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

On the Capacity of Finite State Multiple Access Channels with Asymmetric Partial State Feedback

Giacomo Como[1] and Serdar Yüksel [2]

***Abstract*— We provide a single letter characterization of the capacity region for independent, identically distributed, finite-state channels, with partial (quantized) state information, when the 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. We also consider the case where the channel has memory.**

I. Introduction and Literature Review

Wireless communication channels and Internet are examples of channels where the channel characteristics are variable. 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.

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] 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].

[1]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

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

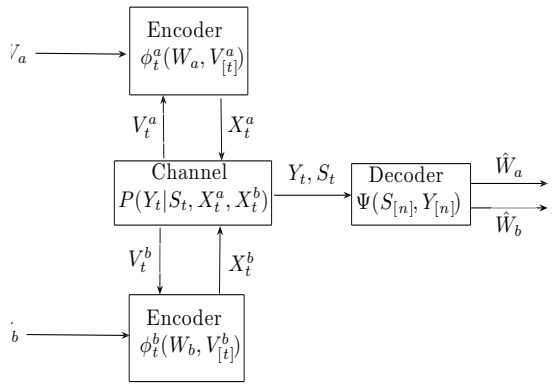

Fig. 1: Finite state multiple access channel with asymmetric partial state feedback.

A related work [9] has studied MAC channels where the encoders have degraded information on the channel state, which is coded to the encoders. In this paper, we consider a setting where the encoders have asymmetric partial state information, where the partial information is obtained through a quantizer.

Another recent related work is [20] which provided an infinite-dimensional characterization for the capacity region for Multiple Access Channels with feedback.

We use tools from team decision theory to develop our result [17], [18]. We also discuss the case where the channel state has Markovian dynamics.

II. 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 $v_{[i]} = (v_1, \ldots, v_i)$ denotes 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 $\mathrm{H}(\,\cdot\,)$, and $I(\,\cdot\,;\,\cdot\,)$ (respectively $\mathrm{H}(\,\cdot\,|\,\cdot\,)$, and $I(\,\cdot\,;\,\cdot\,|\,\cdot\,)$) for the (conditional) binary entropy and mutual information of r.v.s. For $0 \leq x \leq 1$, $\mathrm{H}(x)$ will also denote the binary 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

$$\mathcal{A}^{(n)} := \bigcup_{0 \leq s < t} \mathcal{A}^s$$

the set of $\mathcal{A}$-strings of length smaller than n.

We shall consider a finite state, multiple access channel with two transmitters, indexed by $i = a, b$, 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 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 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 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 $\mathcal{V} := \mathcal{V}_a \times \mathcal{V}_b$. The channel output at time t, Y_t, takes values in a finite set $\mathcal{Y}$; its conditional distribution satisfies

$$P(Y_t | W, X_{[N]}, S_{[N]}) = P(Y_t | X_t, S_t)\,. \tag{1}$$

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 $R = (R_a, R_b) \in \mathbb{R}_+^2$, 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

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

such that, for $i = a, b$, $1 \leq t \leq n$:

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

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

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

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

of probability distributions on the two channel input sets conditioned on the quantized observation of each transmitter. For every static team policy π, $\mathcal{R}(\pi)$ will denote the region of all rate pairs $R = (R_a, R_b) \in \mathbb{R}_+^2$ satisfying

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

where S, $X = (X_a, X_b)$, and Y, are r.v.s whose joint distribution factorizes as

$$\begin{aligned} &P(S, X_a, X_b, Y) \\ &= P(S)\pi_a(X_a | q_a(S))\pi_b(X_b | q_b(S))P(Y | S, X_a, X_b)\,. \end{aligned} \tag{4}$$

We can now state the main result of the paper.

Theorem 4: The achievable rate region is given by

$$\overline{\mathrm{co}}\,(\cup_\mu \mathcal{R}(\mu))$$

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

In the sequel, we shall prove this theorem.

III. Converse to the Channel Theorem

For $R \in \mathbb{R}_+^2$, and $0 < \varepsilon \leq 1/2$, let us consider a (R, n, ε)-code. Fano's inequality implies that

$$H(W | Y_{[n]}; S_{[n]}) \quad \leq \quad \mathrm{H}(\varepsilon) + \varepsilon \log(|\mathcal{W}_a||\mathcal{W}_b|)\,. \tag{5}$$

For $t \geq 1$, define

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

and observe that

$$\begin{aligned} \sum_{1 \leq t \leq N} \Delta_t &= \mathrm{H}(W | S_1) - \mathrm{H}(W | S_{[n+1]}, Y_{[n]}) \\ &= \log(|\mathcal{W}_a||\mathcal{W}_b|) - \mathrm{H}(W | S_{[n]}, Y_{[n]})\,, \end{aligned} \tag{6}$$

since S_1 is independent of W, and 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

$$\begin{aligned}\Delta_t &= I(W; Y_t, S_{t+1}|Y_{[t-1]}, S_{[t]}) \\ &= I(W; Y_t|Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{H}(Y_t|Y_{[t-1]}, S_{[t]}) - \mathrm{H}(Y_t|W, Y_{[t-1]}, S_{[t]}) \\ &\leq \mathrm{H}(Y_t|S_{[t]}) - \mathrm{H}(Y_t|W, S_{[t]}) \\ &= I(W; Y_t|S_{[t]}) \,,\end{aligned} \tag{7}$$

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 $(W, S_{[t]}) - X_t - Y_t$ forms a Markov chain, the data processing inequality implies that

$$I(W; Y_t|S_{[t]}) \leq I(X_t; Y_t|S_{[t]}) \,. \tag{8}$$

By combining (5), (6), (7) and (8), we then get

$$\begin{aligned}R_a + R_b &\leq \frac{1}{n}\log(|\mathcal{W}_a||\mathcal{W}_b|) \\ &\leq \frac{1}{1-\varepsilon}\frac{1}{n}\sum_{1\leq t\leq n} I(X_t; Y_t|S_{[t]}) + \frac{\mathrm{H}(\varepsilon)}{n(1-\varepsilon)} \\ &\leq \frac{1}{n}\sum_{1\leq t\leq n} I(X_t; Y_t|S_{[t]}) + \eta(\varepsilon) \\ &\leq \sum_{\boldsymbol{s}\in\mathcal{S}^{(n)}} \alpha_{\boldsymbol{s}} I(X_t; Y_t|S_t, S_{[t-1]} = \boldsymbol{s}) + \eta(\varepsilon) \,,\end{aligned} \tag{9}$$

where

$$\eta(\varepsilon) := \frac{\varepsilon}{1-\varepsilon}\log|\mathcal{Y}| + \frac{\mathrm{H}(\varepsilon)}{1-\varepsilon} \,, \tag{10}$$

is such that

$$\lim_{\varepsilon\to 0}\eta(\varepsilon) = 0 \,, \tag{11}$$

and

$$\alpha_{\boldsymbol{s}} := \frac{1}{n}\mathbb{P}(S_{[t-1]} = \boldsymbol{s}) \,, \qquad \boldsymbol{s}\in\mathcal{S}^{(n)} \,, \tag{12}$$

are such that

$$\sum_{\boldsymbol{s}\in\mathcal{S}^{(n)}} \alpha_{\boldsymbol{s}} = 1 \,. \tag{13}$$

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

$$\mathrm{H}(W_a|Y_{[N]}; S_{[N]}) \leq \mathrm{H}(\varepsilon) + \varepsilon\log(|\mathcal{W}_a|) \,. \tag{14}$$

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

$$\begin{aligned}\sum_{1\leq t\leq n}\Delta_t^a &= \mathrm{H}(W_1|S_1, W_2) - \mathrm{H}(W_a|W_b, S_{[n+1]}, Y_{[n]}) \\ &\geq \log|\mathcal{W}_a| - \mathrm{H}(W_a|S_{[n]}, Y_{[n]}) \,,\end{aligned} \tag{15}$$

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

$$\begin{aligned}\Delta_t^a &= I(W_a; Y_t, S_{t+1}|W_b, Y_{[t-1]}, S_{[t]}) \\ &= I(W_a; Y_t|W_b, Y_{[t-1]}, S_{[t]}) \\ &= \mathrm{H}(Y_t|W_b, Y_{[t-1]}, S_{[t]}) - \mathrm{H}(Y_t|W, Y_{[t-1]}, S_{[t]}) \\ &\leq \mathrm{H}(Y_t|W_b, S_{[t]}) - \mathrm{H}(Y_t|W, S_{[t]}) \\ &= I(W_a; Y_t|W_b, S_{[t]}) \,,\end{aligned} \tag{16}$$

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

$$I(W_a; Y_t|W_b, S_{[t]}) \leq I(X_t^a; Y_t|X_t^b, S_{[t]}) \,. \tag{17}$$

By combining (14), (15), (16), and (17), one gets

$$\begin{aligned}R_a &\leq \frac{1}{n}\log|\mathcal{W}_a| \\ &\leq \frac{1}{1-\varepsilon}\frac{1}{n}\sum_t I(X_t^a; Y_t|X_t^b, S_{[t]}) + \frac{1}{n}\frac{\mathrm{H}(\varepsilon)}{1-\varepsilon} \\ &\leq \frac{1}{n}\sum_t I(X_t^a; Y_t|X_t^b, S_{[t]}) + \eta(\varepsilon) \\ &= \sum_{\boldsymbol{s}\in\mathcal{S}^{(n)}} \alpha_{\boldsymbol{s}} I(X_t^a; Y_t|X_t^b, S_t, S_{[t-1]} = \boldsymbol{s}) + \eta(\varepsilon) \,.\end{aligned} \tag{18}$$

In the same way, by reversing the roles of encoder a and b, one obtains

$$R_b \leq \frac{1}{n}\sum_{\boldsymbol{s}\in\mathcal{S}^{(n)}} I(X_t^a; Y_t|X_t^b, S_t, S_{[t-1]} = \boldsymbol{s}) + \eta(\varepsilon) \,. \tag{19}$$

Observe that, for all $s_{[t]} \in \mathcal{S}^t$, $x = (x_a, x_b) \in \mathcal{X}$, and $y \in \mathcal{Y}$,

$$\begin{aligned}&\mathbb{P}(S_t = s_t, X_t = x, Y_t = y|S_{[t-1]} = s_{[t-1]}) \\ =\ &\mathbb{P}(S_t = s_t|S_{[t-1]} = s_{t-1}) \\ &\quad \mathbb{P}(X_t = x|S_{[t]} = \boldsymbol{s}_{[t]})P(y|s_t, x) \\ =\ &P(s_t)\mathbb{P}(X_t = x|S_{[t]} = \boldsymbol{s}_{[t]})P(y|s_t, x) \,,\end{aligned} \tag{20}$$

the former above equality following from (1), the latter being implied by the assumption that the channel state sequence is i.i.d..

Now, recall that $X_t^i = \phi_t^i(W_i, V_{[t]}^i)$, for $i = a, b$. For $x_i \in \mathcal{X}_i$, $v_i \in \mathcal{V}_i$, and $\boldsymbol{s} \in \mathcal{S}^{t-1}$, let us consider the set $\Upsilon_{\boldsymbol{s}}^i(x_i, v_i) \subseteq \mathcal{W}_i$,

$$\Upsilon_{\boldsymbol{s}}^i(x_i, v_i) := \left\{w_i : \phi_t^i(w_i, q_i(s_1), \ldots, q(s_{t-1}), v_i) = x_i\right\} \,,$$

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

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

Then, we have

$$\begin{aligned}
&\mathbb{P}(X_t = x|S_{[t]} = s_{[t]}) \\
&= \sum_w \mathbb{P}(X_t = x|S_{[t]} = s_{[t]}, W = w)\mathbb{P}(W = w|s_{[t]}) \\
&= \sum_w \mathbb{P}(X_t = x|S_{[t]} = s_{[t]}, W = w)|\mathcal{W}_a|^{-1}|\mathcal{W}_b|^{-1} \\
&= \sum_{w_a \in \Upsilon^a_{\boldsymbol{s}}(x_a, q_a(s_t))} |\mathcal{W}_a|^{-1} \sum_{w_b \in \Upsilon^b_{\boldsymbol{s}}(x_b, q_b(s_t))} |\mathcal{W}_b|^{-1} \\
&= \pi^a_{\boldsymbol{s}}(x_a|q_a(s_t))\pi^b_{\boldsymbol{s}}(x_b|q_b(s_t))\,,
\end{aligned} \tag{21}$$

the second inequality above following from the mutual independence of $S_{[t]}$, W_a, and W_b.

It thus follows from (20) and (21) that the joint distribution of S_t, $X_t = (X^a_t, X^b_t)$, and Y_t, factorizes as in (4). Hence, (9), (18), and (19), together with (11) and (13), imply that any achievable data rate $R = (R_a, R_b)$ can be written as a convex combination of rate pairs satisfying (3). Hence, any achievable rate pair R belongs to $\overline{\text{co}}(\cup_\mu \mathcal{R}(\mu))$.

Remark: For the validity of the arguments above, a critical step is (20), where the hypothesis of i.i.d.channel state sequence has been used. ◇

IV. Achievability

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 $\text{co}(\cup_\pi \mathcal{R}(\pi))$ 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 $\mathcal{U}_i := \{u_i : \mathcal{V}_i \to \mathcal{X}_i\}$, for $i = a, b$, 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

$$f_i : \mathcal{W}_i \to \mathcal{U}_i^n\,, \qquad i = a, b\,,$$

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

$$\phi^i_t : \mathcal{W}_i \times \mathcal{V}^t_i \to \mathcal{X}_i\,, \quad \phi^i_t(w_i, \boldsymbol{s}_{[t]}) = [f_i(w_i)](s_t)$$

for $i = a, b$, 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 $(\phi^a_t, \phi^b_t, \psi)$ 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 $(\phi^a_t, \phi^b_t, \psi)$ 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 $R = (R_a, R_b) \in \mathbb{R}^2_+$, such that

$$\begin{aligned}
R_a &< I(U_a; Z|U_b) \\
R_b &< I(U_b; Z|U_a) \\
R_a + R_b &< I(U; Z)\,,
\end{aligned} \tag{22}$$

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

$$P(U_a, U_b, Z) = \mu_a(U_a)\mu_b(U_b)Q(Z|U_a, U_b)\,. \tag{23}$$

Consider a sequence of random codes

$$\left\{f^{(n)}_a : \mathcal{W}^{(n)}_a \to \mathcal{U}^n_a\,,\ f^{(n)}_b : \mathcal{W}^{(n)}_b \to \mathcal{U}^n_b\right\}_n$$

where $|\mathcal{W}^{(n)}_a| = \lceil \exp(R_a n)\rceil$ and $|\mathcal{W}^{(n)}_b| = \lceil \exp(R_b n)\rceil$ and $\{f^{(n)}_a(w_a), f^{(n)}_b(w_b)|w_a \in \mathcal{W}^{(n)}_a, w_b \in \mathcal{W}^{(n)}_b\}$ is a collection of independent r.v.s, with $f^{(n)}_i(w_i)$ taking values in $\mathcal{U}^n_i$ with product distribution $\mu_i \otimes \ldots \otimes \mu_i$, for $i = a, b$ 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

$$\mu_i(u_i) = \prod_{v_i \in \mathcal{V}_i} \pi_i(u_i(v_i)|v_i)\,, \quad u_i : \mathcal{V}_i \to \mathcal{X}_i\,,\ i = a, b\,, \tag{24}$$

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 (23), 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

$$\begin{aligned}
I(X_a; Y|S, X_b) &= I(U_a; Z|U_b) \\
I(X_b; Y|S, X_a) &= I(U_b; Z|U_a) \\
I(X; Y|S) &= I(U; Z)\,.
\end{aligned} \tag{25}$$

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 (22) for the product probability distributions μ_a, μ_b defined by (24). As observed above, the direct coding theorem for memoryless MACs implies that such a rate pair is achievable on the MAC Q. As argued above in this section, 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 $\overline{\text{co}}(\cup_\mu \mathcal{R}(\mu))$ then follows from the aforementioned time-sharing principle.

Remark: 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. $\diamond$

V. Discussion on When the Channel State is Markovian

In this section, we briefly comment on the case where the channel state has Markovian dynamics. Consider a Markov chain $\{S_t, t = 1, 2, \dots\}$ taking values in finite state space $\mathcal{S}$ with a stationary transition matrix A such that $P(S_{t+1} = j | S_t = i) = A_{i,j}$. Suppose this Markovian process is observed through a quantized observation process as discussed earlier. In [2] such a setting was studied when there is only one coder and a single letter characterization was obtained as follows:

$$C := \int_{s,\tilde{\pi}} P(d\tilde{\pi}) \sup_{P(X|\tilde{\pi})} \{P(s|\tilde{\pi}) I(X;Y|s,\tilde{\pi})\},$$

where $\tilde{\pi}$ denotes the conditional distribution of the channel state given the quantized observation history.

In 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 case there is only one encoder, then it was shown that the optimal encoder policy uses the conditional probability measure on the state as a sufficient statistic. As such, the optimization problem is tractable.

For the sum-rate, one might consider the mutual information cost $c : \mathcal{S} \times \mathcal{P}(\mathcal{X}_a) \times \mathcal{P}(\mathcal{X}_b) \to \mathbb{R}$ defined by

$$\begin{aligned} c(s, \pi_a, \pi_b) \quad = \quad & \sum_{x_a, x_b, y} \pi_a(x_a) \pi_b(x_b) P(y|s, x_a, x_b) \\ & \log \frac{P(y|s, x_a, x_b)}{\sum_{x_a, x_b} P(y|s, x_a, x_b) \pi_a(x_a) \pi_b(x_b)}, \end{aligned}$$

where the admissible coding policies map, decentrally, the observation histories to $\mathcal{P}(\mathcal{X}_a)$ and $\mathcal{P}(\mathcal{X}_b)$.

For such a setting, however, 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 $(W, V^a_{[t]}, V^b_{[t]})$ to (X^a_t, X^b_t) 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.

For the i.i.d. case, by first showing that the past information is irrelevant, we observed that we could limit the memory space on which the optimization is performed, and as such have a problem which is tractable.

VI. Concluding Remarks

This paper studied the problem of multi-access coding with asymmetric, imperfect channel state information. We observed that one can provide a single letter characterization when the channel state is i.i.d., but when the channel state is Markovian, the problem is more complicated.

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 Trans. Inf. Theory*, vol. 45, pp. 2007-2019, 1999.

[5] S.I. Gelfand and M.S. Pinsker, "Coding for Channel with Random Parameters," *Prob. of Cont. and Info. Theory*, vol. 9, pp. 19-31, 1980.

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

[7] O. Hernndez-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 Trans. Inf. Theory*, vol. 51, pp. 799 - 810, 2005.

[9] 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.

[10] S.A. Jafar "Capacity with causal and non-causal side information - A Unified View", *IEEE Trans. Inf. 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 Trans. Inf. Theory*, vol. 52, pp. 3052-3062, 2006.

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

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

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

[17] S. Yüksel, "Stochastic nestedness and the belief sharing information pattern", *IEEE Trans. Autom. 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.