On the Capacity of Memoryless Finite-State Multiple Access Channels with Asymmetric State Information at the Encoders[1]

Giacomo Como[2] and Serdar Yüksel [3]

***Abstract*—A single-letter characterization is provided for the capacity region of finite-state multiple access channels, when the channel state process is an independent and identically distributed sequence, the transmitters have access to partial (quantized) state information, and complete channel state information is available at the receiver. The partial channel state information is assumed to be asymmetric at the encoders. As a main contribution, a tight converse coding theorem is presented. A simple proof of achievability is reported as well. The difficulties associated with the case when the channel state has memory are discussed and connections to decentralized stochastic control theory are presented.**

I. Introduction and Literature Review

Wireless communication channels and Internet type networks are examples of channels where the channel characteristics are time-varying. In wireless channels, the mobility of users and changes in landscape as well as interference may lead to temporal variations in the channel quality. In network applications, user demand and node failure may affect the channel reliability. Such channel variation models may include fast fading and slow fading; in fast fading, the channel state is assumed to be changing for each use of the channel. On the other hand, in slow fading, the channel is assumed to be constant for each finite block length.

In such problems, the channel state can be transmitted to the encoders either explicitly, or through output feedback. Typically the feedback is not perfect, that is the encoder has only partial information regarding the state or the output variables. The present paper studies a particular case, finite-state multiple access channels (MACs), where partial channel state information (CSI) is provided to the encoders causally. What makes such setup particularly interesting is the fact that the partial CSI available to the two transmitters is in general *asymmetric*, i.e., none of the transmitters' CSI contains the CSI available to the other one. On the other hand, we assume that the receiver has access to perfect state information.

A single-letter characterization of the capacity region is provided for the case of independent and identically distributed (i.i.d.) channel state sequences. As we shall review shortly, results in the literature have already provided achievability results for such problems. The main contribution of this paper consists in providing a tight converse theorem, in addition to a simple discussion of achievability of the capacity region. Our proof of the converse theorem involves showing that restricting to encoders using only the quantized CSI on the current state does not cause any loss of optimality with respect to the most general class of admissible encoders using all the quantized CSI causally observed until a given time.

The problem at hand can be thought of as a decentralized stochastic control problem. We shall elaborate on this connection in the concluding section, where we shall also discuss in what our arguments fail when trying to extend them to a proof of the converse theorem for finite-state MACs with memory, and asymmetric CSI at the transmitters.

Let us now present a brief literature review. Capacity with partial channel state information at the transmitter is related to the problem of coding with unequal side information at the encoder and the decoder. The capacity of memoryless channels, with various cases of state information being available at neither, either or both the transmitter and receiver, has been studied in [16] and [8]. Reference [17] develops a stochastic control framework for the computation of the capacity of channels with memory and complete noiseless output feedback via the properties of the directed mutual information. Reference [9] considers fading channels with perfect channel state information at the transmitter, and shows that with instantaneous and perfect state information, the transmitter can adjust the data rates for each channel state to

[1]A preliminary version of this paper was presented at the 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt) as part of the Control over Communication Channels (ConCom) Workshop, Seoul, South Korea, in June 2009, and has appeared in its proceedings.

[2]G. Como is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, MA, USA. Email: giacomo@mit.edu.

[3]S. Yüksel is with the Mathematics and Engineering Program, Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6. Email: yuksel@mast.queensu.ca. S. Yüksel's research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

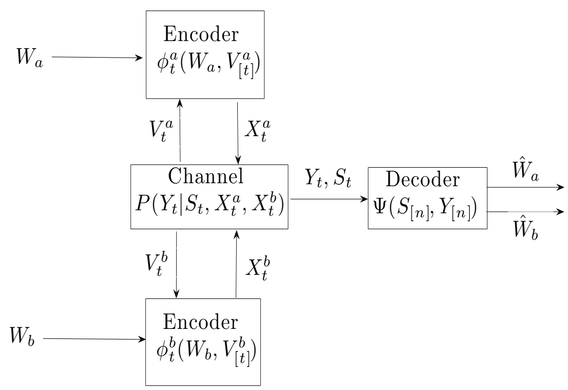

Fig. 1: Finite-state multiple access channel with asymmetric partial state information at the transmitters.

maximize the average transmission rate. Viswanathan [19] relaxes this assumption of perfect instantaneous state information, and studies the capacity of Markovian channels with delayed information. Reference [5] studies the capacity of Markov channels with perfect causal state information. The capacity of Markovian, finite-state channels with quantized state information available at the transmitter is studied in [20].

The works most closely related to ours are [7] and [13]. In [7], the capacity of general finite-state MACs with different levels of causal CSI at the transmitters is characterized in terms of non-single-letter formulas. Moreover, single-letter characterizations are provided for the capacity of finite-state MACs when the decoder has perfect CSI and the encoders are restricted to use only a finite window of, possibly limited, CSI; the capacity region without any such restriction is recovered in the limit of large window size. Reference [13] develops a general framework for approximating, and possibly characterizing, the capacity of channels with causal, and non-causal CSI: in particular, Theorem 4 therein provides a single-letter characterization of the capacity region of a MAC with independent CSI at the transmitters. With respect to [7], [13], the present paper considers the somewhat simpler case of a MAC with i.i.d. state, where the encoders have causal, asymmetric, partial CSI, which is obtained through fixed quantizers acting componentwise. In contrast to [7], a single-letter expression for the capacity region is obtained in this case without any finite window restriction on the CSI available to the transmitters, while, differently from [13], the CSI available to the transmitters is not assumed to be independent. Recent related work also includes [14], providing an infinite-dimensional characterization for the capacity region for Multiple Access Channels with feedback, and [4], studying the case of MAC channels where the encoders have access to coded non-causal state information.

The rest of the paper is organized as follows. In Section II a formal statement of the problem and the main results are presented, consisting in a single-letter characterization of the capacity region of finite-state MACs with i.i.d. state. Section III contains the proof of achievability of the capacity region, while Section IV presents a proof of the converse coding theorem. Finally, in Section V, we discuss the issues arising when trying to generalize our arguments to the memory case, and present some final remarks on the connections of this problem with the decentralized stochastic control literature.

II. Capacity of i.i.d. finite-state MAC with asymmetric partial CSI

In the following, we shall present some notation, before formally stating the problem. For a vector v, and a positive integer i, v_i will denote the i-th entry of v, while $v_{[i]} = (v_1, \dots, v_i)$ will denote the vector of the first i entries of v. Following a common convention, capital letters will be used to denote random variables (r.v.s), and small letters denote particular realizations. We shall use the standard notation $\mathrm{H}(\,\cdot\,)$, and $\mathrm{I}(\,\cdot\,;\,\cdot\,)$ (respectively $\mathrm{H}(\,\cdot\,|\,\cdot\,)$, and $\mathrm{I}(\,\cdot\,;\,\cdot\,|\,\cdot\,)$) for the (conditional) entropy and mutual information of r.v.s. With a slight abuse of notation, for $0 \leq x \leq 1$, we shall write $\mathrm{H}(x)$ for the entropy of x. For a finite set $\mathcal{A}$, $\mathcal{P}(\mathcal{A})$ will denote the simplex of probability distributions over $\mathcal{A}$. Finally, for a positive integer n, we shall denote by $\mathcal{A}^{(n)} := \bigcup_{0 \leq s < n} \mathcal{A}^s$ the set of $\mathcal{A}$-strings of length smaller than n. [1]

We shall consider a finite-state MAC with two transmitters, indexed by $i \in \{a, b\}$, and one receiver. Transmitter i aims at reliably communicating a message W_i, uniformly distributed over some finite message set $\mathcal{W}_i$, to the receiver. The two messages W_a and W_b are assumed to be mutually independent. We shall use the notation $W := (W_a, W_b)$ for the vector of the two messages.

The channel state process is modeled by a sequence $S = \{S_t : t = 1, 2, \dots\}$ of independent, identically distributed (i.i.d.) r.v.s, taking values in some finite-state space $\mathcal{S}$, and independent from W; the probability distribution of any S_t is denoted by $P(\,\cdot\,) \in \mathcal{P}(\mathcal{S})$. The two encoders have access to causal, partial state information: at each time $t \geq 1$, encoder i observes $V_t^{(i)} = q_i(S_t)$, where $q_i : \mathcal{S} \to \mathcal{V}_i$ is a quantizer modeling the imperfection in the state information. We shall denote by $V_t := (V_t^{(a)}, V_t^{(b)})$ the vector of quantized

[1] This includes the empty string, conventionally assumed to be the only element of $\mathcal{A}^0$.

state observations, taking values in $\mathcal{V} := \mathcal{V}_a \times \mathcal{V}_b$. The channel input of encoder i at time t, $X_t^{(i)}$, takes values in a finite set $\mathcal{X}_i$, and is assumed to be a function of the locally available information $(W_i, V_{[t]}^{(i)})$. The symbol $X_t = (X_t^{(a)}, X_t^{(b)})$ will be used for the vector of the two channel inputs at time t, taking values in $\mathcal{X} := \mathcal{X}_a \times \mathcal{X}_b$. The channel output at time t, Y_t, takes values in a finite set $\mathcal{Y}$; its conditional distribution satisfies

$$\mathbb{P}(Y_t = y | W = w, X_{[t]} = x_{[t]}, S_{[t]} = s_{[t]}) = P(y_t | s_t, x_t) \tag{1}$$

where, for any $s \in \mathcal{S}$, and $x \in \mathcal{X}$, $P(\,\cdot\,|s,x) \in \mathcal{P}(\mathcal{Y})$ is an output probability distribution. Finally, the decoder is assumed to have access to perfect causal state information (which may be known causally or non-causally); the estimated message pair will be denoted by $\hat{W} = (\hat{W}_a, \hat{W}_b)$.

We now present the class of transmission systems.

Definition 1: For a rate pair $R = (R_a, R_b)$, a block-length $n \geq 1$, and a target error probability $\varepsilon \geq 0$, an (R, n, ϵ)-coding scheme consists of two sequences of functions

$$\{\phi_t^{(i)} : \mathcal{W}_i \times \mathcal{V}_i^t \to \mathcal{X}_i\}_{1 \leq t \leq n} \,,$$

and a decoding function

$$\psi : \mathcal{S}^n \times \mathcal{Y}^n \to \mathcal{W}_a \times \mathcal{W}_b \,,$$

such that, for $i \in \{a, b\}$, $1 \leq t \leq n$:

- $|\mathcal{W}_i| \geq 2^{R_i n}$;
- $X_t^{(i)} = \phi_t^{(i)}(W_i, V_{[t]}^{(i)})$;
- $\hat{W} := \psi(S_{[n]}, Y_{[n]})$;
- $\mathbb{P}(\hat{W} \neq W) \leq \varepsilon$.

We now proceed with the characterization of the capacity region.

Definition 2: A rate pair $R = (R_a, R_b)$ is achievable if, for all $\varepsilon > 0$, there exists, for some $n \geq 1$, an (R, n, ε)-coding scheme. The capacity region of the finite-state MAC is the closure of the set of all achievable rate pairs.

We now introduce what we call *memoryless stationary team policies* and their associated rate regions.

Definition 3: A memoryless stationary team policy is a family

$$\pi = \{\pi_i(\,\cdot\,|v_i) \in \mathcal{P}(\mathcal{X}_i) |\, i \in \{a, b\},\, v_i \in \mathcal{V}_i\} \tag{2}$$

of probability distributions on the two channel input sets conditioned on the quantized observation of each transmitter. For every memoryless stationary team policy π, $\mathcal{R}(\pi)$ will denote the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$\begin{array}{rclcl} 0 & \leq & R_a & < & \mathrm{I}(X_a; Y | X_b, S) \\ 0 & \leq & R_b & < & \mathrm{I}(X_b; Y | X_a, S) \\ 0 & \leq & R_a + R_b & < & \mathrm{I}(X; Y | S) \,, \end{array} \tag{3}$$

where S, $X = (X_a, X_b)$, and Y, are r.v.s taking values in $\mathcal{S}$, $\mathcal{X}$, and $\mathcal{Y}$, respectively, and whose joint probability distribution

$$\nu(s, x, y) := P(S = s, X = x, Y = y)$$

factorizes as

$$\nu(s, x, y) = P(s) \pi_a(x_a | q_a(s)) \pi_b(x_b | q_b(s)) P(y | s, x) \,. \tag{4}$$

We can now state the main result of the paper.

Theorem 4: The achievable rate region is given by

$$\overline{\mathrm{co}} \left(\bigcup\nolimits_\pi \mathcal{R}(\pi) \right) \,,$$

the closure of the convex hull of the rate regions associated to all possible memoryless stationary team policies π as in (2).

In Section III we shall prove the direct part of Theorem 4, namely that every rate pair $R \in \overline{\mathrm{co}} \left(\cup_\pi \mathcal{R}(\pi) \right)$ is achievable. In Section IV we shall prove the converse part, i.e. that no rate pair $R \in \mathbb{R}_+^2 \setminus \overline{\mathrm{co}} \left(\cup_\pi \mathcal{R}(\pi) \right)$ is achievable.

III. Achievability of the capacity region

The result on achievability is known, and one could derive it from the arguments in [7]. For convenience, we present a brief discussion, with a different approach. Such an approach was suggested at the beginning of [13, Sect. VI], and consists in considering an equivalent MAC having as input mappings form the CSI information available at the transmitters to the original MAC's input.

Specifically, we shall consider an equivalent memoryless MAC having output space $\mathcal{Z} := \mathcal{S} \times \mathcal{Y}$ coinciding with the product of the state and output space of the original MAC, input spaces $\mathcal{U}_i := \{u_i : \mathcal{V}_i \to \mathcal{X}_i\}$, for $i \in \{a, b\}$, and transition probabilities

$$Q(z | u_a, u_b) := P(s) P(y | u_a(q_a(s)), u_b(q_b(s))) \,,$$

where $z = (s, y)$. A coding scheme for such a MAC consists of a pair of encoders $f^{(i)} : \mathcal{W}_i \to \mathcal{U}_i^n$, $i \in \{a, b\}$, and a decoder $g : \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{W}_a \times \mathcal{W}_b$. To any such coding scheme it is natural to associate a coding scheme for the original finite-state MAC, by defining the encoders

$$\phi_t^{(i)} : \mathcal{W}_i \times \mathcal{V}_i^t \to \mathcal{X}_i \,, \quad \phi_t^{(i)}(w_i, v_{[t]}) = [f^{(i)}(w_i)](v_t^{(i)})$$

and letting the decoder $\psi : \mathcal{Y}^n \times \mathcal{S}^n \to \mathcal{W}_a \times \mathcal{W}_b$ coincide with g. It is not hard to verify that the probability measure induced on the product space $\mathcal{W}_a \times \mathcal{W}_b \times \mathcal{S}^n \times \mathcal{Y}^n$ by the coding scheme $(f^{(a)}, f^{(b)}, g)$ and the memoryless MAC Q coincides with that induced by the corresponding coding scheme $(\phi_t^{(a)}, \phi_t^{(b)}, \psi)$ and the finite-state MAC P. Hence, in this way, to any (R, n, ε)-coding scheme on the memoryless MAC Q, it is possible to associate an (R, n, ε)-coding scheme $(\phi_t^{(a)}, \phi_t^{(b)}, \psi)$ on the original finite-state MAC P.

Now, let $\mu_a \in \mathcal{P}(\mathcal{U}_a)$, and $\mu_b \in \mathcal{P}(\mathcal{U}_b)$, be probability distributions on the input spaces of the new memoryless MAC, and fix an arbitrary rate pair $R = (R_a, R_b)$, such that

$$\begin{array}{rcl} R_a & < & \mathrm{I}(U_a; Z|U_b) \\ R_b & < & \mathrm{I}(U_b; Z|U_a) \\ R_a + R_b & < & \mathrm{I}(U; Z)\,, \end{array} \tag{5}$$

where $U = (U_a, U_b)$ and Z are random variables whose joint distribution factorizes as

$$P(U_a, U_b, Z) = \mu_a(U_a)\mu_b(U_b)Q(Z|U_a, U_b)\,. \tag{6}$$

For a positive integer n, let $f_a^{(n)} : \mathcal{W}_a^{(n)} \to \mathcal{U}_a^n$, and $f_b^{(n)} : \mathcal{W}_b^{(n)} \to \mathcal{U}_b^n$ be random encoders, with $|\mathcal{W}_a^{(n)}| = \lceil \exp(R_a n) \rceil$, $|\mathcal{W}_b^{(n)}| = \lceil \exp(R_b n) \rceil$, and

$$\left\{ f_a^{(n)}(w_a), f_b^{(n)}(w_b) :\, w_a \in \mathcal{W}_a^{(n)}, w_b \in \mathcal{W}_b^{(n)} \right\}$$

is a collection of independent r.v.s, with $f_i^{(n)}(w_i)$ taking values in $\mathcal{U}_i^n$ with product distribution $\mu_i \otimes \ldots \otimes \mu_i$, for each $i \in \{a, b\}$ and $w_i \in \mathcal{W}_i$. Then, it follows from the direct coding theorem for memoryless MACs [6, Th.3.2, p. 272] that the average error probability of such a code ensemble converges to zero as n grows large.

Now, we apply the arguments above to the special class of probability distributions $\mu_i \in \mathcal{P}(\mathcal{U}_i) = \mathcal{P}(\mathcal{X}_i^{\mathcal{S}_i})$ with the product structure

$$\mu_i(u_i) = \prod_{v_i \in \mathcal{V}_i} \pi_i(u_i(v_i)|v_i)\,, \quad u_i : \mathcal{V}_i \to \mathcal{X}_i\,, \tag{7}$$

where $i \in \{a, b\}$, and π is some memoryless stationary team policy, as in (2). Observe that, for such μ_a and μ_b, to any triple of r.v.s (U_a, U_b, Z), with joint distribution as in (6), one can naturally associate random variables S, $X_a := U_a(q_a(S))$, $X_b := U_b(q_b(S))$, and Y, whose joint probability distribution satisfies (4). Moreover, it can be readily verified that

$$\begin{array}{rcl} \mathrm{I}(X_a; Y|S, X_b) & = & \mathrm{I}(U_a; Z|U_b) \\ \mathrm{I}(X_b; Y|S, X_a) & = & \mathrm{I}(U_b; Z|U_a) \\ \mathrm{I}(X; Y|S) & = & \mathrm{I}(U; Z)\,. \end{array} \tag{8}$$

Hence, if a rate pair $R = (R_a, R_b)$ belongs to the rate region $\mathcal{R}(\pi)$ associated to some memoryless stationary team policy π (i.e. if it satisfies (3)), that R satisfies (5) for the product probability distributions μ_a, μ_b defined by (7). As observed above, the direct coding theorem for memoryless MACs implies that such a rate pair is achievable on the MAC Q. This in turn implies that the rate pair is achievable on the original finite-state MAC P. The proof of achievability of the capacity region $\overline{\mathrm{co}}(\cup_\pi \mathcal{R}(\pi))$ then follows from a standard time-sharing principle (see, e.g., [6, Lemma 2.2, p.272]).

IV. Converse to the coding theorem

In this section, we shall prove that no rate outside $\overline{\mathrm{co}}(\cup_\pi \mathcal{R}(\pi))$ is achievable. Lemma 5 shows that any achievable rate pair can be approximated by convex combinations of (conditional) mutual information terms. For $\varepsilon \in [0, 1]$, define

$$\eta(\varepsilon) := \frac{\varepsilon}{1-\varepsilon} \log |\mathcal{Y}| + \frac{\mathrm{H}(\varepsilon)}{1-\varepsilon}\,, \tag{9}$$

and observe that

$$\lim_{\varepsilon \to 0} \eta(\varepsilon) = 0\,. \tag{10}$$

For every $t \geq 1$, and $\boldsymbol{\sigma} \in \mathcal{S}^{t-1}$, define

$$\alpha_{\boldsymbol{\sigma}} := \frac{1}{n} \mathbb{P}(S_{[t-1]} = \boldsymbol{\sigma})\,. \tag{11}$$

Clearly, $\alpha_{\boldsymbol{\sigma}} \geq 0$, and

$$\sum_{\boldsymbol{\sigma} \in \mathcal{S}^{(n)}} \alpha_{\boldsymbol{\sigma}} = \frac{1}{n} \sum_{1 \leq t \leq n} \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{t-1}} \mathbb{P}(S_{[t-1]} = \boldsymbol{\sigma}) = 1\,. \tag{12}$$

Lemma 5: For a rate pair $R \in \mathbb{R}_+^2$, a block-length $n \geq 1$, and a constant $\varepsilon \in (0, 1/2)$, assume that there exists a (R, n, ε)-code. Then,

$$R_a + R_b \leq \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{(n)}} \alpha_{\boldsymbol{\sigma}}\, \mathrm{I}(X_t; Y_t|S_t, S_{[t-1]} = \boldsymbol{\sigma}) + \eta(\varepsilon) \tag{13}$$

$$R_a \leq \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{(n)}} \alpha_{\boldsymbol{\sigma}}\, \mathrm{I}(X_t^{(a)}; Y_t|X_t^{(b)}, S_t, S_{[t-1]} = \boldsymbol{\sigma}) + \eta(\varepsilon)\,. \tag{14}$$

$$R_b \leq \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{(n)}} \alpha_{\boldsymbol{\sigma}}\, \mathrm{I}(X_t^{(b)}; Y_t|X_t^{(a)}, S_t, S_{[t-1]} = \boldsymbol{\sigma}) + \eta(\varepsilon)\,. \tag{15}$$

Proof: By Fano's inequality we have the following estimate of the residual uncertainty on the messages given the full decoder's observation

$$H(W|Y_{[n]}; S_{[n]}) \quad \leq \quad \mathrm{H}(\varepsilon) + \varepsilon \log(|\mathcal{W}_a||\mathcal{W}_b|)\,. \tag{16}$$

For $1 \leq t \leq n$, we consider the conditional mutual information term

$$\Delta_t := \mathrm{I}(W; Y_t, S_{t+1}|Y_{[t-1]}, S_{[t]})\,,$$

and observe that

$$\begin{array}{rcl} \sum_{1 \leq t \leq n} \Delta_t & = & \mathrm{H}(W|S_1) - \mathrm{H}(W|S_{[n+1]}, Y_{[n]}) \\ & = & \log(|\mathcal{W}_a||\mathcal{W}_b|) - \mathrm{H}(W|S_{[n]}, Y_{[n]})\,, \end{array} \tag{17}$$

since the initial state S_1 is independent of the message pair W, and the final state S_{n+1} is conditionally independent of W given $(S_{[n]}, Y_{[n]})$. On the other hand, using the conditional independence of W from S_{t+1} given $(S_{[t]}, Y_{[t]})$, one gets

$$\begin{aligned}\Delta_t &= \mathrm{I}(W; Y_t, S_{t+1} | Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{I}(W; Y_t | Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{H}(Y_t | Y_{[t-1]}, S_{[t]}) - \mathrm{H}(Y_t | W, Y_{[t-1]}, S_{[t]}) \\ &\leq \mathrm{H}(Y_t | S_{[t]}) - \mathrm{H}(Y_t | W, S_{[t]}) \\ &= \mathrm{I}(W; Y_t | S_{[t]}) \,,\end{aligned} \tag{18}$$

where the above inequality follows from the fact that $\mathrm{H}(Y_t|Y_{[t-1]}, S_{[t]}) \leq \mathrm{H}(Y_t|S_{[t]})$, since removing the conditioning does not decrease the entropy, while $\mathrm{H}(Y_t|W, Y_{[t-1]}, S_{[t]}) = \mathrm{H}(Y_t|W, S_{[t]})$, as Y_t is conditionally independent from $Y_{[t-1]}$ given $(W, S_{[t]})$, due to the absence of output feedback. Since $(W, S_{[t]}) - (X_t, S_t) - Y_t$ forms a Markov chain, the data processing inequality implies that

$$\mathrm{I}(W; Y_t | S_{[t]}) \leq \mathrm{I}(X_t; Y_t | S_{[t]}) \,. \tag{19}$$

By combining (16), (17), (18) and (19), we then get

$$\begin{aligned}R_a + R_b &\leq \frac{1}{n} \log(|\mathcal{W}_a||\mathcal{W}_b|) \\ &\leq \frac{1}{1-\varepsilon} \frac{1}{n} \sum_{1 \leq t \leq n} \mathrm{I}(X_t; Y_t | S_{[t]}) + \frac{\mathrm{H}(\varepsilon)}{n(1-\varepsilon)} \\ &\leq \frac{1}{n} \sum_{1 \leq t \leq n} \mathrm{I}(X_t; Y_t | S_{[t]}) + \eta(\varepsilon) \,.\end{aligned} \tag{20}$$

Moreover, observe that

$$\begin{aligned}\mathrm{I}(X_t; Y_t | S_{[t]}) &= \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{t-1}} \mathbb{P}(S_{[t-1]} = \boldsymbol{\sigma}) \chi_{\boldsymbol{\sigma}} \\ &= n \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{t-1}} \alpha_{\boldsymbol{\sigma}} \chi_{\boldsymbol{\sigma}} \,,\end{aligned}$$

where $\chi_{\boldsymbol{\sigma}} := \mathrm{I}(X_t; Y_t | S_t, S_{[t-1]} = \boldsymbol{\sigma})$. Substituting into (20) yields (13).

Analogously, let us focus on encoder a: by Fano's inequality, we have that

$$\mathrm{H}(W_a | Y_{[n]}, S_{[n]}) \leq \mathrm{H}(\varepsilon) + \varepsilon \log(|\mathcal{W}_a|) \,. \tag{21}$$

For $t \geq 1$, define

$$\Delta_t^{(a)} := \mathrm{I}(W_a; Y_t, S_{t+1} | W_b, Y_{[t-1]}, S_{[t]}) \,,$$

and observe that

$$\begin{aligned}\sum_{1 \leq t \leq n} \Delta_t^{(a)} &= \mathrm{H}(W_a | S_1, W_b) - \mathrm{H}(W_a | W_b, S_{[n+1]}, Y_{[n]}) \\ &\geq \log |\mathcal{W}_a| - \mathrm{H}(W_a | S_{[n]}, Y_{[n]}) \,,\end{aligned} \tag{22}$$

where the last inequality follows from the independence between W_a, S_1, and W_b, and the fact that removing the conditioning does not decrease the entropy. Now, we have

$$\begin{aligned}\Delta_t^{(a)} &= \mathrm{I}(W_a; Y_t, S_{t+1} | W_b, Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{I}(W_a; Y_t | W_b, Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{H}(Y_t | W_b, Y_{[t-1]}, S_{[t]}) - \mathrm{H}(Y_t | W, Y_{[t-1]}, S_{[t]}) \\ &\leq \mathrm{H}(Y_t | W_b, S_{[t]}) - \mathrm{H}(Y_t | W, S_{[t]}) \\ &= \mathrm{I}(W_a; Y_t | W_b, S_{[t]}) \,,\end{aligned} \tag{23}$$

where the inequality above follows from the fact that $\mathrm{H}(Y_t|W_b, Y_{[t-1]}, S_{[t]}) \leq \mathrm{H}(Y_t|W_b, S_{[t]})$ since removing the conditioning does not decrease the entropy, and that $\mathrm{H}(Y_t|W, Y_{[t-1]}, S_{[t]}) = \mathrm{H}(Y_t|W, S_{[t]})$ due to absence of output feedback. Observe that, since, conditioned on W_b and $S_{[t]}$ (hence, on $X_t^{(b)}$), $W_a - X_t^{(a)} - Y_t$ forms a Markov chain, the data processing inequality implies that

$$\mathrm{I}(W_a; Y_t | W_b, S_{[t]}) \quad \leq \quad \mathrm{I}(X_t^{(a)}; Y_t | X_t^{(b)}, S_{[t]}) \,. \tag{24}$$

By combining (21), (22), (23), and (24), one gets

$$\begin{aligned}R_a &\leq \frac{1}{n} \log |\mathcal{W}_a| \\ &\leq \frac{1}{n(1-\varepsilon)} \sum_{1 \leq t \leq n} \mathrm{I}(X_t^{(a)}; Y_t | X_t^{(b)}, S_{[t]}) + \frac{\mathrm{H}(\varepsilon)}{n(1-\varepsilon)} \\ &\leq \frac{1}{n} \sum_{1 \leq t \leq n} \mathrm{I}(X_t^{(a)}; Y_t | X_t^{(b)}, S_{[t]}) + \eta(\varepsilon) \\ &= \sum_{\boldsymbol{\sigma} \in \mathcal{S}^{(n)}} \alpha_{\boldsymbol{\sigma}} \, \mathrm{I}(X_t^{(a)}; Y_t | X_t^{(b)}, S_t, S_{[t-1]} = \boldsymbol{\sigma}) + \eta(\varepsilon) \,,\end{aligned}$$

which proves (14).

In the same way, by reversing the roles of encoder a and b, one obtains (15). ∎

For $t \geq 1$, let us fix some string $\boldsymbol{\sigma} \in \mathcal{S}^{t-1}$, and focus our attention on the conditional mutual information terms $\mathrm{I}(X_t; Y_t | S_t, S_{[t-1]} = \boldsymbol{\sigma})$, $\mathrm{I}(X_t^{(a)}; Y_t | X_t^{(b)}, S_t, S_{[t-1]} = \boldsymbol{\sigma})$, and $\mathrm{I}(X_t^{(b)}; Y_t | X_t^{(a)}, S_t, S_{[t-1]} = \boldsymbol{\sigma})$, appearing in the rightmost sides of (13), (14), and (15), respectively. Clearly, the three of these quantities depend only on the joint conditional distribution of current channel state S_t, input X_t, and output Y_t, given the past state realization $S_{[t-1]} = \boldsymbol{\sigma}$. Hence, the key step now consists in showing that

$$\boldsymbol{\nu}_{\boldsymbol{\sigma}}(s, x, y) := \mathbb{P}(S_t = s, X_t = x, Y_t = y | S_{[t-1]} = \boldsymbol{\sigma}) \tag{25}$$

factorizes as in (4). This is proved in Lemma 6 below.

For $x_i \in \mathcal{X}_i$, $v_i \in \mathcal{V}_i$, and $\boldsymbol{\sigma} \in \mathcal{S}^{t-1}$, let us consider the set $\Upsilon_{\boldsymbol{\sigma}}^{(i)}(x_i, v_i) \subseteq \mathcal{W}_i$,

$$\Upsilon_{\boldsymbol{\sigma}}^{(i)}(x_i, v_i) := \left\{ w_i \colon \phi_t^{(i)}(w_i, q_i(\sigma_1), \ldots, q(\sigma_{t-1}), v_i) = x_i \right\}$$

and the probability distribution $\pi_{\boldsymbol{\sigma}}^{(i)}(\,\cdot\,|v_i) \in \mathcal{P}(\mathcal{X}_i)$,

$$\pi_{\boldsymbol{\sigma}}^{(i)}(x_i | v_i) := \sum_{w_i \in \Upsilon_{\boldsymbol{\sigma}}^{(i)}(x_i, v_i)} |\mathcal{W}_i|^{-1} \,.$$

Lemma 6: For every $1 \le t \le n$, $\boldsymbol{\sigma} \in \mathcal{S}^{t-1}$, $s \in \mathcal{S}$, $x_a \in \mathcal{X}_a$, and $x_b \in \mathcal{X}_b$,

$$\boldsymbol{\nu}_{\boldsymbol{\sigma}}(s, x, y) = P(s)\pi^{(a)}_{\boldsymbol{\sigma}}(x_a|q_a(s))\pi^{(b)}_{\boldsymbol{\sigma}}(x_b|q_b(s))P(y|s, x)\,. \tag{26}$$

Proof: First, observe that

$$\begin{aligned} \boldsymbol{\nu}_{\boldsymbol{\sigma}}(s, x, y) &= \mathbb{P}(S_t = s|S_{[t-1]} = \boldsymbol{\sigma})\nu_{\boldsymbol{\sigma}}(x|s)P(y|s, x) \\ &= P(s)\nu_{\boldsymbol{\sigma}}(x|s)P(y|s, x) \end{aligned} \tag{27}$$

where $\nu_{\boldsymbol{\sigma}}(x|s) := \mathbb{P}(X_t = x|S_{[t]} = (\boldsymbol{\sigma}, s))$. The former of the equalities in (27) follows from (1), while the latter is implied by the assumption that the channel state sequence is i.i.d..

Now, recall that, for $i \in \{a, b\}$, the current input satisfies $X^{(i)}_t = \phi^{(i)}_t(W_i, V^{(i)}_{[t]})$. For $w \in \mathcal{W}$, let $\xi_w := \mathbb{P}(X_t = x|S_{[t]} = (\boldsymbol{\sigma}, s), W = w)$. Then,

$$\begin{aligned} \nu_{\boldsymbol{\sigma}}(x|s) &= \sum\nolimits_w \xi_w \mathbb{P}(W = w|S_{[t]} = (\boldsymbol{\sigma}, s)) \\ &= \sum\nolimits_w |\mathcal{W}_a|^{-1}|\mathcal{W}_b|^{-1}\xi_w \\ &= \sum_{w_a \in \Upsilon^{(a)}_{\boldsymbol{\sigma}}(x_a, q_a(s))} |\mathcal{W}_a|^{-1} \sum_{w_b \in \Upsilon^{(b)}_{\boldsymbol{\sigma}}(x_b, q_b(s))} |\mathcal{W}_b|^{-1} \\ &= \pi^{(a)}_{\boldsymbol{\sigma}}(x_a|q_a(s))\pi^{(b)}_{\boldsymbol{\sigma}}(x_b|q_b(s))\,, \end{aligned} \tag{28}$$

the second inequality above following from the mutual independence of $S_{[t]}$, W_a, and W_b. The claim now follows from (27) and (28). ■

Let us now fix an achievable rate pair $R = (R_a, R_b)$. By choosing (R, n, ε)-codes for arbitrarily small $\varepsilon > 0$, the inequalities (13), (14), and (15), together with (10) and (12), imply that (R_a, R_b) can be approximated by convex combinations of rate pairs (indexed by $\boldsymbol{\sigma} \in \mathcal{S}^{(n)}$) satisfying (3) for joint state-input-output distributions as in (25). Hence, any achievable rate pair R belongs to $\overline{\mathrm{co}}(\cup_\pi \mathcal{R}(\pi))$.

Remark 1: For the validity of the arguments above, two critical steps were (27), where the hypothesis of i.i.d. channel state sequence has been used, and (28), which only relies on the mutual independence of W and $S_{[t]}$, this being a consequence of the assumption of absence of inter-symbol interference. In particular, the key point in (27) is the fact that the past state realization $\boldsymbol{\sigma}$ appears in $\nu_{\boldsymbol{\sigma}}$ only and not in $P(S_t)$. ⋄

Remark 2: For the validity of the arguments above, it is critical that the receiver observes the channel state. More in general, it would suffice that the state information available at the decoder contains the one available at the two transmitters. In this way, the decoder does not need to estimate the coding policies used in a decentralized time-sharing. ⋄

V. Extensions and Concluding Remarks

The present paper has dealt with the problem of reliable transmission over finite-state multiple access channels with asymmetric, partial channel state information at the encoders. A single-letter characterization of the capacity region has been provided in the special case when the channel state is a sequence of independent and identically distributed random variables.

It is worth commenting to which extent the results above can be generalized to channels with memory. Let us consider the case when the channel state sequence $\{S_t : t = 1, 2, \ldots\}$ forms a Markov chain with transition probabilities $\mathbb{P}(S_{t+1} = s_+|S_t = s) = P(s_+|s)$ which are stationary and satisfy the strongly mixing condition $P(s_+|s) > 0$ for all $s, s_+ \in \mathcal{S}$. Further, assume that there is no inter-symbol interference, i.e. $\{S_t : t = 1, 2, \ldots\}$ is independent from the message W, and that the state process is observed through quantized observations $V^{(i)}_t = q_i(S_t)$, as discussed earlier.

In general, for a multi-person optimization problem, whenever a dynamic programming recursion with a fixed complexity per time stage is possible via the construction of a Markov Chain with a fixed state space (see [21] for a review of information structures in decentralized control), the information structure is said to have a quasi-classical pattern; thus, under such a structure, the optimization problem is computationally feasible and the problem is said to be *tractable*. In a team decision theoretic approach, one may assume that there is a centralized decision maker which designs an optimal team design statically, before the realization of the random variables. This approach is based on Witsenhausen's equivalent model for discrete stochastic control [22].

In the case of finite-state multiple access channels with independent and identically distributed state sequence, by first showing that the past information is irrelevant, we observed that we could limit the memory space on which the optimization is performed. This is because, as observed in Remark 1, in the rightmost side of (27) the past state realization $\boldsymbol{\sigma}$ affects only the control $\nu_{\boldsymbol{\sigma}}(x|s)$, but not the current state distribution $P(S_t)$. In contrast, when the state sequence is a Markov chain, the past state realization $\boldsymbol{\sigma}$ does affect both the control $\nu_{\boldsymbol{\sigma}}(x|s)$ as well as the current state distribution $P(S_t|S_{[t-1]} = \boldsymbol{\sigma})$. It is exactly such a joint dependence which prevents the proof presented here to be generalized to the Markov case.

In case there is only one transmitter, the conditional probability distribution of the state given the observation history, $\Pi_t(\,\cdot\,) := \mathbb{P}(S_t = \,\cdot\,|V_{[t]}) \in \mathcal{P}(\mathcal{S})$, can be shown to be a sufficient statistic, i.e. the optimal coding policy can be shown to depend on it only. As a consequence, the optimization problem is tractable. Such a setting was studied in [20], where the following single-letter characterization was obtained for the capacity of finite-state single-user channels with quantized state observation at the transmitter and full state observation at the receiver:

$$C := \int_{\mathcal{P}(\mathcal{S})} \mathrm{d}\tilde{P}(\pi) \sup_{P(X|\pi) \in \mathcal{P}(\mathcal{X})} \left\{ \sum_{s} I(X;Y|s,\pi) \tilde{P}(s|\pi) \right\}$$

where $\tilde{P}(s,\pi) := \tilde{P}(s|\pi)\tilde{P}(\pi)$ denotes the asymptotic joint distribution of the state S_t and its estimate Π_t, existence and uniqueness of which are ensured by the strong mixing condition.

For finite-state multiple access channels with memory, a similar approach can successfully be undertaken only if the state observation is symmetric, namely if $q_a = q_b$. Indeed, in this case, the conditional state estimation $\Pi_t(\,\cdot\,) = \mathbb{P}(S_t = \cdot\,|V_{[t]}^{(a)}) = \mathbb{P}(S_t = \cdot\,|V_{[t]}^{(b)})$ can be shown to be a sufficient statistic, and a single-letter characterization of the capacity region can be proved.

However, for the general case when the channel state sequence has memory and the state observation is asymmetric (i.e. $q_a \neq q_b$), the construction of a Markov chain is not straightforward. The conditional measure on the channel state is no longer a sufficient statistic: In particular, if one adopts a team decision based approach, where there is a fictitious centralized decision maker, this decision maker should make decisions for all the possible memory realizations, that is the policy is to map the variables $(W, V_{[t]}^{(a)}, V_{[t]}^{(b)})$ to $(X_t^{(a)}, X_t^{(b)})$ decentrally, and the memory cannot be truncated, as every additional bit is essential in the construction of an equivalent Markov chain to which the Markov Decision Dynamic Program can be applied; both for the prediction on the channel state as well as the belief of the coders on each other's memory.

Let us also elaborate a discussion in view of *common knowledge* of Aumann [2]: An information between two decision makers is common knowledge if it is measurable on the sigma-fields generated by both of the local information variables at the decision makers. It is rare in practical applications that all the local information is common knowledge, and requiring such a limitation is too restrictive as elaborated in [21]: For example, if we look for person-by-person optimal policies, a policy of one of the encoders (say Encoder a) which uses the past will force the other encoder (Encoder b) to also use the past to second-guess the action of Encoder a. Encoder a, in turn, will have to guess the guessing of Encoder b with regard to its action, and this will lead to a regress of belief on beliefs ad infinitum. The approach in our paper showed that we can avoid such an approach when the channel state sequence is memoryless. However, when the channel state has memory, the past information provides useful information which is important for estimating the future channel states. As such, we cannot avoid the use of the information on the past channel state realizations. If one is to construct an equivalent state based on which to generate coding policies, the equivalent state needs to keep growing with time: The discussion in [7] provides such a block-level characterization and it seems we cannot go beyond this due to the non-tractability of the optimization problem. We note that if the encoders can exchange their past observations with a fixed delay, if they can exchange their observations periodically, or if they can exchange their beliefs at every time stage, then the optimization problem will be tractable.

One question of important practical interest is the following: If the channel transitions form a Markov chain, which is mixing fast, is it sufficient to use a finite memory construction for practical purposes? This is currently being investigated.

References

[1] E. Altman, *Constrained Markov Decision Processes*, Chapman & Hall/CRC, Boca Raton, FL, 1999.

[2] R. J. Aumann, "Agreeing to Disagree", *Annals of Statistics*, vol. 4, pp. 1236–1239, 1976.

[3] G. Caire and S. Shamai, "On the capacity of some channels with channel state information", *IEEE Trans. Inf. Theory*, vol. 45, pp. 2007–2019, 1999.

[4] Y. Cemal, and Y. Steinberg, "Multiple access channel with partial state information at the encoders", *IEEE Trans. Inf. Theory*, vol. 51, pp. 3992–4003, 2005.

[5] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback", *IEEE Trans. Inf. Theory*, vol. 51, pp. 780–789, 2005.

[6] I. Csiszar, and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akademiai Kiado, 1981.

[7] A. Das, and P. Narayan, "Capacities of time-varying multiple-access channels with side information", *IEEE Trans. Inf. Theory*, vol. 48, pp. 4–25, 2002.

[8] S.I. Gelfand, and M.S. Pinsker, "Coding for channel with random parameters," *Prob. Control Information*, vol. 9, pp. 19–31, 1980.

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

[10] O. Hernandez-Lerma, and J.B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.

[11] S. Tatikonda, S. Yang, and A. Kavcic, "Feedback capacity of finite-state machine channels", *IEEE Trans. Inf. Theory*, vol. 51, pp. 799–810, 2005.

[12] M. Medard, and R. Srikant, "Capacity of nearly-decomposable Markov chains under asymmetric receiver-sender side information", *IEEE Trans. Inf. Theory*, vol. 52, pp. 3052–3062, 2006.

[13] S.A. Jafar, "Capacity with causal and non-causal side information – A Unified View", *IEEE Trans. Inf. Theory*, vol. 52, pp. 5468–5475, 2006.

[14] H. Permuter, T. Weissman, and J. Chen, "New bounds for the capacity region of the finite-state multiple access channel", in *Proc. IEEE Int. Symp. Information Theory*, Toronto, Canada, June 2008, pp. 394–398.

[15] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder", *Proc. Inst. Elec. Eng. Pt.*, vol. 139, pp. 113–117, April 1992.

[16] C. Shannon, "Channels with side information at the transmitter," *IBM J. Research Develop.*, vol. 2, pp. 289–293, 1958.

[17] S. Tatikonda, and S.K. Mitter, "The Capacity of channels with feedback", *IEEE Trans. Inf. Theory*, vol. 55, pp. 323-349, 2009.

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

[19] H. Viswanathan, “Capacity of Markov Channels with Receiver CSI and Delayed Feedback”, *IEEE Trans. Inf. Theory*, vol. 45, pp. 761–771, March 1999.

[20] S. Yüksel, and S. Tatikonda, “Capacity of Markov Channels with Partial State Feedback”, in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Nice, France, June 2007, pp. 1861–1865.

[21] S. Yüksel, “Stochastic nestedness and the belief sharing information pattern”, *IEEE Trans. Autom. Control*, vol. 55, pp. 2773–2786, 2009.

[22] H.S. Witsenhausen, “Equivalent stochastic control problems”, *Math. Control Signals Systems*, vol. 1, pp. 3–11, 1988.