

On the capacity of finite state multiple access channels with asymmetric partial state feedback¹

Giacomo Como² and Serdar Yüksel³

June 19, 2009

Abstract

A single letter characterization is provided for the capacity region of finite-state multiple access channels, when the channel state is an independent and identically distributed process, the transmitters have access to partial (quantized) state information, and complete channel state information is available at the receiver. The partial state information is asymmetric at the encoders. The problem is practically relevant, and provides a tractable optimization problem. The case where the channel state process is Markovian is also discussed.

1 Introduction and Literature Review

Wireless communication channels and the Internet are examples of channels where the channel characteristics are time-varying. Channel fading models for wireless communications include fast fading and slow fading; in fast fading the channel state is assumed to be changing for each use of the channel, whereas in slow fading, the channel is assumed to be constant for each finite block length. In fading channels, the channel fade might not be transmitted to the transmitter over a perfect channel, but via reducing the data rate, the error in feedback transmission can be improved.

The present paper studies finite state multiple access channels (MACs) with asymmetrically quantized state information at the transmitters, and perfect state information at the receiver. A single letter characterization of the capacity region is proved for the case of independent and identically distributed channel state sequence. Generalizations to the case where the channel state dynamics is Markovian are discussed as well. Our approach is inspired by analogies with team decision theory [18, 17].

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 side information being available at neither, either or both the transmitter and receiver have been studied in [12] and [5]. [1] studied the capacity of channels with memory and complete noiseless output feedback and introduced a stochastic control formulation for the computation via the properties of the directed mutual information. Reference [6]

¹This paper appeared in part 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.

²G. 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.

³S. 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. Research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant.

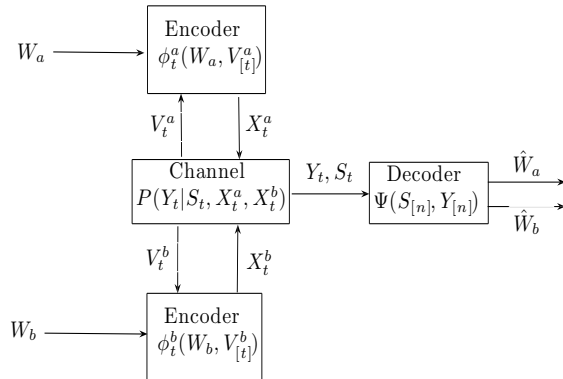


Figure 1: Finite state multiple access channel with asymmetric partial state feedback.

considered fading channels with perfect channel state information at the transmitter, and showed that with instantaneous and perfect state feedback, the transmitter can adjust the data rates for each channel state to maximize the average transmission rate. Viswanathan [15] relaxed this assumption of perfect instantaneous feedback, and studied the capacity of Markovian channels with delayed feedback. Reference [16] studied the capacity of Markov channels with perfect causal state feedback. Capacity of Markovian, finite state channels with quantized state feedback available at the transmitter was studied in [2].

A related work [9] has studied MAC channels where the encoders have degraded information on the channel state, which is coded to the encoders. Authors in [9] also considered outer and inner bounds on the capacity region when the information at the encoders is asymmetric, that is not necessarily degraded. In contrast, the present paper considers and obtains a precise result for a setting where the encoders have asymmetric, partial state information, which is obtained through fixed quantizers acting componentwise, rather than encoded in blocks (that is, the side information is causal as opposed to non-causal). Another recent related work is [20] which provided an infinite-dimensional characterization for the capacity region for Multiple Access Channels with feedback.

The rest of the paper is organized as follows. In Sect. 2 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. Sect. 3 contains the proof of achievability of the capacity region, while Sect. 4 presents a proof of the converse coding theorem. Finally, in Sect. 5 we discuss generalizations to the memory case and present some final remarks.

2 Capacity of i.i.d. Finite-State MAC Channel with Asymmetric Partial State Feedback

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 the usual 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 $H(\cdot)$, and $I(\cdot; \cdot)$ (respectively $H(\cdot | \cdot)$, and $I(\cdot; \cdot | \cdot)$) 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 $H(x)$ 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, multiple access channel 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 $S = (S_t)$ 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 $q_i : \mathcal{S} \rightarrow \mathcal{V}_i$ 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 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

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

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; 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

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

and a decoding function

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

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

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

Upon the description of the channel and transmission systems, we now proceed with the characterization of the capacity region.

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

Definition 2 A rate pair $R \in \mathbb{R}_+^2$ 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 static team policies and their associated rate regions.

Definition 3 A static team policy is a family

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

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

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

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

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

We can now state the main result of the paper.

Theorem 4 The achievable rate region is given by

$$\overline{\text{co}} \left(\bigcup_{\pi} \mathcal{R}(\pi) \right)$$

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

In Sect.3 we shall prove the direct part of Theorem 4, namely that every rate pair $R \in \overline{\text{co}}(\bigcup_{\pi} \mathcal{R}(\pi))$ is achievable. In Sect. 4 we shall prove the converse part, i.e. that no rate pair $R \in \mathbb{R}_+^2 \setminus \overline{\text{co}}(\bigcup_{\pi} \mathcal{R}(\pi))$ is achievable.

3 Achievability of the capacity region

In this section, we shall show that any rate pair $R = (R_a, R_b)$ belonging to the region $\mathcal{R}(\pi)$, for some static policy π , is achievable. Achievability of any rate pair R in $\text{co}(\bigcup_{\pi} \mathcal{R}(\pi))$ will then follow by a standard time-sharing argument (see e.g. [19, Lemma 2.2, p.272]).

In order to prove achievability on the original finite state MAC, 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 $\mathcal{U}_i := \{u_i : \mathcal{V}_i \rightarrow \mathcal{X}_i\}$, 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

$$f^{(i)} : \mathcal{W}_i \rightarrow \mathcal{U}_i^n, \quad i \in \{a, b\},$$

and a decoder

$$g : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{W}_a \times \mathcal{W}_b.$$

To any such coding scheme it is natural to associate a coding scheme for the original finite state MAC, by defining the encoders

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

for $i \in \{a, b\}$, and letting the decoder $\psi : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{W}_a \times \mathcal{W}_b$ coincide with g . It is not hard to verify the following:

Lemma 3.1 *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

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

where $U = (U_a, U_b)$ and Z are r.v.s whose joint distribution factorizes as

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

Consider a sequence of random codes

$$\left\{ f_a^{(n)} : \mathcal{W}_a^{(n)} \rightarrow \mathcal{U}_a^n, f_b^{(n)} : \mathcal{W}_b^{(n)} \rightarrow \mathcal{U}_b^n \right\}_n$$

where

$$|\mathcal{W}_a^{(n)}| = \lceil \exp(R_a n) \rceil, \quad |\mathcal{W}_b^{(n)}| = \lceil \exp(R_b n) \rceil$$

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 $\mu_i \otimes \dots \otimes \mu_i$, for each $i \in \{a, b\}$ and $w_i \in \mathcal{W}_i$. Then, it follows from the direct coding theorem for memoryless MACs [19, Th.3.2, p. 272] that the average error probability of such a code ensemble converges to zero in the limit of n going to infinity.

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

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

for some static team policy π 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 r.v.s $S, X_a := U_a(q_a(S))$,

$X_b := U_b(q_b(S))$, and Y , whose joint probability distribution satisfies (4). Moreover, it can be readily verified that

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

Hence, if a rate pair $R = (R_a, R_b)$ belongs to the rate region $\mathcal{R}(\pi)$ associated to some static 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 . Thanks to Lemma 3.1, 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 $\overline{\text{co}}(\cup_\pi \mathcal{R}(\pi))$ then follows from a standard time-sharing principle.

4 Converse to the coding theorem

In this section, we shall prove that no rate outside $\overline{\text{co}}(\cup_\pi \mathcal{R}(\pi))$ 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

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

and observe that

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

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

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

Clearly,

$$\alpha_\sigma \geq 0, \quad \sum_{\sigma \in \mathcal{S}^{(n)}} \alpha_\sigma = \frac{1}{n} \sum_{1 \leq t \leq n} \sum_{\sigma \in \mathcal{S}^{t-1}} \mathbb{P}(S_{[t-1]} = \sigma) = 1. \tag{12}$$

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,*

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

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

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

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

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

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

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

and observe that

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

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

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

where the above inequality follows from the fact that $H(Y_t | Y_{[t-1]}, S_{[t]}) \leq H(Y_t | S_{[t]})$, since removing the conditioning does not decrease the entropy, while $H(Y_t | W, Y_{[t-1]}, S_{[t]}) = H(Y_t | W, S_{[t]})$, 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

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

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

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

Moreover, observe that

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

Substituting into (20) yields (13).

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

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

For $t \geq 1$, define

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

and observe that

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

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

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

where the inequality above follows from the fact that $\mathbb{H}(Y_t | W_b, Y_{[t-1]}, S_{[t]}) \leq \mathbb{H}(Y_t | W_b, S_{[t]})$ since removing the conditioning does not decrease the entropy, and that $\mathbb{H}(Y_t | W, Y_{[t-1]}, S_{[t]}) = \mathbb{H}(Y_t | W, S_{[t]})$ 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

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

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

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

which proves (14).

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

For $t \geq 1$, let us fix some string $\sigma \in \mathcal{S}^{t-1}$, and focus our attention on the conditional mutual information terms $I(X_t; Y_t | S_t, S_{[t-1]} = \sigma)$, $I(X_t^{(a)}; Y_t | X_t^{(b)}, S_t, S_{[t-1]} = \sigma)$, and $I(X_t^{(b)}; Y_t | X_t^{(a)}, S_t, S_{[t-1]} = \sigma)$, 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 $S_{[t-1]} = \sigma$. Hence, the key step now consists in showing that

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

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

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

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

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

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

$$\nu_{\sigma}(s, x, y) = P(s)\pi_{\sigma}^{(a)}(x_a | q_a(s))\pi_{\sigma}^{(b)}(x_b | q_b(s))P(y | s, x). \quad (26)$$

Proof First, observe that

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

where

$$\nu_{\sigma}(x | s) := \mathbb{P}(X_t = x | S_{[t]} = (\sigma, s)),$$

the former above equality 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 $X_t^{(i)} = \phi_t^{(i)}(W_i, V_t^{(i)})$. Therefore, we have

$$\begin{aligned} \nu_{\sigma}(x | s) &= \sum_w \mathbb{P}(X_t = x | S_{[t]} = (\sigma, s), W = w) \mathbb{P}(W = w | S_{[t]} = (\sigma, s)) \\ &= \sum_w \frac{1}{|\mathcal{W}_a| |\mathcal{W}_b|} \mathbb{P}(X_t = x | S_{[t]} = (\sigma, s), W = w) \\ &= \sum_{w_a \in \Upsilon_{\sigma}^{(a)}(x_a, q_a(s))} \frac{1}{|\mathcal{W}_a|} \sum_{w_b \in \Upsilon_{\sigma}^{(b)}(x_b, q_b(s))} \frac{1}{|\mathcal{W}_b|} \\ &= \pi_{\sigma}^{(a)}(x_a | q_a(s)) \pi_{\sigma}^{(b)}(x_b | q_b(s)), \end{aligned} \quad (28)$$

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 $\sigma \in \mathcal{S}^{(n)}$) satisfying (3) for joint state-input-output distributions as in (25). Hence, any achievable rate pair R belongs to $\overline{\text{co}}(\cup_{\pi} \mathcal{R}(\pi))$.

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 σ appears in ν_{σ} 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. In this way, the decoder does not need to estimate the coding policies used in a decentralized time-sharing. ◇

5 Extensions and concluding Remarks

The present paper has dealt with the problem of reliable transmission over multiple access finite state channels with asymmetric, imperfect channel state information. 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 $\{S_t : t \geq 1\}$ forms a Markov chain with stationary transition probabilities $\mathbb{P}(S_{t+1} = s_+ | S_t = s) = P(s_+ | s)$ satisfying 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. $\{S_t : t \geq 1\}$ is independent from the message W , and that the state process is observed through quantized observations $V_t^{(i)} = q_i(S_t)$ 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 [17] 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 assumes that there is a centralized fictitious decision maker which designs an optimal design statically, before random variables take place. This approach is based on Witsenhausen's equivalent model for discrete stochastic control [18].

For case of finite state MACs with i.i.d. 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 σ affects only the control $\nu_\sigma(x|s)$, but not the current state distribution $P(S_t)$. In contrast, when the state sequence is a Markov chain, the past state realization σ does affect both the control $\nu_\sigma(x|s)$ as well as the current state distribution $P(S_t | S_{[t-1]} = \sigma)$. 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 [2], 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:

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

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 belief of the coders on each other's memory. 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 [17]. One question of 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] S. Tatikonda and S. Mitter, "The Capacity of channels with feedback", *IEEE Trans. Inf. Theory*, vol. 55, pp. 323-349, 2009.
- [2] S. Yüksel and S. Tatikonda, "Capacity of Markov Channels with Partial State Feedback", in *Proc. of the IEEE International Symposium on Information Theory*, June 2007, Nice, France.
- [3] E. Altman, *Constrained Markov Decision Processes*, Chapman & Hall/CRC, Boca Raton, FL, 1999.
- [4] G. Caire and S. Shamai "On the capacity of some channels with channel state information", *IEEE Transactions on Information Theory*, vol. 45, pp. 2007-2019, 1999.
- [5] S.I. Gelfand and M.S. Pinsker, "Coding for Channel with Random Parameters," *Prob. of Control and Information*, vol. 9, pp. 19-31, 1980.
- [6] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information", *IEEE Transactions on Information Theory*, vol. 43, pp. 1986-1992, 1997.
- [7] O. Hernandez-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.
- [8] S. Tatikonda, S. Yang and A. Kavcic, "Feedback capacity of finite-state machine channels", *IEEE Transactions on Information Theory*, vol. 51, pp. 799 - 810, 2005.
- [9] Y. Cemal and Y. Steinberg, "Multiple access channel with partial state information at the encoders", *IEEE Transactions on Information Theory*, vol. 51, pp. 3992-4003, 2005.
- [10] S.A. Jafar "Capacity with causal and non-causal side information - A Unified View", *IEEE Transactions on Information Theory*, vol. 52, pp. 5468-5475, 2006.
- [11] 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.
- [12] C. Shannon, "Channels with side information at the transmitter," *IBM Journal Research and Development*, vol. 2, pp. 289-293, 1958.
- [13] M. Medard and R. Srikant, "Capacity of nearly-decomposable Markov chains under asymmetric receiver-sender side information", *IEEE Transactions on Information Theory*, vol. 52, pp. 3052-3062, 2006.

- [14] S. Verdú and T. S. Han, “A general formula for channel capacity”, *IEEE Transactions on Information Theory*, vol. 40, pp. 1147-1157, 1994.
- [15] H. Viswanathan, “Capacity of Markov Channels with Receiver CSI and Delayed Feedback”, *IEEE Transactions on Information Theory*, vol. 45, pp. 761-771, March 1999.
- [16] J. Chen and T. Berger, “The capacity of finite-state Markov channels with feedback”, *IEEE Transactions on Information Theory*, vol. 51, pp. 780-789, 2005.
- [17] S. Yüksel, “Stochastic nestedness and the belief sharing information pattern”, *IEEE Transactions on Automatic Control*, to appear, 2009.
- [18] H. S. Witsenhausen, “Equivalent stochastic control problems”, *Mathematics of Control, Signals, and Systems*, vol. 1, pp. 3-11, Springer-Verlag, 1988.
- [19] I. Csiszar and J. Korner, “Information Theory: Coding Theorems for Discrete Memoryless Systems”, Akademiai Kiado, 1981.
- [20] H. Permuter, T. Weissman and J. Chen, “New bounds for the capacity region of the finite-state multiple access channel”, in *Proc. of the IEEE International Symposium on Information Theory*, June 2008, Toronto, Canada.