law

$$P_{Y|FS}(y^n | f^n, s^n) = W(y^n | x^n, s^n), \quad \text{where} \\ x_t = f_t(h(s^t)), \quad t = 1, \dots, n. \quad (13)$$

For $j = 1, 2$, let us denote by \underline{F}_j a sequence of finite-dimensional rvs $\{F_j^n = (F_{j,1}^{(n)}, F_{j,2}^{(n)}, \dots, F_{j,n}^{(n)})\}_{n=1}^\infty$ where each component $F_{j,t}^{(n)}$, $t = 1, \dots, n$, takes values in the finite set $\mathcal{F}_j(t)$. Unless otherwise mentioned, $F_{j,t}$ will denote an $\mathcal{F}_j(t)$ -valued rv and $f_{j,t}$ a realization of it. We do not require that the distribution of F_j^n satisfy any consistency conditions. In addition to denoting by \underline{F}_j the sequence of rvs as above, we shall often denote by \underline{F}_j the corresponding sequence of pmfs when the usage is clear from the context; here, \underline{F}_j represents an input sequence of finite-dimensional pmfs $\underline{F}_j = \{F_j^n\}_{n=1}^\infty$, $j = 1, 2$, for the TVMAC in (12). The corresponding output sequence of finite-dimensional pmfs induced by $\underline{F}_1, \underline{F}_2$ through the channel in (12) will be denoted by \underline{Y} . Similarly, $\underline{E}_j = \{E_j^n\}_{n=1}^\infty$ (resp., $\underline{D} = \{D^n\}_{n=1}^\infty$) is a sequence of rvs with $E_j^n = (E_{j,1}, \dots, E_{j,n})$ (resp., $D^n = (D_1, \dots, D_n)$), where each component $E_{j,t} = h_j(S_t)$ (resp., $D_t = g(S_t)$) takes values in the set \mathcal{E}_j (resp., \mathcal{D}); \underline{E}_j (resp., \underline{D}) will also be used to denote the corresponding sequences of pmfs.

The following notion of ‘‘liminf in probability,’’ is needed in order to state our results. This notion, along with its operational significance, originated in the work of Han and Verdú [20]. Precisely, given a sequence of \mathbb{R} -valued rvs $\{A_n\}_{n=1}^\infty$, its *liminf in probability*, $\underline{\lim-p} A_n$, is defined [20] as

$$\underline{\lim-p} A_n = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha: \lim_n \Pr(A_n < \alpha) = 0 \right\} \quad (14)$$

where \lim_n denotes $\lim_{n \rightarrow \infty}$. In the context of the TVMAC in (1), consider a sequence of rvs F_1^n, F_2^n, Y^n, D^n (with F_1^n, F_2^n, Y^n, D^n taking values in $\mathcal{F}_1^n, \mathcal{F}_2^n, \mathcal{Y}^n, \mathcal{D}^n$, respectively) generated according to a given sequence of joint pmfs $\underline{F}_1, \underline{F}_2, \underline{Y}, \underline{D}$. Denote by $i(F_1^n, F_2^n \wedge Y^n | D^n)$ the rv

$$\left(\log \frac{\Pr(Y^n | F_1^n, F_2^n, D^n)}{\Pr(Y^n | D^n)} \right).$$

The corresponding $\underline{\lim-p} \frac{1}{n} i(F_1^n, F_2^n \wedge Y^n | D^n)$ will be written as $\underline{I}(\underline{F}_1, \underline{F}_2 \wedge \underline{Y} | \underline{D})$. The quantities $i(F_1^n \wedge Y^n | F_2^n, D^n)$, $i(F_2^n \wedge Y^n | F_1^n, D^n)$, and $\underline{I}(\underline{F}_1 \wedge \underline{Y} | \underline{F}_2, \underline{D})$, $\underline{I}(\underline{F}_2 \wedge \underline{Y} | \underline{F}_1, \underline{D})$ are similarly defined.

The following proposition provides the capacity region \mathcal{C} of the TVMAC in (11) with CSI at the encoders and decoder. This result can be obtained by a straightforward modification of the

³The model in (8) considers only frequency nonselective fading effects [21].

⁴Although $f_{j,t}$ is defined in terms of all past and present CSI $e_j^t = (e_{j,1}, \dots, e_{j,t})$, it may depend only on a subset of these CSI symbols.

relevant treatment in Han [19].⁵ It will prove useful in establishing Theorems 2 and 3 below.

Proposition 1: The capacity region of the TVMAC in (1) with CSI at the encoders and decoder is given by

$$\mathcal{C} = \bigcup_{(\mathbf{E}_1, \mathbf{E}_2)} \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq \underline{I}(\mathbf{F}_1 \wedge \mathbf{Y} | \mathbf{E}_2, \mathbf{D}) \\ 0 \leq R_2 \leq \underline{I}(\mathbf{E}_2 \wedge \mathbf{Y} | \mathbf{E}_1, \mathbf{D}) \\ R_1 + R_2 \leq \underline{I}(\mathbf{F}_1, \mathbf{E}_2 \wedge \mathbf{Y} | \mathbf{D}) \end{array} \right\} \quad (15)$$

where the joint pmf of the rvs $F_1^n, F_2^n, S^n, Y^n, n \geq 1$, is given by

$$P_{F_1 F_2 S Y}(f_1^n, f_2^n, s^n, y^n) = P_{F_1}(f_1^n) P_{F_2}(f_2^n) P_S(s^n) P_{Y|F_1 F_2 S}(y^n | f_1^n, f_2^n, s^n) \quad (16)$$

with $P_{Y|F_1 F_2 S}(y^n | f_1^n, f_2^n, s^n)$ as in (12), whence the joint pmf of the rvs F_1^n, F_2^n, D^n, Y^n is given by

$$P_{F_1 F_2 D Y}(f_1^n, f_2^n, d^n, y^n) = \sum_{s^n \in \mathcal{S}^n} 1(g(s^n) = d^n) P_{F_1 F_2 S Y}(f_1^n, f_2^n, s^n, y^n) \quad (17)$$

where $1(\cdot)$ denotes the indicator function.

Remarks:

- A1. The region on the right-hand side of (15) is closed but need not be convex in general [19].
- A2. It readily follows that the capacity of the single-sender TVC in (2) with CSI at the encoder and decoder is

$$C = \sup_{\mathbf{F}} \underline{I}(\mathbf{F} \wedge \mathbf{Y} | \mathbf{D}) \quad (18)$$

where the joint pmf of the rvs $F^n, D^n, Y^n, n \geq 1$, is given by

$$P_{F D Y}(f^n, d^n, y^n) = P_F(f^n) \sum_{s^n \in \mathcal{S}^n} P_S(s^n) 1(g(s^n) = d^n) P_{Y|F S}(y^n | f^n, s^n) \quad (19)$$

where $P_{Y|F S}(y^n | f^n, s^n)$ is given by (13).

In Proposition 1, finding a “good” input distribution in (15) entails a search over the (abstract) space of map-valued rvs or “strategies.” On the other hand, we see below that if the CSI available to the encoder is contained in that available to the decoder, such a search is reduced to that over distributions of (familiar) channel input-valued rvs. From a practical standpoint, the latter situation—in contrast with the former—may provide useful insights concerning the correlation properties of “good” codes and CSI.

Proposition 2: For the TVMAC in (1) with CSI at the encoders and decoder, if there exist mappings ψ_1, ψ_2 such that

⁵Formally, the situation in which encoder CSI is additionally present, differs slightly from that in [19]. Now, each code “symbol” in a codeword can be regarded as a mapping from the available encoder CSI to the channel input alphabet. Note that the space of these mappings grows exponentially with available encoder CSI.

$e_{1,t} = \psi_1(d_t), e_{2,t} = \psi_2(d_t), t \geq 1$, the capacity region is given by

$$\mathcal{C} = \bigcup_{\mathbf{X}_1 | \mathbf{E}_1, \mathbf{X}_2 | \mathbf{E}_2} \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq \underline{I}(\mathbf{X}_1 \wedge \mathbf{Y} | \mathbf{X}_2, \mathbf{D}) \\ 0 \leq R_2 \leq \underline{I}(\mathbf{X}_2 \wedge \mathbf{Y} | \mathbf{X}_1, \mathbf{D}) \\ R_1 + R_2 \leq \underline{I}(\mathbf{X}_1, \mathbf{X}_2 \wedge \mathbf{Y} | \mathbf{D}) \end{array} \right\} \quad (20)$$

where the joint pmf of the rvs $X_1^n, X_2^n, D^n, Y^n, n \geq 1$, is given by

$$P_{X_1 X_2 D Y}(x_1^n, x_2^n, d^n, y^n) = \sum_{s^n \in \mathcal{S}^n, e_1^n \in \mathcal{E}_1^n, e_2^n \in \mathcal{E}_2^n} P_S(s^n) 1(g(s^n) = d^n) \times 1(\psi_1(d^n) = e_1^n, \psi_2(d^n) = e_2^n) \times P_{X_1 | E_1}(x_1^n | e_1^n) P_{X_2 | E_2}(x_2^n | e_2^n) W(y^n | x_1^n, x_2^n, s^n) \quad (21)$$

with the union in (20) being taken over all sequences of conditional pmfs of X_j^n given $E_j^n, j = 1, 2, n \geq 2$, satisfying

$$P_{X_j | E_j}(x_j^n | e_j^n) = P_{X_j | E_j}(x_{j,1} | e_{j,1}) \times \prod_{t=2}^n \Pr(X_{j,t} = x_{j,t} | X_j^{t-1} = x_j^{t-1}, E_j^t = e_j^t). \quad (22)$$

Remarks:

- B1. Remark A1 following Proposition 1 applies here as well.
- B2. A specialization of Proposition 1 to the single-sender TVC in (2) is obvious.
- B3. The heuristics behind Proposition 2, as a consequence of Proposition 1, are as follows. For each $n \geq 1$, D^n determines E^n , so that F_j^n and D^n together determine $X_j^n, j = 1, 2$. It is then to be expected that the conditional pmf of Y^n given F_1^n, F_2^n, D^n (resp., F_j^n, D^n) coincides, with probability 1, with the conditional pmf of Y^n given X_1^n, X_2^n, D^n (resp., X_j^n, D^n). This results in the reduction of (15) to (20).

Further simplifications of the capacity region of the TVMAC in (1) with CSI at the encoders and decoder obtain under the following two assumptions. First, we assume that the channel law (1) satisfies the “memorylessness” property

$$W(y^n | x_1^n, x_2^n, s^n) = \prod_{t=1}^n W(y_t | x_{1,t}, x_{2,t}, s_t), \quad n \geq 1. \quad (23)$$

Now, the TVMAC can be described in terms of a family $\mathcal{W} = \{W(\cdot | \cdot, \cdot, s), s \in \mathcal{S}\}$ of channels $\mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}$, together with the probability law P_S of the state process $\{S_t\}_{t=1}^\infty$; the state $s \in \mathcal{S}$ can be interpreted as an index identifying a particular $W \in \mathcal{W}$. Second, assume that the state process $\{S_t\}_{t=1}^\infty$ is a time-invariant, indecomposable, aperiodic Markov chain (TIAMC), with initial pmf denoted by P_{S_1} (i.e., the pmf of the rv S_1), transition probability matrix Γ , and stationary pmf π_S . Thus, $\Gamma(s|s')$ denotes the (one-step) conditional probability of a state transition from s' to $s, s', s \in \mathcal{S}$. Under these two assumptions, the TVMAC in (23) is tantamount to a “finite-state” multiple-access channel (cf., e.g., [6]). It should be noted that

the channel between the encoders and the receiver will, in general, possess “memory” which is induced by the state process; the assumed “memorylessness” property in (23) involves conditioning on the prevailing state sequence. Its capacity region, in general, depends on the initial pmf P_{S_1} , and will be denoted by $\mathcal{C}_{P_{S_1}}$. In some special situations, however, the capacity region is invariant with respect to P_{S_1} . In order to describe one such situation we introduce the following.

Definition: Given nonempty subsets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{S}$, we say that \mathcal{A} connects to \mathcal{B} , denoted $\mathcal{A} \rightarrow \mathcal{B}$, if for every $s \in \mathcal{B}$ there exists $s' \in \mathcal{A}$ such that $\Gamma(s|s') > 0$.

For convenience, we denote

$$\mathcal{S}_{j,e} = \{s \in \mathcal{S}: h_j(s) = e\}, \quad e \in \mathcal{E}_j, \quad j = 1, 2. \quad (24)$$

Further, for any pmf P on \mathcal{S} denote by σ_P the support of P , i.e., $\sigma_P = \{s \in \mathcal{S}: P(s) > 0\}$.

Theorem 1: Consider the memoryless TVMAC in (23) with CSI at the encoders and decoder, when the state process $\{S_t\}_{t=1}^\infty$ is a TIAMC.

- 1) For every P_{S_1} , it holds that

$$\mathcal{C}_{P_{S_1}} \supseteq \mathcal{C}_{\pi_S} \quad (25)$$

where the inclusion can be strict.

- 2) The capacity region $\mathcal{C}_{P_{S_1}}$ is the same for all initial pmfs P_{S_1} which are restricted to be point masses on σ_{π_S} .
- 3) If there exists $\bar{s} \in \sigma_{\pi_S}$ satisfying the condition

$$\{\bar{s}\} \rightarrow \sigma_{\pi_S} \quad (26)$$

then, $\mathcal{C}_{P_{S_1}} = \mathcal{C}_{\pi_S}$ for every P_{S_1} .

Corollary 1: Under the hypothesis of Theorem 1 the following are true.

- 1) The assertion in Theorem 1 part 3 also holds if (26) is replaced by the following condition. There exist $\bar{s} \in \sigma_{\pi_S}$ and sets $\mathcal{A}_l \subseteq \mathcal{S}_{1,e_{1,l}} \cap \mathcal{S}_{2,e_{2,l}}$, for some $e_{j,l} \in \mathcal{E}_j$, $j = 1, 2$, and $l = 1, \dots, L$, for some finite integer L , such that

$$\{\bar{s}\} \rightarrow \mathcal{A}_1 \rightarrow \dots \rightarrow \mathcal{A}_L \rightarrow \sigma_{\pi_S}. \quad (27)$$

- 2) If the initial state is known to both transmitters, the capacity region is the same for all initial pmfs P_{S_1} whose support is contained in σ_{π_S} .

Remarks:

- C1. That the inclusion in (25) can be strict is illustrated by Example 1 below.
- C2. The hypothesis in (27) is weaker than that in (26).

In Theorem 1 part 1, the inclusion in (25) is not a surprising result. Loosely speaking, the proof entails showing that a code which achieves any rate pair \mathcal{C}_{π_S} can also be used to achieve the same rate pair for any initial pmf P_{S_1} simply by waiting (finitely) long enough for the behavior of the state process to approach stationarity. What is, however, surprising is that the inclusion in (25) can be *strict* for certain initial pmfs P_{S_1} ; see Example 1 below. This behavior is in contrast with that of a similar single-sender (indecomposable) FSC with no CSI at the encoder or decoder [6], [7], with perfect CSI at the encoder alone [9], or with perfect CSI at the encoder and decoder [6], [15].

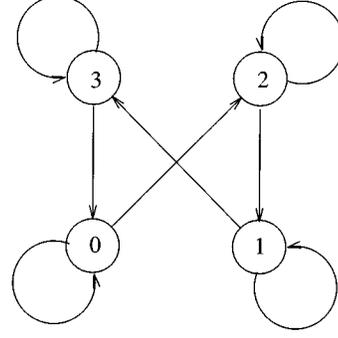


Fig. 2. Transition diagram for the state process $\{S_t\}_{t=1}^\infty$ in Example 1.

An explanation for a strict inclusion in (25) is as follows. As will be seen in Example 1 below, for certain initial pmfs, the transmitters and receiver can infer, in conjunction with the CSI, additional detailed information concerning the evolution of the state process, than when the initial pmf is the stationary one. For instance, Example 1 considers an extreme case where for a specific choice of P_{S_1} , the CSI leads to a *perfect* knowledge of the state process.

However, the sufficient “mixing” condition (26) limits, in effect, the amount of additional inferred information concerning the state process beyond that provided by π_S ; consequently, under (26), $\mathcal{C}_{P_{S_1}} = \mathcal{C}_{\pi_S}$ for every P_{S_1} .

Example 1: Consider the single-sender memoryless TVC in (2) characterized by

$$W(y^n|x^n, s^n) = \prod_{t=1}^n W(y_t|x_t, s_t), \quad n \geq 1 \quad (28)$$

with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $\mathcal{S} = \{0, 1, 2, 3\}$, and

$$\begin{aligned} W(0|0, 0) &= 0.9 & W(1|1, 0) &= 1 \\ W(0|0, 1) &= 1 & W(1|1, 1) &= 0.9 \\ W(0|0, 2) &= 0.5 & W(1|1, 2) &= 1 \\ W(0|0, 3) &= 1 & W(1|1, 3) &= 0.5. \end{aligned}$$

Let the state process $\{S_t\}_{t=1}^\infty$ be a TIAMC with transition diagram given in Fig. 2 and

$$\Gamma(0|0) = \Gamma(1|1) = \Gamma(2|2) = \Gamma(3|3) = 0.5.$$

The encoder and decoder CSI are given in terms of

$$\begin{aligned} g(0) &= g(1) = 0 & g(2) &= g(3) = 1 \\ h(0) &= h(1) = 0 & h(2) &= h(3) = 1. \end{aligned}$$

In specializing Theorem 1 part 1 to this case, we show that it can hold, for some P_{S_1} , that $\mathcal{C}_{P_{S_1}} > \mathcal{C}_{\pi_S}$. Consider first the situation when the initial pmf $P_{S_1} = \bar{P}$ places a point mass at state 0, i.e., $\bar{P}(0) = 1$. Clearly, knowledge of the initial state and e^t (resp., d^t) provides the encoder (resp., decoder) with perfect CSI. The corresponding capacity $\mathcal{C}_{\bar{P}}$ is well known (cf., e.g., [14, p. 2011]), and is given by

$$\mathcal{C}_{\bar{P}} = \sum_{s \in \mathcal{S}} \pi_S(s) \max_{X|S=s} I(X \wedge Y|S=s)$$

which is seen to equal 0.376 bit per channel use. Next, observe that when the initial pmf $P_{S_1} = \pi_S$, the capacity \mathcal{C}_{π_S} does not exceed the capacity $\bar{\mathcal{C}}_{\pi_S}$ for the situation where $P_{S_1} = \pi_S$ and, additionally, $d_t = s_t$, $t \geq 1$, i.e., with *perfect* CSI at the decoder.

In the latter situation, the hypothesis of [14, Proposition 2] is met, and by [14]

$$\bar{C}_{\pi_S} = \max_{X|E} I(X \wedge Y|S, E) = 0.371 \text{ bit per channel use}$$

so that

$$C_{\pi_S} \leq \bar{C}_{\pi_S} < C_{\bar{P}}. \quad \square$$

The capacity region of the TVMAC in (23), when the state process $\{S_t\}_{t=1}^{\infty}$ is a TIAMC, is, of course, given by Proposition 1. However, this region has a simpler characterization when encoder restrictions are imposed, or when the initial pmf $P_{S_1} = \pi_S$, and is stated in the following.

Theorem 2: Consider the memoryless TVMAC in (23) with CSI at the encoders and decoder, when the state process $\{S_t\}_{t=1}^{\infty}$ is a TIAMC with initial pmf P_{S_1} .

- 1) Assume an encoder restriction of window- k for each sender, where k is a fixed positive integer. Then, the capacity region $\mathcal{C}(k)$ does not depend on P_{S_1} , and is given by

$$\mathcal{C}(k) = \bigcup_{\underline{E}_1, \underline{E}_2} \left\{ (R_1, R_2): \begin{cases} 0 \leq R_1 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n) \\ 0 \leq R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n) \\ R_1 + R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n) \end{cases} \right\} \quad (29)$$

where $F_{j,t}$ is a $\mathcal{F}_j(k)$ -valued rv for $t \geq k$, $j = 1, 2$, and the joint pmf of F_1^n, F_2^n, D^n, Y^n is given by (16) and (17), with (12) now replaced by

$$P_{Y|F_1 F_2 S}(y^n | f_1^n, f_2^n, s^n) = W(y^n | x_1^n, x_2^n, s^n), \quad \text{where } x_{j,t} = f_{j,t} \left(h_j \left(s_{\max\{1, t-k+1\}}^t \right) \right), \quad j = 1, 2, \text{ and } t = 1, \dots, n. \quad (30)$$

- 2) If, however, no such restriction is imposed on the encoders, the capacity region $\mathcal{C}_{P_{S_1}}$ satisfies the inclusion⁶

$$\mathcal{C}_{P_{S_1}} \supseteq \lim_k \mathcal{C}(k) \quad (31)$$

for every P_{S_1} . In particular, when $P_{S_1} = \pi_S$, the capacity region \mathcal{C}_{π_S} is given by

$$\begin{aligned} \mathcal{C}_{\pi_S} &= \lim_k \mathcal{C}(k) \quad (32) \\ &= \bigcup_{\underline{E}_1, \underline{E}_2} \left\{ (R_1, R_2): \begin{cases} 0 \leq R_1 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n) \\ 0 \leq R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n) \\ R_1 + R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n) \end{cases} \right\} \quad (33) \end{aligned}$$

⁶Clearly, $\mathcal{C}(k) \subseteq \mathcal{C}(k+1)$ and $\lim_k \mathcal{C}(k) = \bigcup_k \mathcal{C}(k)$.

where the joint pmf of the rvs F_1^n, F_2^n, D^n, Y^n is given by (16) and (17) with $P_{S_1} = \pi_S$.⁷

Remarks:

- D1. The proof of Theorem 2 part 1 will show that the capacity region $\mathcal{C}(k)$ in (29) can, in fact, be achieved by input sequences of mappings which constitute suitable hidden Markov processes (cf. [7]).
- D2. The inclusion in (13) can be strict. This is seen from (32) and Theorem 1.
- D3. Under the additional conditions (26) or (27), it follows from Theorem 1 part 3 that $\mathcal{C}_{P_{S_1}}$ is the same for all initial pmfs P_{S_1} , and is given by (32) or (33), the latter evaluated with $P_{S_1} = \pi_S$.
- D4. By Theorem 2 part 1, the capacity region $\mathcal{C}(k)$ does not depend on P_{S_1} . Hence, upon taking $P_{S_1} = \pi_S$, without loss of generality, the state process $\{S_t\}_{t=1}^{\infty}$ is rendered stationary. Consequently, by time sharing for the TVMAC in (23), convex combinations of given achievable rate pairs are also achievable. Hence the capacity region $\mathcal{C}(k)$ is a closed, convex set. Further, since $\mathcal{C}(k) \subseteq \mathcal{C}(k+1)$, $\lim_k \mathcal{C}(k)$ is also closed and convex. Thus, by (32), \mathcal{C}_{π_S} is a closed and convex set.
- D5. While $\mathcal{C}_{P_{S_1}}$ is closed, it is not clear if it is convex.
- D6. The capacity regions in (29) and (33) simplify if the CSI available to each encoder is contained in that available to the decoder, in a manner akin to Proposition 2. For example, under the hypothesis of Theorem 2 part 1, suppose further that there exist mappings ψ_1, ψ_2 , such that $e_{1,t} = \psi_1(d_t)$ and $e_{2,t} = \psi_2(d_t)$, $t \geq 1$. The capacity region $\mathcal{C}(k)$ in (29) then simplifies as

$$\mathcal{C}(k) = \bigcup_{\underline{X}_1 | \underline{E}_1, \underline{X}_2 | \underline{E}_2} \left\{ (R_1, R_2): \begin{cases} 0 \leq R_1 \leq \underline{\lim}_n \frac{1}{n} I(X_1^n \wedge Y^n | X_2^n, D^n) \\ 0 \leq R_2 \leq \underline{\lim}_n \frac{1}{n} I(X_2^n \wedge Y^n | X_1^n, D^n) \\ R_1 + R_2 \leq \underline{\lim}_n \frac{1}{n} I(X_1^n, X_2^n \wedge Y^n | D^n) \end{cases} \right\} \quad (34)$$

where the joint pmf of the rvs X_1^n, X_2^n, D^n, Y^n is given by (21), and the union in (34) is taken over all sequences of conditional pmfs of X_j^n given E_j^n , $j = 1, 2$, satisfying

$$\begin{aligned} P_{X_j | E_j}(x_j^n | e_j^n) &= P_{X_j | E_j}(x_{j,1} | e_{j,1}) \\ &\times \prod_{t=2}^n \Pr \left(X_{j,t} = x_{j,t} \mid X_j^{t-1} = x_j^{t-1}, \right. \\ &\quad \left. E_{j, \max\{1, t-k+1\}}^t = e_{j, \max\{1, t-k+1\}}^t \right) \quad (35) \end{aligned}$$

for each $n \geq 2$. Furthermore, the counterpart of Theorem 2 part 2 is now obvious, and will not be repeated here.

⁷Although the expressions in (29) and (33) appear to be identical at first blush, they do differ in general since $F_{j,t}$ is $\mathcal{F}_j(k)$ -valued for $t \geq k$ in (29), and $\mathcal{F}_j(t)$ -valued in (33).

For the single-sender memoryless TVC in (28) with CSI at the encoder and decoder, when $\{S_t\}_{t=1}^\infty$ is a TIAMC with initial pmf P_{S_1} , the specializations of Theorems 1 and 2, along with those of the associated remarks, are obvious.

We next turn to another special case of Proposition 1, in which by allowing the receiver to fully observe the state process, we obtain a simpler expression for the capacity region of the TVMAC than in Proposition 1 or Theorem 2. Specifically, consider the memoryless TVMAC in (23) with limited CSI available to each encoder and *perfect* CSI available to the decoder, i.e., $d_t = s_t$, $t \geq 1$. We present below the capacity region of the TVMAC in (23) with and without an encoder restriction of window- k , when the state process $\{S_t\}_{t=1}^\infty$ is additionally assumed to be stationary and ergodic (but is not necessarily a TIAMC).

Theorem 3: Assume an encoder restriction of window- k for each sender, where k is a fixed positive integer, and assume that the decoder has perfect CSI. Then, the capacity region $\mathbf{C}(k)$ of the memoryless TVMAC in (23), when the state process $\{S_t\}_{t=1}^\infty$ is stationary and ergodic, is the convex closure of the region

$$\bigcup_{\substack{X_{1,k}|E_1^k, X_{2,k}|E_2^k \\ X_{1,k}, X_{2,k} \rightarrow E_1^k E_2^k \rightarrow S^k}} \left\{ \begin{array}{l} 0 \leq R_1 \leq I(X_{1,k} \wedge Y_k | X_{2,k}, S^k), \\ (R_1, R_2): 0 \leq R_2 \leq I(X_{2,k} \wedge Y_k | X_{1,k}, S^k), \\ R_1 + R_2 \leq I(X_{1,k}, X_{2,k} \wedge Y_k | S^k) \end{array} \right\} \quad (36)$$

where the joint pmf of $X_{1,k}, X_{2,k}, S^k, Y_k$ is given by

$$P_{X_{1,k} X_{2,k} S^k Y_k}(x_{1,k}, x_{2,k}, s^k, y_k) = P_S(s^k) W(y_k | x_{1,k}, x_{2,k}, s_k) \times \prod_{j=1}^2 \left[\sum_{e_j^k \in \mathcal{E}_j^k} 1(h_j(s^k) = e_j^k) P_{X_j|E_j}(x_{j,k} | e_j^k) \right]. \quad (37)$$

If, however, no such restriction is imposed on the encoders, i.e., encoder θ_j , $j = 1, 2$, at time t is allowed to depend on the entire CSI e_j^t available to it, the capacity region \mathbf{C} of the TVMAC in (23) is given by

$$\mathbf{C} = \lim_k \mathbf{C}(k) = \sup_k \mathbf{C}(k). \quad (38)$$

Remarks:

- E1. The assertions of Theorem 3 clearly hold in the special case when $\{S_t\}_{t=1}^\infty$ is a TIAMC.
- E2. The specialization of Theorem 3 to the single-sender memoryless TVC in (28) is immediate.

An interesting consequence of Theorem 3 for the single-sender memoryless TVC in (28), observed earlier in [14], and proved below in a different manner, is the following.

Corollary 2: Let the state process $\{S_t\}_{t=1}^\infty$ be stationary and ergodic. Assume that $\{S_t\}_{t=1}^\infty$ and the mapping $h: \mathcal{S} \rightarrow \mathcal{E}$ is such that

$$S_t \rightarrow E_{t-l+1}^t \rightarrow E^{t-l}, \quad t > l \quad (39)$$

for a fixed integer $l \geq 1$. Then the capacity of the single-sender memoryless TVC in (28) is given by

$$C = C(l) = \max_{X_l|E^l: X_l \rightarrow E^l \rightarrow S^l} I(X_l \wedge Y_l | S^l) \quad (40)$$

where the joint pmf of X_l, S^l, Y_l is given by

$$P_{XSY}(x_l, s^l, y_l) = P_S(s^l) W(y_l | x_l, s_l) \sum_{e^l \in \mathcal{E}^l} 1(h(s^l) = e^l) P_{X|E}(x_l | e^l). \quad (41)$$

Remarks:

- F1. The condition in (39) is satisfied, for instance, when $\{S_t\}_{t=1}^\infty$ is a Markov process and the encoder is provided a delayed but perfect estimate of the channel state (cf. [13]).
- F2. The analog of Corollary 2 holds for the TVMAC in (23) under the seemingly excessive (sufficient) conditions: for $t > l$

$$S_t \rightarrow E_{1,t-l+1}^t E_{2,t-l+1}^t \rightarrow E_1^{t-l} E_2^{t-l} \\ S_t \rightarrow E_i^{t-l+1} \rightarrow E_j^{t-l}, \quad i, j = 1, 2. \quad (42)$$

Our final concern is with the multiple-access fading channel (TVFC) model in (8). While a capacity result akin to Proposition 1 can, in general, be obtained, its usefulness is limited since the expressions therein do not lend themselves to any significant simplification, and, hence, do not lead to any useful insights. In fact, this is already true in the case of the single-sender version of the fading channel in (8), for which the capacity result is analogous to (18). Therefore, two useful simplifying assumptions will hereafter be made for the TVFC in (8). First, we shall assume that the state process(es) are jointly stationary and ergodic. Secondly, we shall assume that *perfect* CSI is available to the decoder; however, the encoders need not have the same benefits.⁸

Our main result for the TVFC in (8) is presented below. An analogous result for the single-sender TVFC in (10)—but with different assumptions concerning the encoder and decoder CSI—has appeared recently in [14]. It will be convenient to set

$$\zeta(a, b) \triangleq \frac{1}{2} \log \left(1 + \frac{a}{b} \right), \quad a, b \geq 0. \quad (43)$$

Theorem 4: Consider the TVFC in (8), and let the state process

$$\{S_t\}_{t=1}^\infty = \{(\tilde{S}_1, t, \tilde{S}_2, t)\}_{t=1}^\infty$$

be stationary and ergodic. Further, assume that sender- j is subject to an input power constraint \mathcal{P}_j , $j = 1, 2$, as in (9).

- 1) Assume an encoder restriction of window- k for each sender and perfect CSI at the decoder. Then, the capacity region $\mathbf{C}(k, \mathcal{P}_1, \mathcal{P}_2)$ is given by the closure of the

⁸For communication on the uplink of a wireless channel, the receiver at the base station commands measurement and computational resources superior to those of the (mobile) transmitters.

region shown in (44) at the bottom of the page, where $\mu_j: \mathcal{E}_j^k \rightarrow \mathbb{R}^+$ denotes a power allocation policy for sender- j subject to the constraint

$$\mathbb{E}_S[\mu_j(E_j^k)] \leq \mathcal{P}_j, \quad j = 1, 2. \quad (45)$$

- 2) If no such restriction is imposed on the encoders, i.e., encoder θ_j , $j = 1, 2$, is allowed to depend at time t on the entire CSI e_j^t available to it, the capacity region $\mathcal{C}(\mathcal{P}_1, \mathcal{P}_2)$ is given by

$$\mathcal{C}(\mathcal{P}_1, \mathcal{P}_2) = \lim_k \mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2) = \sup_k \mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2). \quad (46)$$

Remarks:

- G1. The capacity regions $\mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2)$ and $\mathcal{C}(\mathcal{P}_1, \mathcal{P}_2)$ given by (44) and (46), respectively, are convex by dint of the assumed stationarity and ergodicity of the state process $\{S_t\}_{t=1}^\infty$. The reason is akin to that in Remark D4 following Theorem 2.
- G2. A specialization of Theorem 4 to the case of a single-sender TVFC in (10) is immediate.

The following consequence of Theorem 4 is analogous to Corollary 2 above (cf. also Remark F2 following Corollary 2).

Corollary 3: Consider the single-sender TVFC in (10), with an input power constraint \mathcal{P} as in (9). Let the state process $\{S_t\}_{t=1}^\infty$ be stationary and ergodic. Assume further that

$$S_t \text{---} E_{t-l+1}^t \text{---} E_{t-l}^{t-l} \quad t > l \quad (47)$$

for some fixed integer $l \geq 1$. Then the capacity is given by

$$C(\mathcal{P}) = C(l, \mathcal{P}) = \sup_\mu \mathbb{E}_S \left[\zeta \left(\mu(E^l) | S_l|^2, \sigma_N^2 \right) \right] \quad (48)$$

where $\mu: \mathcal{E}^l \rightarrow \mathbb{R}^+$ denotes a power allocation policy subject to the constraint

$$\mathbb{E}_S[\mu(E^l)] \leq \mathcal{P}. \quad (49)$$

IV. PROOFS OF RESULTS

Our first concern is with the TVMAC in (1) for which we shall prove the capacity results in Proposition 1, 2, and Theorems 1 through 3.

The proof of Proposition 1, as indicated in Section III, follows from Han [19]. Specifically, the forward part is based on a random coding argument *a la* [19]. Similarly, the proof of the converse part relies on Lemma 1 below, obtained as a simple

modification of [19, Lemma 3], and stated next without proof. Given a length- n block code $(\theta_1, \theta_2, \phi)$ for the TVMAC in (1) with CSI at the encoders and decoder, the output of encoder θ_j , $j = 1, 2$, corresponding to message m_j at time t is $x_j(m_j, e_j^t)$, where e_j^t is the CSI available to encoder θ_j ; as noted earlier, an equivalent representation of the codeword corresponding to message m_j is obtained in terms of mappings in \mathcal{F}_j^n . Let \overline{M}_1 and \overline{M}_2 denote independent message rvs which are uniformly distributed on the message sets $\mathcal{M}_1 = \{1, \dots, M_1\}$ and $\mathcal{M}_2 = \{1, \dots, M_2\}$ of sender-1 and sender-2, respectively. Also, assume that \overline{M}_1 and \overline{M}_2 are independent of the state process $\{S_t\}_{t=1}^\infty$. The corresponding pmf F_j^n on \mathcal{F}_j^n places probability mass $1/M_j$ on each of the codewords $(c_{j,1}^n, \dots, c_{j,M_j}^n)$, where $c_{m_j}^n \in \mathcal{F}_j^n$ is the codeword corresponding to message m_j , $m_j = 1, \dots, M_j$, of sender- j , $j = 1, 2$.

Lemma 1: Let $(\theta_1, \theta_2, \phi)$ be a length- n block code with message sets $\mathcal{M}_1, \mathcal{M}_2$, for the TVMAC in (1) with CSI at the encoders and decoder, and (average) probability of error $P_e(\theta_1, \theta_2, \phi) < \epsilon$, $0 < \epsilon < 1$. Then, for every $\gamma > 0$

$$\begin{aligned} \Pr \left(\frac{\log M_1}{n} \geq \frac{1}{n} i(F_1^n \wedge Y^n | F_2^n, D^n) + \gamma \right) &\leq \epsilon + 2^{-n\gamma} \\ \Pr \left(\frac{\log M_2}{n} \geq \frac{1}{n} i(F_2^n \wedge Y^n | F_1^n, D^n) + \gamma \right) &\leq \epsilon + 2^{-n\gamma} \\ \Pr \left(\frac{\log M_1 M_2}{n} \geq \frac{1}{n} i(F_1^n, F_2^n \wedge Y^n | D^n) + \gamma \right) &\leq \epsilon + 2^{-n\gamma} \end{aligned} \quad (50)$$

where F_j^n places probability mass $1/M_j$ on each of the codewords for sender- j , and the joint pmf of F_1^n, F_2^n, Y^n, D^n is given by (16) and (17).

Before proceeding to the proof of Proposition 2, we present a technical lemma which will be used in the proof.

Lemma 2: For rvs $F_1^n, F_2^n, E_1^n, E_2^n, X_1^n, X_2^n, D^n, Y^n$, where $X_j^n = F_j^n(E_j^n)$, $j = 1, 2$, and with joint pmf determined by (16), (17), and (12), the following Markov relationships hold:

$$Y^n \text{---} X_1^n X_2^n E_1^n E_2^n D^n \text{---} F_1^n F_2^n \quad (51)$$

$$Y^n \text{---} X_j^n E_j^n D^n \text{---} F_j^n, \quad j = 1, 2. \quad (52)$$

Proof: See Appendix A. \square

Proof of Proposition 2: The proof entails a reduction of (15) and (16) to (20)–(22). It will be convenient to paraphrase the hypothesis, with an abuse of notation, by writing

$$E_j^n = \psi_j(D^n), \quad j = 1, 2. \quad (53)$$

$$\bigcup_{(\mu_1, \mu_2)} \left\{ (R_1, R_2): \begin{array}{l} 0 \leq R_1 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(E_1^k) | \tilde{S}_{1,k}^2, \sigma_N^2 \right) \right] \\ 0 \leq R_2 \leq \mathbb{E}_S \left[\zeta \left(\mu_2(E_2^k) | \tilde{S}_{2,k}^2, \sigma_N^2 \right) \right] \\ R_1 + R_2 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(E_1^k) | \tilde{S}_{1,k}^2 + \mu_2(E_2^k) | \tilde{S}_{2,k}^2, \sigma_N^2 \right) \right] \end{array} \right\} \quad (44)$$

We first show the equality of the expressions in (15) and (20) involving liminf in probability. Observe that (51) and (52), by virtue of (53), yield the following Markov relationships:

$$Y^n \text{---} X_1^n X_2^n D^n \text{---} F_1^n F_2^n \quad (54)$$

$$Y^n \text{---} X_j^n D^n \text{---} F_j^n, \quad j = 1, 2. \quad (55)$$

On the other hand, since $X_j^n = F_j^n(\psi_j(D^n))$, $j = 1, 2$, the following Markov relationships hold trivially:

$$Y^n \text{---} F_1^n F_2^n D^n \text{---} X_1^n X_2^n \quad (56)$$

$$Y^n \text{---} F_j^n D^n \text{---} X_j^n, \quad j = 1, 2. \quad (57)$$

From (54), (55) and (56), (57), we see that the following hold with probability 1:

$$\Pr(Y^n | X_1^n, X_2^n, D^n) = \Pr(Y^n | F_1^n, F_2^n, D^n) \quad (58)$$

$$\Pr(Y^n | X_j^n, D^n) = \Pr(Y^n | F_j^n, D^n), \quad j = 1, 2. \quad (59)$$

By virtue of (58) and (59), the definition of liminf in probability yields

$$\underline{I}(\underline{X}_1 \wedge \underline{Y} | \underline{X}_2, \underline{D}) = \underline{I}(\underline{F}_1 \wedge \underline{Y} | \underline{F}_2, \underline{D}) \quad (60)$$

$$\underline{I}(\underline{X}_2 \wedge \underline{Y} | \underline{X}_1, \underline{D}) = \underline{I}(\underline{F}_2 \wedge \underline{Y} | \underline{F}_1, \underline{D}) \quad (61)$$

$$\underline{I}(\underline{X}_1, \underline{X}_2 \wedge \underline{Y} | \underline{D}) = \underline{I}(\underline{F}_1, \underline{F}_2 \wedge \underline{Y} | \underline{D}). \quad (62)$$

Next, we show that (16) and (17), combined with the hypothesis, lead to (21) and (22). Observe that

$$\begin{aligned} & P_{X_1 X_2 D Y}(x_1^n, x_2^n, d^n, y^n) \\ &= \sum_{f_1^n \in \mathcal{F}_1^n, f_2^n \in \mathcal{F}_2^n, s^n \in \mathcal{S}_n} \\ & P_{Y | X_1 X_2 F_1 F_2 S D}(y^n | x_1^n, x_2^n, f_1^n, f_2^n, s^n, d^n) \\ & \times P_{X_1 X_2 F_1 F_2 S D}(x_1^n, x_2^n, f_1^n, f_2^n, s^n, d^n). \end{aligned} \quad (63)$$

From (12) and (15), it readily follows that

$$Y^n \text{---} X_1^n X_2^n S^n \text{---} F_1^n F_2^n \quad (64)$$

which, since $D^n = g(S^n)$, leads to

$$Y^n \text{---} X_1^n X_2^n S^n \text{---} F_1^n F_2^n D^n \quad (65)$$

so that the first factor in the sum in (63) is

$$\begin{aligned} & P_{Y | X_1 X_2 F_1 F_2 S D}(y^n | x_1^n, x_2^n, f_1^n, f_2^n, s^n, d^n) \\ &= W(y^n | x_1^n, x_2^n, s^n). \end{aligned} \quad (66)$$

Further, the second factor in the sum in (63) is

$$\begin{aligned} & P_{X_1 X_2 F_1 F_2 S D}(x_1^n, x_2^n, f_1^n, f_2^n, s^n, d^n) \\ &= \sum_{e_1^n \in \mathcal{E}_1^n, e_2^n \in \mathcal{E}_2^n} P_{F_1 F_2 S E_1 E_2 D}(f_1^n, f_2^n, s^n, e_1^n, e_2^n, d^n) \\ & \times P_{X_1 X_2 | F_1 F_2 S E_1 E_2 D}(x_1^n, x_2^n | f_1^n, f_2^n, s^n, e_1^n, e_2^n, d^n). \end{aligned} \quad (67)$$

Since $X_j^n = F_j^n(E_j^n)$, $j = 1, 2$, the second probability in the right-hand side of (67) equals

$$1(f_1^n(e_1^n) = x_1^n)1(f_2^n(e_2^n) = x_2^n)$$

and the first probability, by (16) and (35), equals

$$\begin{aligned} & P_{F_1}(f_1^n)P_{F_2}(f_2^n)P_S(s^n) \\ & \times 1(\psi_1(d^n) = e_1^n)1(\psi_2(d^n) = e_2^n)1(g(s^n) = d^n) \end{aligned}$$

so that (67) now becomes

$$\begin{aligned} & P_{X_1 X_2 F_1 F_2 S D}(x_1^n, x_2^n, f_1^n, f_2^n, s^n, d^n) \\ &= \sum_{e_1^n \in \mathcal{E}_1^n, e_2^n \in \mathcal{E}_2^n} 1(f_1^n(e_1^n) = x_1^n)1(f_2^n(e_2^n) = x_2^n) \\ & \times P_{F_1}(f_1^n)P_{F_2}(f_2^n)P_S(s^n)1(\psi_1(d^n) = e_1^n) \\ & \times 1(\psi_2(d^n) = e_2^n)1(g(s^n) = d^n). \end{aligned} \quad (68)$$

Upon substituting (66) and (68) in (63), and noting that for $j = 1, 2$

$$\sum_{f_j^n \in \mathcal{F}_j^n} 1(f_j^n(e_j^n) = x_j^n)P_{F_j}(f_j^n) = P_{X_j | E_j}(x_j^n | e_j^n) \quad (69)$$

we obtain (21).

Finally, observe that for $j = 1, 2$, and $t < n$

$$\begin{aligned} & P_{X_j | E_j}(x_j^t | e_j^t) = \sum_{f_j^t \in \mathcal{F}_j^t} P_{F_j X_j | E_j}(f_j^t, x_j^t | e_j^t) \\ &= \sum_{f_j^t \in \mathcal{F}_j^t} P_{F_j}(f_j^t)1(f_j^t(e_j^t) = x_j^t), \quad \text{by (16)} \\ &= P_{X_j | E_j}(x_j^t | e_j^t), \end{aligned} \quad (70)$$

where the last equality is by (69). A straightforward calculation using (70) now shows that the conditional pmf of X_j^n given E_j^n , $j = 1, 2$, satisfies (22), thereby completing the proof of Proposition 2. \square

We next state, as Lemma 3, a weaker converse than that which follows from Proposition 1 above for the capacity region of the TVMAC in (1). The significance of the lemma will be clear in the sequel. We add that Lemma 3 can also be obtained directly from Fano's inequality in a standard manner.

Lemma 3: The capacity region \mathcal{C} of the TVMAC (1) with CSI at the encoders and decoder satisfies the inclusion

$$\mathcal{C} \subseteq \bigcup_{\underline{F}_1, \underline{F}_2} \left\{ \begin{array}{l} 0 \leq R_1 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n) \\ (R_1, R_2): 0 \leq R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n) \\ R_1 + R_2 \leq \underline{\lim}_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n) \end{array} \right\} \quad (71)$$

where the joint pmf of the rvs F_1^n, F_2^n, D^n, Y^n is given by (16) and (17).

Remark: The region on the right-hand side of (71) need not be closed in general.

Proof: Given a sequence of joint pmfs $\underline{F}_1, \underline{F}_2, \underline{D}, \underline{Y}$, the inequalities below follow directly from the definition of liminf in probability (see also [20, Theorem 8(h)])

$$\underline{I}(\underline{F}_1 \wedge \underline{Y} | \underline{F}_2, \underline{D}) \leq \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n)$$

$$\underline{I}(\underline{\mathbf{F}}_2 \wedge \underline{\mathbf{Y}}|\underline{\mathbf{F}}_1, \underline{\mathbf{D}}) \leq \liminf_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n)$$

and

$$\underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{Y}}|\underline{\mathbf{D}}) \leq \liminf_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n). \quad (72)$$

Lemma 3 is now immediate from (15) and (72). \square

Turning next to the proof of Theorem 1, we prove first parts 1 and 3, followed by part 2. Recall that per the hypothesis, the state process $\{S_t\}_{t=1}^\infty$ is now a TIAMC with initial pmf P_{S_1} .

Proof of Theorem 1:

Part 1) In order to establish (25), it suffices to show that if a rate pair (R_1, R_2) is achievable when $P_{S_1} = \pi_S$, it is also achievable for any arbitrary but fixed initial pmf $P_{S_1} = \bar{P}$. By definition, if (R_1, R_2) is an achievable rate pair when $P_{S_1} = \pi_S$, for each $\delta > 0$, there exists a sequence of length- n block codes $(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)})$ with message set $\mathcal{M}_j^{(n)}$ of cardinality $M_j^{(n)}$ for sender- j , $j = 1, 2$, where

$$\frac{\log M_j^{(n)}}{n} > R_j - \delta$$

and

$$P_{e, \pi_S}(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)}) < \epsilon_n, \quad \lim_n \epsilon_n = 0 \quad (73)$$

where the second subscript in P_{e, π_S} makes explicit its dependence on π_S .

Next, when the state process $\{S_t\}_{t=1}^\infty$ is a TIAMC with $P_{S_1} = \bar{P}$, for each $\rho > 0$, there exists a positive integer $T(\bar{P}, \Gamma, \rho) < \infty$, such that

$$\Pr(S_t \in \sigma_{\pi_S}) > 1 - \rho, \quad t \geq T. \quad (74)$$

Consider now a sequence of length- $(T+n)$ block codes

$$(\bar{\theta}_1^{(T+n)}, \bar{\theta}_2^{(T+n)}, \bar{\phi}^{(T+n)})$$

obtained, in effect, by translating $(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)})$ to the right by T (symbol) time units. Precisely

$$\bar{\theta}_{j,t}^{(T+n)}(m_j, h_j(s^t)) = \theta_{j,t-T}^{(n)}(m_j, h_j(s_{t-T}^t)), \quad t > T, j = 1, 2$$

$$\bar{\theta}_{j,t}^{(T+n)}(m_j, h_j(s^t)) = \bar{x}_j, \quad t \leq T, j = 1, 2$$

and

$$\bar{\phi}^{(T+n)}(y^{T+n}, d^{T+n}) = \phi(y_{T+1}^{T+n}, d_{T+1}^{T+n}) \quad (75)$$

where \bar{x}_j is an arbitrary but fixed element of \mathcal{X}_j . It then follows that

$$\begin{aligned} & P_{e, \bar{P}}(\bar{\theta}_1^{(T+n)}, \bar{\theta}_2^{(T+n)}, \bar{\phi}^{(T+n)}) \\ &= \sum_{s \in \mathcal{S}} \Pr(\bar{\phi}^{(T+n)}(Y^{T+n}, D^{T+n}) \\ & \quad \neq (\bar{M}_1, \bar{M}_2) | S_T = s) \Pr(S_T = s) \\ & \leq \Pr(S_T \notin \sigma_{\pi_S}) \\ & \quad + \sum_{s \in \sigma_{\pi_S}} \Pr(\bar{\phi}^{(T+n)}(Y^{T+n}, D^{T+n}) \\ & \quad \neq (\bar{M}_1, \bar{M}_2) | S_T = s) \Pr(S_T = s) \end{aligned}$$

$$\begin{aligned} & \leq \rho + \sum_{s \in \sigma_{\pi_S}} \Pr(\phi(Y_{T+1}^{T+n}, D_{T+1}^{T+n}) \neq (\bar{M}_1, \bar{M}_2) | S_T = s) \\ & \quad \times \Pr(S_T = s), \quad \text{by (75)} \\ & = \rho + \sum_{s \in \sigma_{\pi_S}} \Pr(\phi(Y^n, D^n) \neq (\bar{M}_1, \bar{M}_2) | S_1 = s) \\ & \quad \times \Pr(S_T = s) \end{aligned} \quad (76)$$

where the last equality is a consequence of the memorylessness property in (23). Next, by virtue of the following elementary bound: for each $s \in \sigma_{\pi_S}$

$$\begin{aligned} & \Pr(\phi(Y^n, D^n) \neq (\bar{M}_1, \bar{M}_2) | S_1 = s) \\ & \leq \frac{P_{e, \pi_S}(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)})}{\min_{s' \in \sigma_{\pi_S}} \pi_S(s')} \end{aligned} \quad (77)$$

we obtain from (73), (74), and (76) that

$$P_{e, \bar{P}}(\bar{\theta}_1^{(T+n)}, \bar{\theta}_2^{(T+n)}, \bar{\phi}^{(T+n)}) \leq \rho + \frac{\epsilon_n}{\min_{s' \in \sigma_{\pi_S}} \pi_S(s')}. \quad (78)$$

Since $\lim_n \epsilon_n = 0$, it is clear from (78) that (R_1, R_2) is, for instance, a 2ρ -achievable rate pair when $P_{S_1} = \bar{P}$. Since ρ can be chosen arbitrarily, the proof of (25) is complete.

That the inclusion in (25) can be strict is illustrated by Example 1 above.

Part 3) We rely on the following two propositions, whose proofs are relegated to Appendix B.

Proposition 3: Under the hypothesis of Theorem 1, let (R_1, R_2) be an achievable rate pair when $P_{S_1} = \bar{P}$. Then (R_1, R_2) is also achievable for every initial pmf $P_{S_1} = \check{P}$, with $\sigma_{\check{P}} \subseteq \sigma_{\bar{P}}$.

Proposition 4: Under the hypothesis of Theorem 1, let (R_1, R_2) be an achievable rate pair for the initial pmf $P_{S_1} = \bar{P}$, where \bar{P} is such that $E_{j,1}$, $j = 1, 2$, are constants almost surely (a.s.), and $\sigma_{\bar{P}} \rightarrow \mathcal{B} \subseteq \mathcal{S}$. Then (R_1, R_2) is also an achievable rate pair for every initial pmf $P_{S_1} = \check{P}$ with $\sigma_{\check{P}} \subseteq \mathcal{B}$.

Let (R_1, R_2) be an achievable rate pair for initial pmf $P_{S_1} = \bar{P}$, for some arbitrary pmf \bar{P} on \mathcal{S} . Then, by Proposition 3, it is also an achievable rate pair when P_{S_1} is a point mass on any fixed $\tilde{s}_1 \in \sigma_{\bar{P}}$. Next, since $\{S_t\}_{t=1}^\infty$ is indecomposable, there exists a finite integer $T = T(\Gamma)$ such that

$$\Pr(S_T = r | S_1 = s) > 0, \quad r \in \sigma_{\pi_S}, s \in \mathcal{S}. \quad (79)$$

In particular, this means that there exist $\tilde{s}_2, \dots, \tilde{s}_{T-1} \in \mathcal{S}$, such that

$$\{\tilde{s}_1\} \rightarrow \dots \rightarrow \{\tilde{s}_{T-1}\} \rightarrow \{\bar{s}\}. \quad (80)$$

By Proposition 4, (R_1, R_2) is, therefore, also an achievable rate pair when P_{S_1} is a point mass on \bar{s} . Finally, by the hypothesis (26) and Proposition 4, (R_1, R_2) is also an achievable rate pair for $P_{S_1} = \pi_S$. Consequently, $\mathcal{C}_{P_{S_1}} \subseteq \mathcal{C}_{\pi_S}$, which in conjunction with (25), yields the desired assertion.

Part 2) Let s, r be distinct states in σ_{π_S} . Then, s (resp., r) is connected to r (resp., s) in the sense of (79), (80). This, combined with Proposition 4, yields the desired assertion. \square

Proof of Corollary 1 Part 1: The proof of the corollary is similar to that of Theorem 1 part 3, and entails a repeated application of Propositions 3 and 4. \square

Proof of Corollary 1 Part 2: This is a consequence of Theorem 1 part 2, and the following observation. The transmitters employ “good” codes corresponding to the (known) initial state. Since the state space is finite and fixed, the encoders can inform the decoder of the initial state by adding short suffixes to the original codewords; the overall rate pair of the code is, of course, only negligibly affected by the addition of the suffixes. \square

Proof of Theorem 2:

Part 1) Forward Part: Fix a positive integer k . First, observe that due to the encoder restriction of window- k , each codeword for sender- j , $j = 1, 2$, is a sequence of mappings $f_{j,1}, \dots, f_{j,n}$, where

$$f_{j,t} \in \begin{cases} \mathcal{F}_j(t), & t < k \\ \mathcal{F}_j(k), & t \geq k. \end{cases}$$

Next, from the forward part of Proposition 1, and Lemma 4 in Appendix B, it follows that the set of rate pairs given by

$$\bigcup_{\mathbf{F}_1, \mathbf{F}_2} \left\{ (R_1, R_2): \begin{aligned} 0 &\leq R_1 \leq \underline{\lim}_n p \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | F_{1,k}^n, D_k^n)} \\ 0 &\leq R_2 \leq \underline{\lim}_n p \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | F_{2,k}^n, D_k^n)} \\ R_1 + R_2 &\leq \underline{\lim}_n p \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | D_k^n)} \end{aligned} \right\} \quad (81)$$

is achievable, where the joint pmf of the rvs $F_{1,k}^n, F_{2,k}^n, D_k^n, Y_k^n$ obtains from (16) and (17), with $F_{j,t}, j=1,2$, being $\mathcal{F}_j(k)$ -valued for $t \geq k$.

Let (L_j, μ_j) , $j = 1, 2$, be an indecomposable Markov source (cf., e.g., [7]) consisting of a time-invariant, indecomposable Markov chain⁹ (TIMC) $\{L_j, t\}_{t=k}^\infty$ with state space $\mathcal{L}_j = \{1, \dots, \lambda_j\}$, “initial” pmf $P_{L_j, k}$, and a $\lambda_j \times \lambda_j$ -transition matrix Γ_{L_j} ; and a mapping μ_j with domain \mathcal{L}_j and range $\mathcal{F}_j(k)$. We remark that the resulting process $\{F_{j,t} = \mu_j(L_j, t)\}_{t=k}^\infty$ is often referred to as a hidden Markov process. Further, assume that the processes $\{L_1, t\}_{t=k}^\infty, \{L_2, t\}_{t=k}^\infty$, and the state process $\{S_t\}_{t=1}^\infty$ are mutually independent. The memorylessness property in (23) then implies that $\{L_1, t, L_2, t, S_{t-k+1}^t, Y_t\}_{t=k}^\infty$ is a TIMC. Precisely, for each $t \geq k$

$$\begin{aligned} P_{L_1 L_2 S Y | L_1 L_2 S Y}(l_1, t+1, l_2, t+1, s_{t-k+2}^{t+1}, y_{t+1} | l_1^t, l_2^t, r^t, y^t) \\ = \Gamma_{L_1}(l_1, t+1 | l_1, t) \Gamma_{L_2}(l_2, t+1 | l_2, t) \\ \times \Gamma(s_{t+1} | r_t) \mathbb{1}(s_{t-k+2}^t = r_{t-k+2}^t) \\ \times W(y_{t+1} | \mu_1(l_1, t+1)(h_1(s_{t-k+2}^{t+1})), \\ \mu_2(l_2, t+1)(h_2(s_{t-k+2}^{t+1})), s_{t+1}). \end{aligned} \quad (82)$$

Since

$$\{(F_{1,t}, F_{2,t}, D_t, Y_t)\}_{t=k}^\infty = \{(\mu_1(L_1, t), \mu_2(L_2, t), g(S_t), Y_t)\}_{t=k}^\infty$$

⁹We do not insist that the process $\{L_j, t\}_{t=k}^\infty$ be aperiodic, and hence it is not necessarily a TIAMC.

is obtained as a function of this TIMC, [7, Theorem 2] asserts that the following hold a.s.:

$$\begin{aligned} \lim_t \frac{1}{t} \log \frac{\Pr(Y_k^t | F_{1,k}^t, F_{2,k}^t, D_k^t)}{\Pr(Y_k^t | F_{2,k}^t, D_k^t)} \\ = \lim_n \frac{1}{n} I(F_{1,k}^n \wedge Y_k^n | F_{2,k}^n, D_k^n) \\ \lim_t \frac{1}{t} \log \frac{\Pr(Y_k^t | F_{1,k}^t, F_{2,k}^t, D_k^t)}{\Pr(Y_k^t | F_{1,k}^t, D_k^t)} \\ = \lim_n \frac{1}{n} I(F_{2,k}^n \wedge Y_k^n | F_{1,k}^n, D_k^n) \end{aligned}$$

and

$$\begin{aligned} \lim_t \frac{1}{t} \log \frac{\Pr(Y_k^t | F_{1,k}^t, F_{2,k}^t, D_k^t)}{\Pr(Y_k^t | D_k^t)} \\ = \lim_n \frac{1}{n} I(F_{1,k}^n, F_{2,k}^n \wedge Y_k^n | D_k^n). \end{aligned} \quad (83)$$

Hence, the terms involving $\underline{\lim}_n p$ in (81) are equal to the corresponding limits of mutual information quantities which appear on the right-hand side of (83). Furthermore, the quantities on the right-hand side of (83) do not depend on the initial pmfs $P_{S_1}, P_{L_1, k}$, and $P_{L_2, k}$. Also, the joint pmf of $F_{1,k}^n, F_{2,k}^n, D_k^n, Y_k^n$, which can be easily obtained from (82), is seen to be consistent with (16) and (17), whence, from (81) and (83), we conclude that the closure of the region

$$\bigcup_{(L_1, \mu_1), (L_2, \mu_2)} \left\{ (R_1, R_2): \begin{aligned} 0 &\leq R_1 \leq \lim_n \frac{1}{n} I(F_{1,k}^n \wedge Y_k^n | F_{2,k}^n, D_k^n) \\ 0 &\leq R_2 \leq \lim_n \frac{1}{n} I(F_{2,k}^n \wedge Y_k^n | F_{1,k}^n, D_k^n) \\ R_1 + R_2 &\leq \lim_n \frac{1}{n} I(F_{1,k}^n, F_{2,k}^n \wedge Y_k^n | D_k^n) \end{aligned} \right\} \quad (84)$$

is achievable, where the union is over all independent, indecomposable Markov sources as above. We denote this achievable region by $\underline{\mathcal{C}}(k)$.

Converse Part: Recall first that $F_{j,t}$ is a $\mathcal{F}_j(k)$ -valued rv for $t \geq k$, by virtue of the encoder restriction of window- k . The following proposition, which is established in Appendix C, will be used in our proof.

Proposition 5: Consider the memoryless TVMAC in (23) under the hypothesis of Theorem 2. Then, for any fixed positive integer k , and any sequence of pmfs $\mathbf{F}_1, \mathbf{F}_2$, such that $F_{j,t}$ is a $\mathcal{F}_j(k)$ -valued rv for $t \geq k, j = 1, 2$

$$\begin{aligned} \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n) \\ \underline{\lim}_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n) \end{aligned}$$

and

$$\underline{\lim}_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n)$$

do not depend on P_{S_1} , where the joint pmf of F_1^n, F_2^n, D^n, Y^n is as specified in (16) and (17).

Denote by $\overline{\mathcal{C}}(k)$ the region on the right-hand side of (29). From Proposition 1 and Lemma 3 above, we know that $\mathcal{C}(k) \subseteq \overline{\mathcal{C}}(k)$. We complete the proof of Theorem 2 part 1 by showing that $\overline{\mathcal{C}}(k) = \underline{\mathcal{C}}(k)$. Recall that $\underline{\mathcal{C}}(k)$ does not depend on the

initial pmf P_{S_1} of the state process $\{S_t\}_{t=1}^\infty$. Since Proposition 5 gives that the region $\overline{\mathcal{C}}(k)$ also does not depend on P_{S_1} , we are free to assume hereafter in this proof that $P_{S_1} = \pi_S$, i.e., the state process $\{S_t\}_{t=1}^\infty$ is stationary.

In order to show that $\underline{\mathcal{C}}(k) = \overline{\mathcal{C}}(k)$, for each sequence of pmfs $\underline{F}_1, \underline{F}_2$, with the corresponding joint pmf of F_1^n, F_2^n, D^n, Y^n as in (16) and (17), and every $\epsilon > 0$, we must exhibit Markov sources (L_1, μ_1) and (L_2, μ_2) , with the processes $\{L_{1,t}\}_{t=k}^\infty, \{L_{2,t}\}_{t=k}^\infty$, and $\{S_t\}_{t=1}^\infty$ being mutually independent, and satisfying

$$\begin{aligned} \lim_n \frac{1}{n} I(G_{1,k}^n \wedge Y_k^n | G_{2,k}^n, D_k^n) &\geq I_1 - \epsilon \\ \lim_n \frac{1}{n} I(G_{2,k}^n \wedge Y_k^n | G_{1,k}^n, D_k^n) &\geq I_2 - \epsilon \\ \text{and} \\ \lim_n \frac{1}{n} I(G_{1,k}^n, G_{2,k}^n \wedge Y_k^n | D_k^n) &\geq I_3 - \epsilon \end{aligned} \quad (85)$$

where

$$\begin{aligned} I_1 &\triangleq \underline{\lim}_n \frac{1}{n} I(F_1^n \wedge Y^n | F_2^n, D^n) \\ I_2 &\triangleq \underline{\lim}_n \frac{1}{n} I(F_2^n \wedge Y^n | F_1^n, D^n) \\ \text{and} \\ I_3 &\triangleq \underline{\lim}_n \frac{1}{n} I(F_1^n, F_2^n \wedge Y^n | D^n) \end{aligned} \quad (86)$$

and where $G_{j,t} = \mu_j(L_{j,t})$, $j = 1, 2, t \geq k$, is a $\mathcal{F}_j(k)$ -valued rv. The joint pmf of the rvs $G_{1,k}^n, G_{2,k}^n, D_k^n, Y_k^n$ obtains from (82). Observe that the existence of the limits on the left-hand side of (85), as also their lack of dependence on $P_{L_{1,k}}, P_{L_{2,k}}, P_{S_1}$, is guaranteed by [7, Theorem 2].

Since the alphabets $\{\mathcal{F}_1(t)\}_{t=1}^k, \{\mathcal{F}_2(t)\}_{t=1}^k, \mathcal{D}$, and \mathcal{Y} are finite, for each $\epsilon > 0$, there exists a positive integer $N_0 = N_0(\epsilon)$ such that for every $m \geq N_0$, it holds that

$$\begin{aligned} \frac{1}{m} I(F_{1,k}^{m+k-1} \wedge Y_k^{m+k-1} | F_{2,k}^{m+k-1}, D_k^{m+k-1}) &\geq I_1 - \epsilon \\ \frac{1}{m} I(F_{2,k}^{m+k-1} \wedge Y_k^{m+k-1} | F_{1,k}^{m+k-1}, D_k^{m+k-1}) &\geq I_2 - \epsilon \\ \frac{1}{m} I(F_{1,k}^{m+k-1}, F_{2,k}^{m+k-1} \wedge Y_k^{m+k-1} | D_k^{m+k-1}) &\geq I_3 - \epsilon. \end{aligned} \quad (87)$$

Fix $m \geq N_0$. Let $B_j(m)$, $j = 1, 2$, be the support of the rv $F_{j,k}^{m+k-1}$, and set $N_j = |B_j(m)|$. Clearly, N_1 and N_2 are finite. Let $(c_{j,1}^m, \dots, c_{j,N_j}^m)$, $j = 1, 2$, be an (arbitrary) ordering of the elements of the set $B_j(m)$. Define Markov sources (L_1, μ_1) and (L_2, μ_2) as follows. Set $\lambda_j = mN_j$. For convenience, the ordered pair (i, n_j) shall denote the state $[(n_j - 1)m + i]$, $i = 1, \dots, m$, and $n_j = 1, \dots, N_j$. Define an associated transition matrix Γ_{L_j} by

$$\begin{aligned} \Gamma_{L_j}((i, n_j)|(i+1, n_j)) &= 1, \\ & i = 1, \dots, (m-1), \text{ and } n_j = 1, \dots, N_j \\ \Gamma_{L_j}((m, n'_j)|(1, n_j)) &= \Pr(F_{j,k}^{m+k-1} = c_{j,n_j}^m), \\ & n_j, n'_j = 1, \dots, N_j \end{aligned} \quad (88)$$

with the remaining elements of Γ_{L_j} set to 0. The ‘‘initial’’ pmf $P_{L_{j,k}}$ of the process $\{L_{j,t}\}_{t=k}^\infty$ is chosen to be

$$P_{L_{j,k}}((i, n_j)) = \begin{cases} \Pr(F_{j,k}^{m+k-1} = c_{j,n_j}^m), & \text{if } i = 1 \\ 0, & \text{otherwise.} \end{cases} \quad (89)$$

Finally, define the mapping $\mu_j: \mathcal{L}_j \rightarrow \mathcal{F}_j(k)$ as

$$\mu_j((i, n_j)) = c_{j,n_j}^m(i), \quad i = 1, \dots, m \text{ and } n_j = 1, \dots, N_j \quad (90)$$

where $c_{j,n_j}^m(i)$ denotes the i th symbol of the sequence c_{j,n_j}^m . It is clear that the processes $\{L_{j,t}\}_{t=k}^\infty, j = 1, 2$, are TIMCs. The joint pmf of the rvs $G_{1,k}^n, G_{2,k}^n, D_k^n, Y_k^n$ can be easily seen to exhibit the following properties.

1) *Block Memorylessness*: For each positive integer T , we have

$$\begin{aligned} \Pr(G_{j,k}^{mT+k-1} = f_{j,k}^{mT+k-1}) \\ &= \prod_{t=0}^{T-1} \Pr(G_{j,k}^{m+k-1} = f_{j,mt+k}^{mt+m+k-1}) \\ &= \prod_{t=0}^{T-1} \Pr(F_{j,k}^{m+k-1} = f_{j,mt+k}^{mt+m+k-1}). \end{aligned} \quad (91)$$

2) *Block Stationarity*: The stationarity of the state process $\{S_t\}_{t=1}^\infty$, the memorylessness property (23), and (91), imply that for each positive integer t , the rvs $G_{1,mt+k}^{mt+m+k-1}, G_{2,mt+k}^{mt+m+k-1}, D_k^{mt+m+k-1}, Y_k^{mt+m+k-1}$ and $F_{1,k}^{m+k-1}, F_{2,k}^{m+k-1}, D_k^{m+k-1}, Y_k^{m+k-1}$, are identically distributed.

Next, observe that for every positive integer T

$$\begin{aligned} \frac{1}{mT} I(G_{1,k}^{mT+k-1} \wedge Y_k^{mT+k-1} | G_{2,k}^{mT+k-1}, D_k^{mT+k-1}) \\ &> \frac{1}{mT} \sum_{t=0}^{T-1} I(G_{1,mt+k}^{mt+m+k-1} \wedge Y_{mt+k}^{mt+m+k-1} | G_{2,mt+k}^{mt+m+k-1}, \\ & \quad D_{mt+k}^{mt+m+k-1}), \text{ by ‘‘block memorylessness’’} \\ &= \frac{1}{m} I(F_{1,k}^{m+k-1} \wedge Y_k^{m+k-1} | F_{2,k}^{m+k-1}, D_k^{m+k-1}), \\ & \quad \text{by ‘‘block stationarity’’} \\ &\geq I_1 - \epsilon. \end{aligned} \quad (92)$$

Likewise,

$$\begin{aligned} \frac{1}{mT} I(G_{2,k}^{mT+k-1} \wedge Y_k^{mT+k-1} | G_{1,k}^{mT+k-1}, D_k^{mT+k-1}) \\ &\geq I_2 - \epsilon \\ \text{and} \\ \frac{1}{mT} I(G_{1,k}^{mT+k-1}, G_{2,k}^{mT+k-1} \wedge Y_k^{mT+k-1} | D_k^{mT+k-1}) \\ &\geq I_3 - \epsilon. \end{aligned} \quad (93)$$

The existence of the limits on the left-hand sides of (92) and (93) completes the proof of (85), whereby it follows that $\underline{\mathbf{C}}(k) = \overline{\mathbf{C}}(k) = \mathbf{C}(k)$. \square

Part 2) Denoting by $\overline{\mathbf{C}}_{\pi_S}$ the region on the right-hand side of (33), it follows from Lemma 3, and the definition of $\mathbf{C}(k)$, that

$$\overline{\mathbf{C}}_{\pi_S} \supseteq \mathbf{C}_{\pi_S} \supseteq \lim_k \mathbf{C}(k)$$

whence it suffices to show that

$$\lim_k \mathbf{C}(k) = \overline{\mathbf{C}}_{\pi_S}. \quad (94)$$

In order to establish (94), we show that for each sequence of pmfs $\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2$, with the corresponding joint pmf of F_1^n, F_2^n, D^n, Y^n as in (16) and (17), and $\epsilon > 0$, there exists a positive integer K , and a sequence of pmfs $\underline{\mathbf{G}}_1, \underline{\mathbf{G}}_2$, such that

$$\lim_n \frac{1}{n} I(G_1^n \wedge Y^n | G_2^n, D^n) \geq I_1 - \epsilon$$

$$\lim_n \frac{1}{n} I(G_2^n \wedge Y^n | G_1^n, D^n) \geq I_2 - \epsilon$$

and

$$\lim_n \frac{1}{n} I(G_1^n, G_2^n \wedge Y^n | D^n) \geq I_3 - \epsilon \quad (95)$$

where I_1, I_2 , and I_3 are, respectively, the $\underline{\lim}_n$ terms appearing on the right-hand side of (33), and $G_{j,t}$ is a $\mathcal{F}_j(K)$ -valued rv for $t \geq K$. The joint pmf of the rvs G_1^n, G_2^n, D^n, Y^n corresponds to that in (16), (17) and (30) with F_j^n being replaced by G_j^n , $j = 1, 2$. Choose a positive integer m such that

$$\frac{1}{m} I(F_1^m \wedge Y^m | F_2^m, D^m) \geq I_1 - \epsilon$$

$$\frac{1}{m} I(F_2^m \wedge Y^m | F_1^m, D^m) \geq I_2 - \epsilon$$

and

$$\frac{1}{m} I(F_1^m, F_2^m \wedge Y^m | D^m) \geq I_3 - \epsilon \quad (96)$$

and, for each positive integer T , define

$$\Pr(G_j^{mT} = f_j^{mT}) = \prod_{t=0}^{T-1} \Pr(F_j^m = f_j^{mt+m}). \quad (97)$$

Due to the stationarity of the state process $\{S_t\}_{t=1}^\infty$, it follows from (96) and (97) that

$$\begin{aligned} & \underline{\lim}_T \frac{1}{mT} I(G_1^{mT} \wedge Y^{mT} | G_2^{mT}, D^{mT}) \\ & \geq \frac{1}{m} I(F_1^m \wedge Y^m | F_2^m, D^m) \geq I_1 - \epsilon \end{aligned}$$

$$\begin{aligned} & \underline{\lim}_T \frac{1}{mT} I(G_2^{mT} \wedge Y^{mT} | G_1^{mT}, D^{mT}) \\ & \geq \frac{1}{m} I(F_2^m \wedge Y^m | F_1^m, D^m) \geq I_2 - \epsilon \end{aligned}$$

and

$$\begin{aligned} & \underline{\lim}_T \frac{1}{mT} I(G_1^{mT}, G_2^{mT} \wedge Y^{mT} | D^{mT}) \\ & \geq \frac{1}{m} I(F_1^m, F_2^m \wedge Y^m | D^m) \geq I_3 - \epsilon. \end{aligned}$$

Since our choice of ϵ was arbitrary, it is now evident from Theorem 2 part 1 that $\lim_k \mathbf{C}(k) = \overline{\mathbf{C}}_{\pi_S}$. \square

Proof of Theorem 3: Fix a value of k . For the converse, we apply Proposition 2 and [20, Theorem 8(h)] to obtain the inclusion

$$\mathbf{C}(k) \subseteq \bigcup_{\underline{\mathbf{X}}_1 | \underline{\mathbf{E}}_1, \underline{\mathbf{X}}_2 | \underline{\mathbf{E}}_2} \left\{ \begin{array}{l} 0 \leq R_1 \leq \underline{\lim}_n \frac{1}{n} I(X_1^n \wedge Y^n | X_2^n, S^n) \\ (R_1, R_2): 0 \leq R_2 \leq \underline{\lim}_n \frac{1}{n} I(X_2^n \wedge Y^n | X_1^n, S^n) \\ R_1 + R_2 \leq \underline{\lim}_n \frac{1}{n} I(X_1^n, X_2^n \wedge Y^n | S^n) \end{array} \right\} \quad (98)$$

where the joint pmf of the rvs X_1^n, X_2^n, S^n, Y^n is given by

$$\begin{aligned} P_{X_1 X_2 S Y}(x_1^n, x_2^n, s^n, y^n) &= P_S(s^n) W(y^n | x_1^n, x_2^n, s^n) \\ &\times \sum_{e_1^n \in \mathcal{E}_1^n, e_2^n \in \mathcal{E}_2^n} \prod_{j=1}^n 1(h_j(s^n) = e_j^n) P_{X_j | E_j}(x_j^n | e_j^n) \end{aligned}$$

and the union in (98) is taken over all sequences of conditional pmfs of X_j^n given E_j^n , $j = 1, 2$, satisfying (35). Since the alphabets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{S}, \mathcal{Y}$ are finite, it follows that

$$\underline{\lim}_n \frac{1}{n} I(X_1^n \wedge Y^n | X_2^n, S^n) = \underline{\lim}_n \frac{1}{n} I(X_{1,k}^n \wedge Y_k^n | X_{2,k}^n, S^n).$$

Furthermore, the memorylessness property of the TVMAC in (23) implies that

$$\frac{1}{n} I(X_{1,k}^n \wedge Y_k^n | X_{2,k}^n, S_k^n) \leq \frac{1}{n} \sum_{t=k}^n I(X_{1,t} \wedge Y_t | X_{2,t}, S_{t-k+1}^t).$$

Proceeding likewise, it is now evident from a standard argument (cf., e.g., [3, p. 277]) that $\mathbf{C}(k)$ is contained in the convex closure of the region in (36), where the joint pmf of $X_{1,k}, X_{2,k}, S^k, Y_k$ is given by (37).

For the forward part, fix conditional pmfs $Q_{X_j | E_j}$ of $X_{j,k}$ given E_j^k , $j = 1, 2$, and for each $n \geq k$, define

$$P_{X_j | E_j}(x_j^n | e_j^n) = \prod_{t=k}^n Q_{X_j | E_j}(x_j, t | e_j^t, t-k+1). \quad (99)$$

The joint pmfs of the rvs $X_{1,k}^n, X_{2,k}^n, S^n, Y_k^n$ for $n \geq k$ is given by

$$\begin{aligned} & P_{X_1 X_2 S Y}(x_{1,k}^n, x_{2,k}^n, s^n, y_k^n) \\ &= \sum_{e_1^n \in \mathcal{E}_1^n, e_2^n \in \mathcal{E}_2^n} P_S(s^n) 1(h_1(s^n) = e_1^n) 1(h_2(s^n) = e_2^n) \\ &\quad \times P_{X_1 | E_1}(x_{1,k}^n | e_1^n) P_{X_2 | E_2}(x_{2,k}^n | e_2^n) \\ &\quad \times W(y_k^n | x_{1,k}^n, x_{2,k}^n, s_k^n), \end{aligned} \quad (100)$$

and is seen to be consistent with (21). Combining (99) and (100), it follows from the stationarity of the state process $\{S_t\}_{t=1}^\infty$ and the memorylessness of the TVMAC in (23) that

$$\begin{aligned} & \underline{\lim}_p \frac{1}{n} \log \frac{\Pr(Y_k^n | X_{1,k}^n, X_{2,k}^n, S^n)}{\Pr(Y_k^n | X_{2,k}^n, S^n)} \\ &= \underline{\lim}_p \frac{1}{n} \sum_{t=k}^n \log \frac{\Pr(Y_t | X_{1,t}, X_{2,t}, S_{t-k+1}^t)}{\Pr(Y_t | X_{2,t}, S_{t-k+1}^t)} \\ &= I(X_{1,k} \wedge Y_k | X_{2,k}, S^k). \end{aligned}$$

Proceeding likewise, we conclude from Proposition 1 that the region in (36) is achievable (cf. Lemma 4). Finally, since the state process $\{S_t\}_{t=1}^{\infty}$ is stationary, by time-sharing for the TVMAC in (23), the convex combination of given achievable rate pairs is also achievable, thereby concluding the proof.

It remains to show that in the absence of any encoder restrictions

$$\mathbf{C} = \lim_k \mathbf{C}(k) = \sup_k \mathbf{C}(k).$$

The idea behind the proof is the same as that used to establish an analogous result in Theorem 2 part 2, and is not repeated here.

Proof of Corollary 2: First, observe that as an immediate specialization of Theorem 3, we get that the capacity of the single-sender memoryless TVC in (28), without any encoder restriction, under the remaining hypothesis of Theorem 3, is

$$C = \lim_k C(k) = \sup_k C(k) \quad (101)$$

where

$$C(k) = \max_{X_k|E^k: X_k \text{---} E^k \text{---} S^k} I(X_k \wedge Y_k|S^k) \quad (102)$$

with the joint pmf of $X_k, S^k, Y_k, k \geq 1$, being given by (41) upon replacing l by k . Then this joint pmf implies, for each $k \geq 1$, that

$$I(X_k \wedge Y_k|S^k) = I(X_k \wedge Y_k|S_k, E^k) \quad (103)$$

whence it follows that

$$C(k) = \max_{X_k|E^k: X_k \text{---} E^k \text{---} S_k} I(X_k \wedge Y_k|S_k, E^k). \quad (104)$$

Now, fix a value of $k > l$. Then by the memorylessness property of the TVC in (28), we get that

$$I(X_k \wedge Y_k|S_k, E^k) \leq I(X_k \wedge Y_k|S_k, E_{k-l+1}^k). \quad (105)$$

Next, the Markov condition appearing in the domain of maximization in (104), viz.

$$X_k \text{---} E^k \text{---} S_k \quad (106)$$

together with the hypothesis (39), imply that

$$X_k \text{---} E_{k-l+1}^k \text{---} S_k. \quad (107)$$

This is easily seen, since

$$\begin{aligned} I(X_k \wedge S_k|E_{k-l+1}^k) &\leq I(X_k, E^{k-l} \wedge S_k|E_{k-l+1}^k) \\ &= I(E^{k-l} \wedge S_k|E_{k-l+1}^k) + I(X_k \wedge S_k|E^k) = 0 \end{aligned}$$

where the previous equality is by (39) and (106).

Consequently, by (104), (105), and (107)

$$\begin{aligned} C(k) &\leq \max_{X_k|E_{k-l+1}^k: X_k \text{---} E_{k-l+1}^k \text{---} S_k} I(X_k \wedge Y_k|S_k, E_{k-l+1}^k) \\ &= C(l) \quad (108) \end{aligned}$$

where the previous equality follows by the assumed stationarity of $\{S_t\}_{t=1}^{\infty}$ and the expression for $C(l)$ from (104). Combining

(108) with the fact that $\{C(k)\}_{k=1}^{\infty}$ is nondecreasing in k , yields that $C(k) = C(l)$ for all $k \geq l$. Hence, by (101), $C = C(l)$. \square

Proof of Theorem 4:

Part 1) Converse Part: Let (R_1, R_2) be an achievable rate pair, so that for any $\epsilon, \delta > 0$, there exist length- n block codes $(\theta_1, \theta_2, \phi)$ with

$$P_e(\theta_1, \theta_2, \phi) < \epsilon$$

and

$$\frac{\log M_j}{n} \geq R_j - \delta, \quad j = 1, 2 \quad (109)$$

for all sufficiently large n . Assume that the message rvs \overline{M}_1 and \overline{M}_2 are uniformly distributed on the message sets \mathcal{M}_1 and \mathcal{M}_2 , respectively; and that $\overline{M}_1, \overline{M}_2$, the state process $\{S_t\}_{t=1}^{\infty}$, and the noise process $\{N_t\}_{t=1}^{\infty}$ are mutually independent. Then, it readily follows from (8) that

$$\begin{aligned} Y^n \text{---} x_1^n(\overline{M}_1, h_1(S^n))x_2^n(\overline{M}_2, h_2(S^n))S^n \text{---} \overline{M}_1\overline{M}_2 \\ Y^n \text{---} x_j^n(\overline{M}_j, h_j(S^n))S^n \text{---} \overline{M}_j, \quad j = 1, 2. \quad (110) \end{aligned}$$

Next, Fano's inequality yields

$$\begin{aligned} (1 - \epsilon)(R_j - \delta) \\ \leq \frac{1}{n} [I(\overline{M}_j \wedge Y^n|S^n) + 1], \quad j = 1, 2 \end{aligned}$$

and

$$\begin{aligned} (1 - \epsilon)(R_1 + R_2 - 2\delta) \\ \leq \frac{1}{n} [I(\overline{M}_1, \overline{M}_2 \wedge Y^n|S^n) + 1]. \quad (111) \end{aligned}$$

Considering the first upper bound in (111), observe that

$$\begin{aligned} \frac{1}{n} I(\overline{M}_1 \wedge Y^n|S^n) &\leq \frac{1}{n} I(x_1^n(\overline{M}_1, h_1(S^n)) \wedge Y^n | \\ &\quad x_2^n(\overline{M}_2, h_2(S^n)), S^n), \quad \text{by (110)} \\ &\leq \frac{1}{n} \left[\sum_{t=1}^n \mathbb{H}(Y_t | x_{2,t}(\overline{M}_2, h_2(S^t)), S^t) \right] - \mathbb{H}(N) \\ &= \frac{1}{n} \left[\sum_{t=1}^n \mathbb{H}(x_{1,t}(\overline{M}_1, h_1(S_{\max\{1, t-k+1\}^t})) \tilde{S}_{1,t} + N_t | \right. \\ &\quad \left. S_{\max\{1, t-k+1\}^t}^t) \right] - \mathbb{H}(N) \quad (112) \end{aligned}$$

where $\mathbb{H}(\cdot)$ is the differential entropy function. The last equality follows from (8) and the encoder restriction of window- k . Let $\mu_{j,t}(h_j(s_{\max\{1, t-k+1\}^t}^t))$, $j = 1, 2$, be the "average (input) power at time t " when the CSI at encoder- j is $h_j(s_{\max\{1, t-k+1\}^t}^t)$, i.e.,

$$\begin{aligned} \mu_{j,t} \left(h_j \left(s_{\max\{1, t-k+1\}^t}^t \right) \right) \\ = \frac{1}{M_j} \sum_{m_j=1}^{M_j} \left| x_{j,t} \left(m_j, h_j \left(s_{\max\{1, t-k+1\}^t}^t \right) \right) \right|^2. \end{aligned}$$

Since the noise process $\{N_t\}_{t=1}^{\infty}$, the state process $\{S_t\}_{t=1}^{\infty}$, and the message rv \bar{M}_1 are mutually independent, we obtain

$$\begin{aligned} & \mathbb{E} \left[\left| x_{1,t} \left(\bar{M}_1, h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) \tilde{S}_{1,t} + N_t \right|^2 \right. \\ & \quad \left. S_{\max\{1, t-k+1\}}^t \right] \\ & = \mu_{1,t} \left(h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) |\tilde{S}_{1,t}|^2 + \sigma_N^2. \end{aligned}$$

Consequently,

$$\begin{aligned} & \mathbb{h} \left(x_{1,t} \left(\bar{M}_1, h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) \tilde{S}_{1,t} + N_t \right) \\ & \quad S_{\max\{1, t-k+1\}}^t \\ & \leq \frac{1}{2} \mathbb{E}_S \left[\log \left[2\pi e \left(\mu_{1,t} \left(h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) |\tilde{S}_{1,t}|^2 + \sigma_N^2 \right) \right] \right]. \end{aligned}$$

Continuing with the chain of inequalities in (111) and (112)

$$\begin{aligned} & (1-\epsilon)(R_1 - \delta) \\ & \leq \frac{1}{n} \left[\sum_{t=1}^n \mathbb{E}_S \left[\zeta \left(\mu_{1,t} \left(h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) |\tilde{S}_{1,t}|^2, \right. \right. \right. \\ & \quad \left. \left. \left. \sigma_N^2 \right) \right] + 1 \right] \\ & = \frac{1}{n} \left[\sum_{t=1}^n \mathbb{E}_S \left[\zeta \left(\mu_{1,t} \left(h_1 \left(S_{\max\{1, k-t+1\}}^k \right) \right) |\tilde{S}_{1,k}|^2, \right. \right. \right. \\ & \quad \left. \left. \left. \sigma_N^2 \right) \right] + 1 \right], \text{ since } \{S_t\}_{t=1}^{\infty} \text{ is stationary} \\ & \leq \frac{1}{n} + \mathbb{E}_S \left[\zeta \left(\left\{ \frac{1}{n} \sum_{t=1}^n \mu_{1,t} \left(h_1 \left(S_{\max\{1, k-t+1\}}^k \right) \right) \right\} \right. \right. \\ & \quad \left. \left. \times |\tilde{S}_{1,k}|^2, \sigma_N^2 \right) \right], \text{ by Jensen's inequality} \\ & = \frac{1}{n} + \mathbb{E}_S \left[\zeta \left(\mu_1(h_1(S^k)) |\tilde{S}_{1,k}|^2, \sigma_N^2 \right) \right] \quad (113) \end{aligned}$$

where $\zeta(\cdot, \cdot)$ is defined in (43), and we have set

$$\mu_1(h_1(S^k)) \triangleq \frac{1}{n} \sum_{t=1}^n \mu_{1,t} \left(h_1 \left(S_{\max\{1, k-t+1\}}^k \right) \right).$$

The power constraint \mathcal{P}_1 on the codewords for sender-1 (9) implies that

$$\begin{aligned} & \mathbb{E}_S [\mu_1(h_1(S^k))] \\ & = \frac{1}{n} \sum_{t=1}^n \mathbb{E}_S \left[\mu_{1,t} \left(h_1 \left(S_{\max\{1, t-k+1\}}^t \right) \right) \right] \leq \mathcal{P}_1. \end{aligned}$$

Clearly, μ_1 is a valid power-allocation policy in the sense of (45). Since ϵ and δ are arbitrary, it follows that

$$R_1 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(h_1(S^k)) |\tilde{S}_{1,k}|^2, \sigma_N^2 \right) \right]. \quad (114)$$

Proceeding similarly, we conclude that all achievable rate pairs (R_1, R_2) are contained in the closure of the region in (44) subject to the constraints in (45).

Forward Part: The forward part is a straightforward consequence of known results [14]–[16], combined with a simple observation. Specifically, fix power allocation policies $\mu_j: \mathcal{E}_j^k \rightarrow \mathbb{R}^+$ for sender- j , $j = 1, 2$, satisfying (45). Now consider a channel defined by

$$Y_t = \sum_{j=1}^2 x_{j,t} \tilde{S}_{j,t} \sqrt{\mu_j(h_j(S_{t-k+1}^t))} + N_t, \quad t \geq k \quad (115)$$

where the \mathbb{R} -valued transmitted sequences $\{x_{j,t}\}_{t=k}^{\infty}$ satisfy a ‘‘unity’’ power constraint, *viz.*

$$\frac{1}{n} \sum_{t=k}^n |x_{j,t}|^2 \leq 1, \quad j = 1, 2, \quad n \geq 1 \quad (116)$$

and where $\{\tilde{S}_{j,t}\}_{t=1}^{\infty}$, $j = 1, 2$, and $\{N_t\}_{t=k}^{\infty}$ are as in the hypothesis. Assume now that the encoder has no CSI while the decoder has CSI S_t at each time instant t , $t \geq 1$. Note that for the channel model in (115), the fade experienced by sender- j is

$$\tilde{S}_{j,t} \sqrt{\mu_j(h_j(S_{t-k+1}^t))}, \quad j = 1, 2$$

of which the decoder has *perfect* knowledge. Then the capacity region of the channel in (115) under these circumstances, denoted by $\tilde{\mathcal{C}}(\mu_1, \mu_2)$, is the closure of the region [14]–[16], as shown in (117) at the bottom of the page. We observe now that any achievable rate pair (R_1, R_2) for the channel in (115) is surely contained in $\mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2)$, whereby we get that $\tilde{\mathcal{C}}(\mu_1, \mu_2) \subseteq \mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2)$. It is then evident that the closure of the region in (44) is also contained in $\mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2)$, establishing the forward part.

We remark that the previous argument suggests that the transmitted signals of sender- j , $j = 1, 2$, can be designed in the following manner. First, fix a suitable power-allocation policy μ_j for sender- j as above, satisfying (45). Then, generate independent sets of \mathbb{R} -valued sequences $\{c_j^n(m_j), m_j \in \mathcal{M}_j\}$, $j = 1, 2$, in terms of such independent and identically distributed (i.i.d.) realizations of a Gaussian rv with mean zero and unit variance, as satisfy $\frac{1}{n} \|c_j^n(m_j)\|^2 \leq 1$ with $\|\cdot\|$ denoting Euclidean norm. The transmitted signal for sender- j corresponding to message $m_j \in \mathcal{M}_j$ is then the product of $c_{j,t}(m_j)$ and $\mu_j(h_j(S_{t-k+1}^t))$, $t = k, \dots, n$. \square

Part 2) Clearly,

$$\mathcal{C}(\mathcal{P}_1, \mathcal{P}_2) \supseteq \lim_k \mathcal{C}(k, \mathcal{P}_1, \mathcal{P}_2). \quad (118)$$

$$\left\{ (R_1, R_2): \begin{aligned} & 0 \leq R_1 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(h_1(S^k)) |\tilde{S}_{1,k}|^2, \sigma_N^2 \right) \right] \\ & 0 \leq R_2 \leq \mathbb{E}_S \left[\zeta \left(\mu_2(h_2(S^k)) |\tilde{S}_{2,k}|^2, \sigma_N^2 \right) \right] \\ & R_1 + R_2 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(h_1(S^k)) |\tilde{S}_{1,k}|^2 + \mu_2(h_2(S^k)) |\tilde{S}_{2,k}|^2, \sigma_N^2 \right) \right] \end{aligned} \right\}. \quad (117)$$

In order to show the reverse inequality

$$\mathbf{C}(\mathcal{P}_1, \mathcal{P}_2) \subseteq \lim_k \mathbf{C}(k, \mathcal{P}_1, \mathcal{P}_2) \quad (119)$$

while we could proceed, in principle, along the lines of the proof of Theorem 2 part 2, a simpler proof obtains here. Specifically, let (R_1, R_2) be an achievable rate pair in the absence of any encoder restriction. Then, it follows as in the proof of part 1 that there exist length- n block codes $(\theta_1, \theta_2, \phi)$ satisfying (109), and, furthermore, satisfying, for instance (cf. (114))

$$R_1 \leq \mathbb{E}_S \left[\zeta \left(\mu_1(h_1(S^n)) |\tilde{S}_{1,n}|^2, \sigma_N^2 \right) \right] + \delta \quad (120)$$

where

$$\mu_1(h_1(s^n)) = \frac{1}{n} \sum_{t=1}^n \mu_{1,t}(h_1(s_{n-t+1}^n))$$

and

$$\mu_{1,t}(h_1(s^t)) = \frac{1}{M_1} \sum_{m_1=1}^{M_1} |x_{1,t}(m_1, h_1(s^t))|^2.$$

Analogous bounds hold for R_2 and $R_1 + R_2$. Recalling that δ can be chosen arbitrarily small (with a resulting choice of n), and upon comparing said bounds with (44) for $k = n$, it is clear that $(R_1, R_2) \in \lim_k \mathbf{C}(k, \mathcal{P}_1, \mathcal{P}_2)$. This establishes (119).

The idea behind the proof of Corollary 3 is the same as that of Corollary 2, and is not repeated here. \square

V. DISCUSSION

We have undertaken a systematic study of the capacity problem for a class of TVMACs when varying degrees of CSI are provided to the transmitters and the receiver. The first result, in Proposition 1, which provides a general formula for the capacity region, is based largely on the techniques and results of [19], [20]. Unlike in [19], [20], where no CSI is involved, our general formula is in terms of random sequences of mappings rather than channel input symbols. These mappings can be dispensed with when the CSI available to the transmitters is embedded in that provided to the receiver, as shown in Proposition 2. We then consider special cases where this general formula admits simpler forms from which some useful insights can often be drawn.

Of particular interest is the case when the TVMAC satisfies the ‘‘memorylessness’’ property, and the state process $\{S_t\}_{t=1}^\infty$ is a TIAMC. As follows from Example 1, the capacity region can, in general, depend on the initial pmf of the state process $\{S_t\}_{t=1}^\infty$. A sufficient—but by no means necessary—condition is provided in Theorem 1 under which the capacity region does not depend on the initial pmf. The problem of determining a less stringent and more ‘‘natural’’ condition than ours, and indeed, a necessary and sufficient condition on the state process $\{S_t\}_{t=1}^\infty$ and CSI under which the effect on capacity of the initial pmf vanishes, remains unresolved. When window restrictions are imposed on the encoders, the capacity region, which is shown in Theorem 2 to be achieved by hidden Markov processes, does not depend on the initial state pmf.

When window restrictions are imposed on the encoders, and the decoder is provided with perfect CSI, and, furthermore, the state process $\{S_t\}_{t=1}^\infty$ is assumed to be stationary and ergodic, the capacity region can be expressed in a simple and useful form.

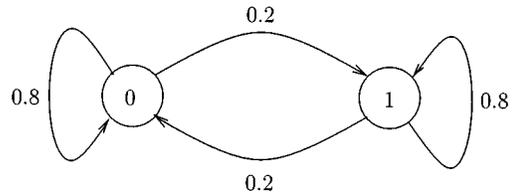


Fig. 3. Transition diagram for the state process $\{S_t\}_{t=1}^\infty$ in Example 2.

This is seen for channels with finite input, state, and output alphabets in Theorem 3, and for the TVFC model with \mathbb{R} -valued alphabets in Theorem 4.

In all our models, we describe the CSI at the transmitters and receiver at each time instant in terms of deterministic mappings of the state rv at that time instant (cf. (3)). Perhaps a more realistic situation arises when the encoder and decoder CSI at each time instant $t \geq 1$ are noisy versions of the current state s_t . This is illustrated by the following example.

Example 2 (Gilbert–Elliott channel [22]–[24]): Consider the single-sender memoryless TVC in (28) with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, $\mathcal{S} = \{0, 1\}$

$$\begin{aligned} W(0|0, 0) &= W(1|1, 0) = 0.99 \\ W(0|0, 1) &= W(1|1, 1) = 0.01. \end{aligned} \quad (121)$$

Let the state process $\{S_t\}_{t=1}^\infty$ be a TIAMC with transition diagram given in Fig. 3, and $\Gamma(0|0) = \Gamma(1|1) = 0.8$. Assume that the encoder receives no CSI, while the decoder CSI is given by

$$D_t = S_t + Z_t \bmod 2 \quad (122)$$

where $\{Z_t\}_{t=1}^\infty$ is a $\{0, 1\}$ -valued i.i.d. process, which is independent of the state process $\{S_t\}_{t=1}^\infty$, with $\Pr(Z_t = 1) = \epsilon$, $t \geq 1$. Clearly,

$$\Pr(D_t \neq S_t) = \Pr(Z_t = 1) = \epsilon.$$

Furthermore, we assume that for each $t \geq 1$, Z_t does not depend on the previous inputs and outputs of the channel. Then the augmented state process $\{\hat{S}_t\}_{t=1}^\infty$, defined by

$$\hat{S}_t = (S_t, D_t), \quad t \geq 1 \quad (123)$$

is also such that \hat{S}_t does not depend on the previous inputs and outputs of the channel. Define a new channel $\{\hat{W}(y^n|x^n, \hat{s}^n)\}_{n=1}^\infty$ by

$$\begin{aligned} \hat{W}(y^n|x^n, \hat{s}^n) &\triangleq P_{Y|X\hat{S}}(y^n|x^n, \hat{s}^n) \\ &= W(y^n|x^n, s^n) \end{aligned} \quad (124)$$

where $\hat{s}^n = (s^n, d^n)$. Observe that the decoder CSI can be described in terms of a deterministic mapping $g: \hat{\mathcal{S}} \rightarrow \mathcal{D}$, where $\hat{\mathcal{S}} = \mathcal{S} \times \mathcal{D}$. Clearly, the capacity C of the memoryless TVC in (124) with decoder CSI given by (122) equals that of the TVC in (121) with the same decoder CSI. It can be seen as a consequence of Theorem 2 that

$$C = \lim_n \max_{X^n} \frac{1}{n} I(X^n \wedge Y^n | D^n)$$

where the joint pmf of the rvs X^n, D^n, Y^n is given by

$$\begin{aligned} P_{XDY}(x^n, d^n, y^n) \\ = P_X(x^n) \sum_{s^n \in \mathcal{S}^n} P_S(s^n) P_{D|S}(d^n|s^n) W(y^n|x^n, s^n). \end{aligned}$$

Fig. 4 shows a plot of C as a function of ϵ , $0 \leq \epsilon \leq 1$. \square

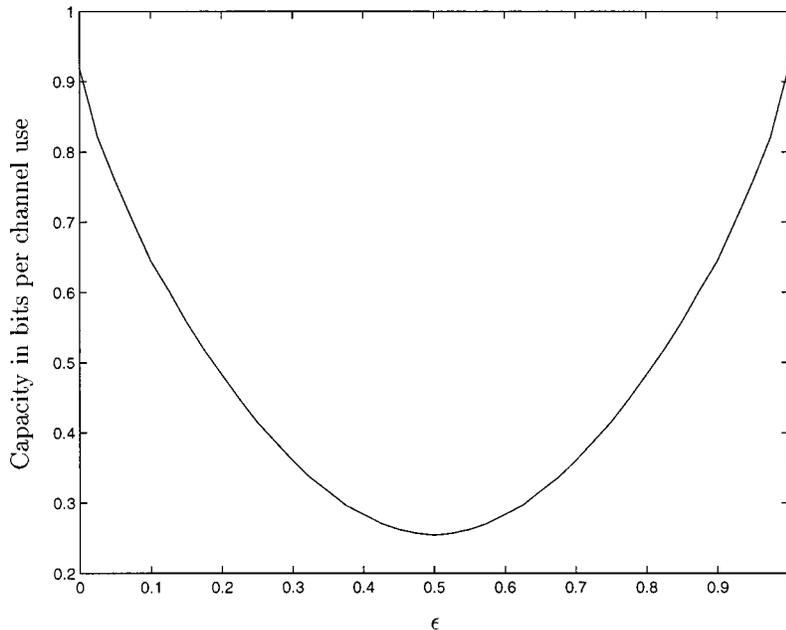


Fig. 4. Capacity of the Gilbert–Elliott channel in (121), when the decoder CSI is given by (122), as a function of ϵ .

In general, the situation where the CSI at the encoders and decoder at time t , *viz.* $E_{j,t}$, $j = 1, 2$, and D_t are noisy versions of S_t , $t \geq 1$, can be accommodated by our models. To this end, we can consider an augmented state $\hat{S}_t = (E_{1,t}, E_{2,t}, D_t, S_t)$ and the corresponding channel $\{\hat{W}(y^n|x_1^n, x_2^n, \hat{s}^n)\}_{n=1}^\infty$ defined by

$$\hat{W}(y^n|x_1^n, x_2^n, \hat{s}^n) = W(y^n|x_1^n, x_2^n, s^n)$$

where $\hat{s}^n = (s^n, e_1^n, e_2^n, d^n)$. Clearly, $E_{j,t}$, $j = 1, 2$, and D_t are now deterministic mappings of \hat{S}_t , $t \geq 1$, in conformity with our models. However, we note that if the noises corrupting transmitter and receiver CSI are different, then the encoder CSI will, in general, not be contained in the decoder CSI.

We conclude by examining a few implications of our results for the TVFC model in (8) with Rayleigh fading noise, when varying degrees of CSI are provided to the transmitters. (Recall that for this model, the receiver has access to perfect CSI.) Assume that the fading processes $\{\tilde{S}_{1,t}\}_{t=1}^\infty$ and $\{\tilde{S}_{2,t}\}_{t=1}^\infty$ in (8) are independent of each other, and that $\tilde{S}_{j,t}$, $j = 1, 2$, is a Rayleigh-distributed rv with pdf $p_{\tilde{S}_j}$ given by

$$p_{\tilde{S}_j}(s) = 2se^{-s^2}, \quad s \geq 0, \quad (125)$$

Also, assume that each sender is subject to the same input power constraint \mathcal{P} . We first consider the case where the CSI for sender- j is given in terms of

$$E_{j,t} = \sum_{i=1}^{2^L} i 1(\eta_{i-1} \leq \tilde{S}_{j,t} < \eta_i) \quad (126)$$

where $\eta_0 = 0 < \eta_1 < \dots < \eta_{2^L-1} < \eta_{2^L} = \infty$ are fixed constants. Here, the range of values of the fade of sender- j at time t is partitioned into 2^L contiguous intervals. The CSI provided to sender- j at time t , which can be represented using L bits, specifies the interval containing the value of $\tilde{S}_{j,t}$. Denoting by $\eta^{(L)}$ a fixed set of constants $(\eta_1, \dots, \eta_{2^L-1})$ as above, and by $\mathcal{C}^{\eta^{(L)}}(1, \mathcal{P}, \mathcal{P})$, the corresponding capacity region under an encoder restriction of window-1, the “throughput” of the system

is defined as (cf., for instance, [25], for the definition of an analogous “capacity”)

$$T^{\eta^{(L)}}(1, \mathcal{P}, \mathcal{P}) = \max \left\{ R_1 : R_1 = R_2 \text{ and } (R_1, R_2) \in \mathcal{C}^{\eta^{(L)}}(1, \mathcal{P}, \mathcal{P}) \right\}. \quad (127)$$

We are interested in determining the maximum value of throughput as a function of $\eta^{(L)}$, denoted $T^{(L)}(\mathcal{P})$, *i.e.*,

$$T^{(L)}(\mathcal{P}) = \max_{\eta^{(L)}} T^{\eta^{(L)}}(1, \mathcal{P}, \mathcal{P}). \quad (128)$$

Next, consider the case where the CSI available to each sender is given in terms of

$$E_{1,t} = E_{2,t} = 1(\tilde{S}_{1,t} < \tilde{S}_{2,t}). \quad (129)$$

Thus, $E_{1,t}$, $E_{2,t}$ identify the sender whose transmitted signal experiences a “deeper” fade at each time t . The corresponding “throughput,” denoted $T^{(><)}(\mathcal{P})$, can be seen to be achieved by a power-allocation policy given by

$$\begin{aligned} \mu_1(e_{1,t} = 0) &= 2\mathcal{P} & \mu_2(e_{2,t} = 0) &= 0 \\ \mu_1(e_{1,t} = 1) &= 0 & \mu_2(e_{2,t} = 1) &= 2\mathcal{P} \end{aligned} \quad (130)$$

upon using the fact that $\Pr(E_{j,t} = 1) = \frac{1}{2}$, $j = 1, 2$. In other words, the sender experiencing the deeper fade remains silent, while the other sender transmits at power $2\mathcal{P}$. This strategy is a variant of conventional time division multiple access (TDMA), where the senders take turns to transmit.¹⁰

In Fig. 5, we present a comparison, based on numerical computations, of $T^{(0)}(\mathcal{P})$, $T^{(1)}(\mathcal{P})$, $T^{(2)}(\mathcal{P})$, $T^{(\infty)}(\mathcal{P})$, $T^{(><)}(\mathcal{P})$, and $T^{(\text{TDMA})}(\mathcal{P})$ where $T^{(0)}(\mathcal{P})$ is the “throughput” when no CSI is available to the senders, while $T^{(\infty)}(\mathcal{P})$ is the “throughput” when each sender has perfect CSI [26]. Since the TVFC in (8) is interference-limited, it is not surprising that $T^{(><)}(\mathcal{P}) > T^{(2)}(\mathcal{P})$. Further, it is interesting to note

¹⁰Note that the conventional TDMA “throughput” $T^{(\text{TDMA})}(\mathcal{P})$, where each sender has perfect CSI, *i.e.*, $E_{1,t} = E_{2,t} = (\tilde{S}_{1,t}, \tilde{S}_{2,t})$, equals half the capacity of a single-sender TVFC with (input) power constraint $2\mathcal{P}$. Clearly, $T^{(\text{TDMA})}(\mathcal{P}) \leq T^{(><)}(\mathcal{P})$.

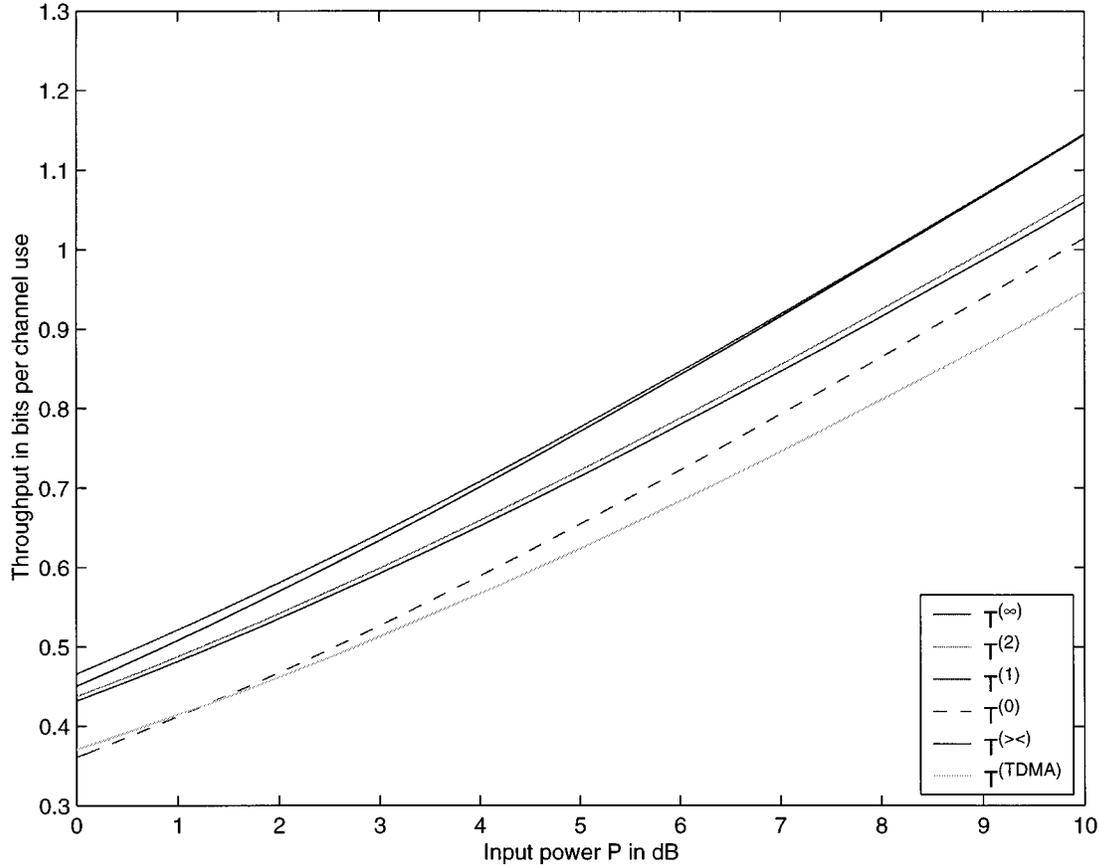


Fig. 5. The “throughput” of the two-sender TVFC in (8), as defined in (128), when various degrees of CSI are available to each sender.

that $T^{(><)}(\mathcal{P})$ is comparable to $T^{(\infty)}(\mathcal{P})$. While, $T^{(><)}(\mathcal{P})$ is achieved by a policy which lets the “better” sender alone transmit at each time instant, $T^{(\infty)}(\mathcal{P})$ is achieved, as shown in [26], by a similar policy combined with a step involving power adaptation based on fade values. That the latter step does not afford a significant improvement in “throughput” can be explained as follows. Loosely speaking, in both cases, the multiple-access channel at each time instant is tantamount to a single-sender channel, by virtue of said transmission policy. Power adaptation by the single transmitting sender does not lead to a significant increase in coding rate, which is redolent of the corresponding result in [11] for the single-sender TVFC in (10).

APPENDIX A PROOF OF LEMMA 2

For convenience, let A^n denote the collection of rvs

$$(X_1^n, X_2^n, E_1^n, E_2^n, D^n, F_1^n, F_2^n)$$

and a^n a realization of A^n . For each a^n with positive probability, we have

$$P_{Y|A}(y^n|a^n) = \frac{\sum_{s^n \in \mathcal{S}^n} P_{Y|AS}(y^n|a^n, s^n) P_{AS}(a^n, s^n)}{\sum_{s^n \in \mathcal{S}^n} P_{AS}(a^n, s^n)}. \quad (131)$$

Since a^n has positive probability, the numerator on the right-hand side of (131), upon using (65), (66), and proceeding as in (67), (68), is seen to equal

$$P_{F_1}(f_1^n) P_{F_2}(f_2^n) \sum_{s^n \in \mathcal{S}^n} W(y^n|x_1^n, x_2^n, s^n) P_S(s^n) \times 1(h_1(s^n) = e_1^n) 1(h_2(s^n) = e_2^n) 1(g(s^n) = D^n)$$

while the denominator likewise equals

$$P_{F_1}(f_1^n) P_{F_2}(f_2^n) \sum_{s^n \in \mathcal{S}^n} P_S(s^n) 1(h_1(s^n) = e_1^n) \times 1(h_2(s^n) = e_2^n) 1(g(s^n) = d^n).$$

Substitution into (131) followed by cancellation of terms involving f_1^n, f_2^n then shows that the left-hand side of (131) does not depend on f_1^n, f_2^n . This establishes (51).

In order to prove (52), we proceed in a similar manner as above. Specifically, for each collection $(x_1^n, e_1^n, d^n, f_1^n)$ with positive probability, we can express

$$P_{Y|X_1 E_1 D F_1}(y^n|x_1^n, e_1^n, d^n, f_1^n)$$

in terms of the ratio on the right-hand side of (131) with the summations in the numerator and denominator now being with respect to $s^n \in \mathcal{S}^n, x_2^n \in \mathcal{X}_2^n, e_2^n \in \mathcal{E}_2^n, f_2^n \in \mathcal{F}_2^n$. A cancellation of terms involving f_1^n then indicates that

$$P_{Y|X_1 E_1 D F_1}(y^n|x_1^n, e_1^n, d^n, f_1^n)$$

does not depend on f_1^n , thereby establishing (52).

APPENDIX B
PROOFS OF PROPOSITIONS 3 AND 4

Proof of Proposition 3: Let $(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)})$ be a length- n code with

$$P_{e, \bar{P}}(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)}) \leq \epsilon, \quad 0 < \epsilon < 1$$

where the second subscript in $P_{e, \bar{P}}$ makes explicit its dependence on \bar{P} . From the following straightforward inequality: for each $s \in \sigma_{\bar{P}}$

$$\begin{aligned} \Pr(\phi^{(n)}(Y^n, D^n) \neq (\bar{M}_1, \bar{M}_2) | S_1 = s) \\ \leq \frac{P_{e, \bar{P}}(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)})}{\min_{s' \in \sigma_{\bar{P}}} \bar{P}(s')} \end{aligned}$$

it follows that when $P_{S_1} = \bar{P}$, with $\sigma_{\bar{P}} \subseteq \sigma_{\bar{P}}$

$$P_{e, \bar{P}}(\theta_1^{(n)}, \theta_2^{(n)}, \phi^{(n)}) \leq \frac{\epsilon}{\min_{s \in \sigma_{\bar{P}}} \bar{P}(s)}$$

leading to the desired assertion. \square

Before we turn to the proof of Proposition 4, we first present the following simple and relevant technical result.

Lemma 4: For every sequence of pmfs $\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2, \underline{\mathbf{D}}, \underline{\mathbf{Y}}$, with the joint pmfs of F_1^n, F_2^n, D^n, Y^n given by (16) and (17), it holds that for every (finite) integer $k \geq 2$

$$\begin{aligned} \underline{I}(\underline{\mathbf{F}}_1 \wedge \underline{\mathbf{Y}} | \underline{\mathbf{F}}_2, \underline{\mathbf{D}}) &= \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | F_{2,k}^n, D_k^n)} \\ \underline{I}(\underline{\mathbf{F}}_2 \wedge \underline{\mathbf{Y}} | \underline{\mathbf{F}}_1, \underline{\mathbf{D}}) &= \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | F_{1,k}^n, D_k^n)} \\ \underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{Y}} | \underline{\mathbf{D}}) &= \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(Y_k^n | F_{1,k}^n, F_{2,k}^n, D_k^n)}{\Pr(Y_k^n | D_k^n)}. \end{aligned} \quad (132)$$

Proof: The following inequalities from [20] will be used in our proof. Consider a sequence of $\mathbb{R} \times \mathbb{R}$ -valued rvs $\{A_t, B_t\}_{t=1}^\infty$. Then

$$\begin{aligned} \underline{\lim-p} A_t + \underline{\lim-p} B_t &\leq \underline{\lim-p} (A_t + B_t) \\ &\leq \underline{\lim-p} A_t + \overline{\lim-p} B_t \end{aligned} \quad (133)$$

where $\overline{\lim-p} B_t$, the ‘‘limsup in probability,’’ of the \mathbb{R} -valued sequence of rvs $\{B_t\}_{t=1}^\infty$, is defined [20] as

$$\overline{\lim-p} B_t = \inf_{\alpha \in \mathbb{R}} \left\{ \alpha : \lim_t \Pr(B_t > \alpha) = 0 \right\}. \quad (134)$$

We shall only prove the last equality in (132); the proof of the remainder of the lemma is similar. Furthermore, it suffices to prove the last equality in (132) for $k = 2$. To this end, first observe that the joint pmf of F_1^n, F_2^n, D^n, Y^n in (16) and (17) implies that

$$\underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{Y}} | \underline{\mathbf{D}}) = \underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{D}}, \underline{\mathbf{Y}}) \quad (135)$$

$$\begin{aligned} \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(Y_2^n | F_{1,2}^n, F_{2,2}^n, D_2^n)}{\Pr(Y_2^n | D_2^n)} \\ = \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)}. \end{aligned} \quad (136)$$

Next, for every $n \geq 2$, the decomposition

$$\begin{aligned} \frac{1}{n} \log \frac{\Pr(D^n, Y^n | F_1^n, F_2^n)}{\Pr(D^n, Y^n)} \\ = \frac{1}{n} \log \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)} \\ + \frac{1}{n} \log \frac{\Pr(F_{1,1}, F_{2,1}, D_1, Y_1 | F_{1,2}^n, F_{2,2}^n, D_2^n, Y_2^n)}{\Pr(D_1, Y_1 | D_2^n, Y_2^n) \Pr(F_{1,1}, F_{2,1} | F_{1,2}^n, F_{2,2}^n)} \end{aligned}$$

leads, by (133), to

$$\begin{aligned} \underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{D}}, \underline{\mathbf{Y}}) \\ \leq \underline{\lim-p} \frac{1}{n} \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)} + \overline{\lim-p} \frac{1}{n} \\ \times \log \frac{\Pr(F_{1,1}, F_{2,1}, D_1, Y_1 | F_{1,2}^n, F_{2,2}^n, D_2^n, Y_2^n)}{\Pr(D_1, Y_1 | D_2^n, Y_2^n) \Pr(F_{1,1}, F_{2,1} | F_{1,2}^n, F_{2,2}^n)} \\ (137) \\ \leq \underline{\lim-p} \frac{1}{n} \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)} + \overline{\lim-p} \frac{1}{n} \\ \times \log \frac{1}{\Pr(D_1, Y_1 | D_2^n, Y_2^n) \Pr(F_{1,1}, F_{2,1} | F_{1,2}^n, F_{2,2}^n)} \\ \leq \underline{\lim-p} \frac{1}{n} \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)} + \overline{\lim-p} \frac{1}{n} \\ \times \log \frac{1}{\Pr(D_1, Y_1 | D_2^n, Y_2^n)} + \overline{\lim-p} \frac{1}{n} \\ \times \log \frac{1}{\Pr(F_{1,1}, F_{2,1} | F_{1,2}^n, F_{2,2}^n)}, \quad \text{by (133)}. \end{aligned} \quad (138)$$

The second term on the right-hand side of (137) can be further bounded above as

$$\begin{aligned} \overline{\lim-p} \frac{1}{n} \log \frac{1}{\Pr(D_1, Y_1 | D_2^n, Y_2^n)} \\ \leq \overline{\lim-p} \frac{1}{n} \log \frac{1}{\Pr(D_1, Y_1)} \\ - \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(D_1, Y_1 | D_2^n, Y_2^n)}{\Pr(D_1, Y_1)} \\ \leq \overline{\lim-p} \frac{1}{n} \log \frac{1}{\Pr(D_1, Y_1)} = 0 \end{aligned} \quad (139)$$

where the first inequality above follows from (133), and the last equality follows from the finiteness of the sets \mathcal{Y} and \mathcal{D} . Proceeding similarly to bound the third term on the right-hand side of (137) as

$$\overline{\lim-p} \frac{1}{n} \log \frac{1}{\Pr(F_{1,1}, F_{2,1} | F_{1,2}^n, F_{2,2}^n)} \leq 0 \quad (140)$$

we then obtain from (137), (139), and (140) that

$$\underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{D}}, \underline{\mathbf{Y}}) \leq \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)}.$$

The reverse inequality, viz.

$$\underline{I}(\underline{\mathbf{F}}_1, \underline{\mathbf{F}}_2 \wedge \underline{\mathbf{D}}, \underline{\mathbf{Y}}) \geq \underline{\lim-p} \frac{1}{n} \log \frac{\Pr(D_2^n, Y_2^n | F_{1,2}^n, F_{2,2}^n)}{\Pr(D_2^n, Y_2^n)}$$

is also true by [20, Theorem 8]. Combined with the observations in (135) and (136), the proof of the last equality in (132) is now complete. \square

Proof of Proposition 4: The proof is based on the simple idea that since $E_{1,1}$ and $E_{2,1}$ are constants a.s., loosely speaking, reliable communication can also be assured “beginning” at time $t = 2$, without any loss of rate. Precisely, for the initial pmf $P_{S_1} = \bar{P}$, by Proposition 1 and Lemma 4, the capacity region $\mathcal{C}_{\bar{P}}$ is given by (141), shown at the bottom of the page, where the joint pmf of the rvs $F_{1,2}^n, F_{2,2}^n, D_2^n, Y_2^n$ obtains from (16) and (17) with $P_{S_1} = \bar{P}$. Now, since \bar{P} is such that $E_{1,1}$ and $E_{2,1}$ are constants a.s., the union in (141) is, in effect, over rvs $\{F_{1,2}^n\}_{n=2}^\infty, \{F_{2,2}^n\}_{n=2}^\infty$, where $F_{j,t}$, $t \geq 2$, is an $\mathcal{F}_j(t-1)$ -valued rv, $j = 1, 2$. This observation, in conjunction with the memorylessness property of the TVMAC in (23), yields that the right-hand side of (141) can be expressed as

$$\bigcup_{(\mathbf{E}_1, \mathbf{E}_2)} \left\{ \begin{array}{l} 0 \leq R_1 \leq I(\mathbf{E}_1 \wedge \mathbf{Y} | \mathbf{E}_2, \mathbf{D}) \\ (R_1, R_2): 0 \leq R_2 \leq I(\mathbf{E}_2 \wedge \mathbf{Y} | \mathbf{E}_1, \mathbf{D}) \\ R_1 + R_2 \leq I(\mathbf{E}_1, \mathbf{E}_2 \wedge \mathbf{Y} | \mathbf{D}) \end{array} \right\} \quad (142)$$

where the joint pmf of the rvs F_1^n, F_2^n, D^n, Y^n is given by (16) and (17) with $P_{S_1} = \bar{P} \circ \Gamma$. By Proposition 1, the region in (142) equals $\mathcal{C}_{\bar{P} \circ \Gamma}$, therefore, $\mathcal{C}_{\bar{P}} = \mathcal{C}_{\bar{P} \circ \Gamma}$. The assertion in Proposition 4 now readily follows upon noting that $\sigma_{\bar{P} \circ \Gamma} \supseteq \mathcal{B}$ and by a use of Proposition 3.

APPENDIX C PROOF OF PROPOSITION 5

The main idea behind the proof of Proposition 5 is already contained in its specialization to the case of a single-sender memoryless TVC in (28). We shall, therefore, restrict ourselves to this special case as its treatment is less cumbersome notationally. Specifically, considering the single-sender memoryless TVC in (28) under the hypothesis of Theorem 2 on $\{S_t\}_{t=1}^\infty$, for any fixed positive integer k , and any sequence of pmfs \underline{F} such that F_t is a $\mathcal{F}(k)$ -valued rv for $t \geq k$, we shall show that

$$\liminf_n \frac{1}{n} I(F^n \wedge Y^n | D^n)$$

does not depend on P_{S_1} , where the joint pmf of F^n, D^n, Y^n is as specified in (19).

To that end, note first that the joint pmf of F^n, D^n, Y^n in (19) implies that

$$\liminf_n \frac{1}{n} I(F^n \wedge Y^n | D^n) = \liminf_n \frac{1}{n} I(F^n \wedge D^n, Y^n).$$

From [6, Ch. 4, Lemma 1], we also have

$$|I(F^n \wedge D^n, Y^n) - I(F^n \wedge D^n, Y^n | S_1)| \leq \log |\mathcal{S}|$$

and it follows that

$$\liminf_n \frac{1}{n} I(F^n \wedge D^n, Y^n) = \liminf_n \frac{1}{n} I(F^n \wedge D^n, Y^n | S_1). \quad (143)$$

Consequently, it suffices to show that

$$J(s) \triangleq \liminf_n \frac{1}{n} I(F^n \wedge D^n, Y^n | S_1 = s)$$

is the same for every $s \in \mathcal{S}$.

Since the state process $\{S_t\}_{t=1}^\infty$ is a TIAMC, given $\epsilon > 0$, there exists a positive integer $N_0(\epsilon) \geq k$, such that for every $j \geq N_0$, and $r, s \in \mathcal{S}$

$$\sum_{s_{j-k+2} \in \mathcal{S}} |\Pr(S_{j-k+2} = s_{j-k+2} | S_1 = s) - \Pr(S_{j-k+2} = s_{j-k+2} | S_1 = r)| < \epsilon$$

hence it also holds that

$$\sum_{s_{j-k+2}^{j+1} \in \mathcal{S}^k} \left| \Pr(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} | S_1 = s) - \Pr(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} | S_1 = r) \right| < \epsilon. \quad (144)$$

Fix $j \geq N_0$. Let l be a positive integer, and set $n = j + l$. For notational convenience, denote F^j and F_{j+1}^n by \underline{E}_1 and \underline{E}_2 , respectively; similarly, define $\underline{D}_1, \underline{D}_2, \underline{Y}_1$, and \underline{Y}_2 . Then, from the chain rule for mutual information we get

$$\begin{aligned} & I(\underline{E}_1, \underline{E}_2 \wedge \underline{D}_1, \underline{Y}_1, \underline{D}_2, \underline{Y}_2, \underline{D}_2 | S_1 = s) \\ &= I(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 | S_1 = s) \\ &+ I(\underline{E}_1 \wedge \underline{D}_2, \underline{Y}_2 | \underline{E}_2, S_1 = s) \\ &+ I(\underline{E}_1, \underline{E}_2 \wedge \underline{D}_1, \underline{Y}_1 | \underline{D}_2, \underline{Y}_2, S_1 = s). \end{aligned} \quad (145)$$

The second and third terms on the right-hand side of (145) are bounded above by $j \log |\mathcal{F}(k)|$ and $j \log |\mathcal{D}||\mathcal{Y}|$, respectively, while the first term can be bounded as

$$\begin{aligned} & I(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 | S_1 = s) \\ & \leq I(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 | S_{j-k+2}^{j+1}, S_1 = s) + k \log |\mathcal{S}| \end{aligned}$$

whence, it follows that

$$\begin{aligned} & I(\underline{E}_1, \underline{E}_2 \wedge \underline{D}_1, \underline{Y}_1, \underline{D}_2, \underline{Y}_2 | S_1 = s) \\ & \leq I(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 | S_{j-k+2}^{j+1}, S_1 = s) \\ & + k \log |\mathcal{S}| + j \log (|\mathcal{F}(k)||\mathcal{D}||\mathcal{Y}|) \end{aligned}$$

$$\mathcal{C}_{\bar{P}} = \bigcup_{\{F_{1,2}^n\}_{n=2}^\infty, \{F_{2,2}^n\}_{n=2}^\infty} \left\{ \begin{array}{l} 0 \leq R_1 \leq \liminf-p \frac{1}{n} \log \frac{\Pr(Y_2^n | F_{1,2}^n, F_{2,2}^n, D_2^n)}{\Pr(Y_2^n | F_{1,2}^n, D_2^n)} \\ (R_1, R_2): 0 \leq R_2 \leq \liminf-p \frac{1}{n} \log \frac{\Pr(Y_2^n | F_{1,2}^n, F_{2,2}^n, D_2^n)}{\Pr(Y_2^n | F_{2,2}^n, D_2^n)} \\ R_1 + R_2 \leq \liminf-p \frac{1}{n} \log \frac{\Pr(Y_2^n | F_{1,2}^n, F_{2,2}^n, D_2^n)}{\Pr(Y_2^n | D_2^n)} \end{array} \right\} \quad (141)$$

$$\begin{aligned}
&= \sum_{s_{j-k+2}^{j+1} \in \mathcal{S}^k} \Pr \left(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \middle| S_1 = s \right) \\
&\quad \times I \left(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 \middle| S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \right) \\
&\quad + k \log |\mathcal{S}| + j \log(|\mathcal{F}(k)| |\mathcal{D}| |\mathcal{Y}|). \tag{146}
\end{aligned}$$

The equality above results from the memorylessness property of the TVC in (28), and the assumption that F_t is an $\mathcal{F}(k)$ -valued rv for $t \geq k$. Proceeding similarly, we obtain the lower bound

$$\begin{aligned}
&I(\underline{E}_1, \underline{E}_2 \wedge \underline{Y}_1, \underline{D}_1, \underline{Y}_2, \underline{D}_2 | S_1 = r) \\
&\geq \sum_{s_{j-k+2}^{j+1} \in \mathcal{S}^k} \Pr \left(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \middle| S_1 = r \right) \\
&\quad \times I \left(\underline{E}_2 \wedge \underline{Y}_2, \underline{D}_2 \middle| S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \right) - k \log |\mathcal{S}|. \tag{147}
\end{aligned}$$

Combining (146) and (147), we obtain

$$\begin{aligned}
&\frac{1}{n} |I(\underline{E}_1, \underline{E}_2 \wedge \underline{D}_1, \underline{Y}_1, \underline{D}_2, \underline{Y}_2 | S_1 = s) \\
&\quad - I(\underline{E}_1, \underline{E}_2 \wedge \underline{D}_1, \underline{Y}_1, \underline{D}_2, \underline{Y}_2 | S_1 = r)| \\
&\leq \frac{1}{n} \left[\sum_{s_{j-k+2}^{j+1} \in \mathcal{S}^k} \left| \Pr \left(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \middle| S_1 = s \right) \right. \right. \\
&\quad \left. \left. - \Pr \left(S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \middle| S_1 = r \right) \right| \right. \\
&\quad \times I \left(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 \middle| S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1} \right) \\
&\quad \left. + j \log(|\mathcal{F}(k)| |\mathcal{D}| |\mathcal{Y}|) + 2k \log |\mathcal{S}| \right] \\
&\leq \frac{1}{n} [\epsilon l \log |\mathcal{D}| |\mathcal{Y}| + j \log(|\mathcal{F}(k)| |\mathcal{D}| |\mathcal{Y}|) + 2k \log |\mathcal{S}|] \\
&\leq \epsilon \log |\mathcal{D}| |\mathcal{Y}| + \epsilon \tag{148}
\end{aligned}$$

for all sufficiently large n , where we have used (144) and the bound

$$I(\underline{E}_2 \wedge \underline{D}_2, \underline{Y}_2 | S_{j-k+2}^{j+1} = s_{j-k+2}^{j+1}) < l \log |\mathcal{D}| |\mathcal{Y}|$$

to arrive at the second inequality. Since ϵ is arbitrary, it is immediate that $J(s) = J(r)$ for every $r, s \in \mathcal{S}$.

ACKNOWLEDGMENT

The authors gratefully acknowledge that this paper has benefited from discussions with G. Caire, S. Shamai, and Associate Editor I. Csiszár.

REFERENCES

- [1] J. Wolfowitz, *Coding Theorems of Information Theory*, 3rd ed. Berlin, Germany: Springer-Verlag, 1978.
- [2] A. Lapidoth and I. E. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 973–983, May 1998.
- [3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [4] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2148–2177, Oct. 1998.
- [5] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.
- [6] R. Gallager, *Information Theory and Reliable Communications*. New York: Springer-Verlag, 1969.
- [7] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, no. 4, pp. 1209–1220, 1958.
- [8] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: Wiley, 1957.
- [9] F. Jelinek, "Indecomposable channels with side information at the transmitter," *Inform. Contr.*, vol. 8, pp. 36–55, 1965.
- [10] H. 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.
- [11] A. Goldsmith, "Design and performance of high-speed communication systems," Ph.D. dissertation, Univ. California, Berkeley, 1994.
- [12] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *Proc. Inst. Elec. Eng.—Pt. I*, vol. 139, pp. 113–117, Apr. 1992.
- [13] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inform. Theory*, vol. 45, pp. 761–771, Mar. 1999.
- [14] 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.
- [15] D. Tse and S. Hanly, "Multiaccess fading channels—Parts I: Polymatroidal structure, optimal resource allocation and throughput capacities; Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2831, Nov. 1998.
- [16] R. Gallager, *Communications and Cryptography—Two Sides of One Tapestry*. Norwell, MA: Kluwer, 1994, ch. An Inequality on the Capacity Region of Multiaccess Fading Channels, pp. 129–139.
- [17] S. Shamai (Shitz), L. Ozarow, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.
- [18] S. Shamai (Shitz) and A. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels: Parts I & II," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1911, Nov. 1997.
- [19] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2773–2795, Nov. 1998.
- [20] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
- [21] T. S. Rappaport, *Wireless Communications, Principles and Practice*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [22] E. N. Gilbert, "Capacity of burst-noise channels," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.
- [23] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, pp. 1977–1997, Sept. 1963.
- [24] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert–Elliott channel," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.
- [25] A. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.
- [26] R. Knopp and P. Humblet, "Information capacity and power-control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Communications (ICC'95)*, Seattle, WA, June 1995, pp. 331–335.