An interpretation of the Cover and Leung capacity region for the MAC with feedback through stochastic control

Achilleas Anastasopoulos and Kihyuk Sohn
EECS Department, University of Michigan

***Abstract*—We consider the problem of communication over a multiple access channel (MAC) with noiseless feedback. A single-letter characterization of the capacity of this channel is not currently known in general. We formulate the MAC with feedback capacity problem as a stochastic control problem for a special class of channels for which the capacity is known to be the single-letter expression given by Cover and Leung. This approach has been recently successful in finding channel capacity for point-to-point channels with noiseless feedback but has not yet been fruitful in the study of multi-user communication systems. Our interpretation provides an understanding of the role of auxiliary random variables and can also hint at on-line capacity-achieving transmission schemes.**

I. Introduction

Shannon showed in his early work [1] that the capacity of single-user discrete memoryless channel (DMC) does not increase with output feedback. Feedback, however, was shown to be useful in the sense of improving the error performance or simplifying the transmission scheme. In the case of multiple-access channels (MACs), the improvement due to feedback is more dramatic since the the capacity region can be expanded with feedback as Gaarder and Wolf [2] showed. Subsequently, Cover and Leung [3] proposed a block Markov superposition coding scheme for the discrete memoryless MAC (DM-MAC) with feedback and it was shown to be tight for a class of channels [4], while it was shown to be strictly smaller than the capacity region for other channels [5]. Along this line of research, Bross and Lapidoth [6] and Venkataramanan and Pradhan [7] independently improved the Cover and Leung achievable region. In a different line of research, the capacity region was determined by Kramer [8] in terms of directed information. However, this expression is in an incomputable multi-letter form, and thus, a single-letter characterization of the capacity region for the DM-MAC with feedback is still an open problem.

Recently, there has been a significant progress in simplifying multi-letter capacity expressions utilizing a stochastic control framework [9]–[13]. In particular, for finite-state channels (FSCs) with feedback The authors in [10], [13] provided a general stochastic control framework for evaluating the capacity of the FSC with feedback starting from the capacity expressions using directed information. Some multi-user channels have also been studied in a similar way; DM-MAC with feedback was considered in [14]; physically degraded broadcast channel with nested feedback was considered in [15]. These works however, concentrated on finding structural results that simplify the construction of the encoder and decoder, and they didn't address directly the simplification of capacity regions.

In this paper, we attempt an interpretation of the single-letter capacity region for the class of DM-MAC with feedback whose feedback capacity region is known [4]. Towards this goal, we develop a stochastic control framework starting from the feedback capacity region characterization of Kramer [8], and proceed through a two-step simplification process. This process allows us to interpret the role of the auxiliary variable in the capacity expression as some "quantized" state of a dynamical system that is partially controlled by the choice of input distributions and provides a potential methodology for deriving single-letter expressions for other single- or multi-user capacity results.

The rest of the paper is organized as follows. In Section II, the channel model and several known facts are introduced. We formulate, simplify and discuss the stochastic control problem in Section III. Most of the proofs are relegated to the appendices.

In the following, we denote random variables with capital letters $(X, Y, Z, ...)$, their realizations with small letters $(x, y, z, ...)$, and alphabets with calligraphic letters $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, ...)$. A sequence of random variables is denoted with $X^t = (X_1, ..., X_t)$. The space of probability distributions (or equivalently probability mass functions) on the finite alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$.

II. Channel and system model

We consider a two-user DM-MAC. The input symbols X, Y and the output symbol Z take values in the finite alphabets $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, respectively. The channel is memoryless in the sense that the current channel output is independent of all the past channel inputs and the channel outputs, i.e.,

$$P(Z_t|X^t, Y^t, Z^{t-1}) = W(Z_t|X_t, Y_t) \qquad (1)$$

Our model considers feedback, that is the transmission of the channel output from the decoder to both encoders with unit delay. We further assume that the feedback channel is noiseless.

Encoders generate their channel inputs based on their private messages and past outputs. Thus

$$X_t = f_t^1(W_1, Z^{t-1}) \quad (2a)$$
$$Y_t = f_t^2(W_2, Z^{t-1}) \quad (2b)$$

The decoder estimates the messages W_1 and W_2 based on T channel outputs. Hence,

$$(\hat{W}_1, \hat{W}_2) = g(Z^T). \quad (3)$$

Unfortunately, a single-letter capacity expression is not known for this channel. A multi-letter capacity expression for DM-MAC with feedback has been established in [8] and can be stated as follows.

Fact 1 (Theorem 5.1 in [8], [16])**.** *The capacity region of the DM-MAC with feedback is*

$$\mathcal{C}_{FB} = \bigcup_{T=1}^{\infty} \mathcal{C}_T \quad (4)$$

*where $\mathcal{C}_T$, the **directed information Tth inner bound region** (or Tth inner bound region), is defined as $\mathcal{C}_T = co\,(\mathcal{R}_T)$, where $co(A)$ denotes the convex hull of a set A, and*

$$\begin{aligned}\mathcal{R}_T = \bigcup_{\mathcal{P}_T}\{(R_1, R_2) :& 0 \le R_1 \le I_T(X \to Z||Y),\\ & 0 \le R_2 \le I_T(Y \to Z||X),\\ & 0 \le R_1 + R_2 \le I_T(X, Y \to Z)\}, \end{aligned} \quad (5)$$

where $I_T(A \to B||C) = \frac{1}{T}\sum_{t=1}^{T} I(A_t; B_t|C^t, B^{t-1})$. All information quantities are evaluated using the joint distribution

$$P(x^T, y^T, z^T) = \prod_{t=1}^{T} W(z_t|x_t, y_t) q_1(x_t|x^{t-1}, z^{t-1}) q_2(y_t|y^{t-1}, z^{t-1}), \quad (6)$$

and the union is over all input joint distributions on x_t, y_t that are conditionally factorizable as

$$P(x_t, y_t|x^{t-1}, y^{t-1}, z^{t-1}) = q_1(x_t|x^{t-1}, z^{t-1}) \cdot q_2(y_t|y^{t-1}, z^{t-1}) \in \mathcal{P}_T \quad (7)$$

for $t = 1, 2, ..., T$.

Furthermore, the regions $\mathcal{C}_T$ can be expressed in the form

$$\mathcal{C}_T = \Big\{(R_1, R_2) \ge 0 : \forall\, \underline{\lambda} = (\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}_+^3, \lambda_1 R_1 + \lambda_2 R_2 + \lambda_3(R_1 + R_2) \le C_T^{\underline{\lambda}}\Big\}, \quad (8)$$

where

$$C_T^{\underline{\lambda}} \triangleq \sup_{\mathcal{P}_T} \{\lambda_1 I_T(X \to Z||Y) + \lambda_2 I_T(Y \to Z||X) + \lambda_3 I_T(X, Y \to Z)\}. \quad (9)$$

Observe that the problem of evaluating capacity is essentially (at least) as hard as the problem of evaluating the quantity $C_T^{\underline{\lambda}}$ for a given $\underline{\lambda}$. Also note that the optimization problem involved in evaluating $C_T^{\underline{\lambda}}$ can be thought of as a decentralized optimization problem involving two agents: the first is choosing the distribution q_1 on x_t after observing the common information z^{t-1} and his private information x^{t-1}, while the second is choosing the distribution q_2 on y_t after observing the common information z^{t-1} and his private information y^{t-1}. This decentralized nature contributes to the difficulty of this optimization problem. However, as we plan to show, even in the special case when the problem can be transformed to a centralized one, there are difficulties of other nature that need to be overcome before arriving to a single-letter expression from a control-theoretic viewpoint.

We say that the channel is in the family $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$ when the second user can perfectly determine the first user's channel inputs based on its own inputs and the channel outputs. In other words,

$$\mathcal{C}_{\mathcal{Y}\to\mathcal{X}} \triangleq \{W : H(X|Y, Z) = 0\}. \quad (10)$$

For the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$, the capacity region is known to be given by the Cover and Leung region [3], [4], which is in single-letter form.

Fact 2 ([3], [4], [16])**.** *The capacity of DM-MAC in the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$ is the set $\mathcal{C}_{CL} = co(\mathcal{R}_{CL})$, where*

$$\begin{aligned}\mathcal{R}_{CL} = \bigcup_{\mathcal{P}}\{(R_1, R_2) :& 0 \le R_1 \le I(X; Z|Y, V)\\ & 0 \le R_2 \le I(Y; Z|X, V)\\ & 0 \le R_1 + R_2 \le I(X, Y; Z)\}. \end{aligned} \quad (11)$$

All information quantities are evaluated using the joint distribution

$$P_{VXYZ}(v, x, y, z) = P_V(v)P_{X|V}(x|v)P_{Y|V}(y|v)W(z|x, y), \quad (12)$$

where $|\mathcal{V}| \le \min\{|\mathcal{X}||\mathcal{Y}|, |\mathcal{Z}|\}$, and the union is over all input joint distributions on x, y that are conditionally factorizable as

$$P_{XY|V}(x, y|v) = P_{X|V}(x|v)P_{Y|V}(y|v) \in \mathcal{P}. \quad (13)$$

Furthermore, the capacity region $\mathcal{C}_{CL}$ can be expressed in the form

$$\mathcal{C}_{CL} = \Big\{(R_1, R_2) \ge 0 : \forall(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}_+^3, \lambda_1 R_1 + \lambda_2 R_2 + \lambda_3(R_1 + R_2) \le C_{CL}^{\underline{\lambda}}\Big\}, \quad (14)$$

where

$$C_{CL}^{\underline{\lambda}} \triangleq \sup_{\mathcal{P}} \{\lambda_1 I(X; Z|Y, V) + \lambda_2 I(Y; Z|X, V) + \lambda_3 I(X, Y; Z)\}. \quad (15)$$

III. Stochastic Control Problem Formulation

Recently, there have been several results in the literature involving computing the capacity of certain communication channels by formulating the information theory problems into a stochastic control framework [10], [13]. The basic procedure to find a single-letter capacity expression is the following: we start with a multi-letter capacity expression in the form of directed information for the channel of interest. We then formulate a stochastic control problem by introducing an appropriate information state. The optimal solution to this problem (whenever it can be found) implies a single-letter expression for the capacity. This is exactly the methodology we are planning to use in this work.

We first note that for the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$, encoder 2 can perfectly figure out the past input history of user 1 since knowledge of Y^{t-1} and Z^{t-1} gives X^{t-1}. This observation leads to the following simplification of the general capacity expression presented in Fact 1.

Lemma 1. *For the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$, the directed information Tth inner bound region is $\mathcal{C}_T = co(\mathcal{R}'_T)$, where*

$$\mathcal{R}'_T = \bigcup_{\mathcal{P}'_T} \{(R_1,R_2):$$
$$0 \le R_1 \le \frac{1}{T}\sum_{t=1}^{T} I(X_t;Z_t|Y_t,X^{t-1},Z^{t-1})$$
$$0 \le R_2 \le \frac{1}{T}\sum_{t=1}^{T} I(Y_t;Z_t|X_t,X^{t-1},Z^{t-1})$$
$$0 \le R_1+R_2 \le \frac{1}{T}\sum_{t=1}^{T} I(X_t,Y_t;Z_t|Z^{t-1})\ \}. \tag{16}$$

All information quantities are evaluated using the joint distribution

$$P(x^T,y^T,z^T) = \prod_{t=1}^{T} W(z_t|x_t,y_t)q_1(x_t|x^{t-1},z^{t-1})q_2(y_t|x^{t-1},z^{t-1}), \tag{17}$$

and the union is over all input distributions on x_t, y_t that are conditionally factorizable as

$$P(x_t,y_t|x^{t-1},y^{t-1},z^{t-1}) = q_1(x_t|x^{t-1},z^{t-1}) \cdot q_2(y_t|x^{t-1},z^{t-1}) \in \mathcal{P}'_T \tag{18}$$

for $t = 1,2,...,T$, Furthermore, the function $C^{\underline{\lambda}}_T$ in (9) can be simplified as

$$C^{\underline{\lambda}}_T = \sup_{\mathcal{P}'_T}\left\{\frac{1}{T}\sum_{t=1}^{T}\lambda_1 I(X_t;Z_t|Y_t,X^{t-1},Z^{t-1}) + \lambda_2 I(Y_t;Z_t|X_t,X^{t-1},Z^{t-1}) + \lambda_3 I(X_t,Y_t;Z_t|Z^{t-1})\right\}. \tag{19}$$

Proof: See appendix A. ■

The only difference between the above lemma and Fact 1 is that the conditional probability of y_t now depends on x^{t-1} instead of y^{t-1}. The above lemma implies that we can restrict attention to channel input distributions of the form (18) without losing optimality. This in turn means that we can think of this problem as an optimization problem involving a single agent, i.e., a centralized problem: in this setup a single agent chooses both distributions q_1 on x_t and q_2 on y_t after observing the information x^{t-1} and z^{t-1}. We now proceed to formulate an equivalent centralized stochastic control problem in order to further simplify the capacity region expression. Towards this end we introduce the following dynamic system.

- **state** at time t: $(X^{t-1},Z^{t-1}) \in \mathcal{X}^{t-1}\times\mathcal{Z}^{t-1}$
- **observation** at time t: $Z_{t-1} \in \mathcal{Z}$
- **action** at time t: $U_t = (U^1_t,U^2_t) : \mathcal{X}^{t-1} \to \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$. Actions at time t can depend on the observations up to time t and the interpretation is

$$u^1_t[z^{t-1}](x_t|x^{t-1}) = q_1(x_t|x^{t-1},z^{t-1}),$$
$$u^2_t[z^{t-1}](y_t|x^{t-1}) = q_2(y_t|x^{t-1},z^{t-1}) \tag{20}$$

- **instantaneous reward** at time t (given $\underline{\lambda} = (\lambda_1,\lambda_2,\lambda_3)$):

$$R^{\underline{\lambda}}_t = \lambda_1 \log\frac{W(Z_t|X_t,Y_t)}{P(Z_t|Y_t,X^{t-1},Z^{t-1})} + \lambda_2 \log\frac{W(Z_t|X_t,Y_t)}{P(Z_t|X_t,X^{t-1},Z^{t-1})} + \lambda_3 \log\frac{W(Z_t|X_t,Y_t)}{P(Z_t|Z^{t-1})} \tag{21}$$

The control problem is to determine the optimal policy $g = \{g_t\}_{t=1}^T$ (such that $u_t = g_t[z^{t-1}]$) that maximizes the average expected reward $\frac{1}{T}\sum_{t=1}^{T} E^g\{R_t\}$.

Several observations are in order. First, the action U^1_t (and similarly U^2_t) is actually a function; one such action is determined by a probability distribution on x_t for every possible realization of x^{t-1}. This implies that the action space is time varying. Further, the policy g_t defines one such function for every possible realization of z^{t-1}. Second, one might ask why we have not defined the observation at time t as the pair (X_{t-1},Z_{t-1}), and subsequently the action U_t as the distributions $U_t = (U^1_t,U^2_t) \in \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$. Although such definitions would definitely be desirable (since they would lead to time invariant action space) they are inappropriate for this problem because they do not result in expressing the instantaneous reward $R^{\underline{\lambda}}_t$ as a function of the current state and current action. This point is crucial throughout this work and will be made clearer in subsequent proofs.

The introduction of the above dynamical system allows us to view the optimization problem in (19) as an equivalent partially observed Markov decision process (POMDP) and thus provide a characterization/simplification of the solution. Towards this end let us define a random variable $\Theta_t \in \mathcal{P}(\mathcal{X}^t)$ to be a probability distribution on x^t conditioned on Z^t, U^t, i.e., $\Theta_t(x^t) \triangleq P(x^t|Z^t,U^t), \forall x^t \in \mathcal{X}^t$. The following lemma establishes important properties of the random quantity Θ_t and its evolution.

Lemma 2. *There exists a mapping Ψ such that θ_t can be recursively generated as $\theta_t = \Psi(\theta_{t-1},u^1_t,u^2_t,z_t)$. Furthermore, $(\Theta_t)_t$ is a controlled Markov chain with control $u_t = (u^1_t,u^2_t)$, and instantaneous reward $\Xi^{\underline{\lambda}}(\theta_{t-1},u_t)$ i.e., $P(\theta_t|\theta^{t-1},u^t) = P_\Psi(\theta_t|\theta_{t-1},u_t)$, and $E\{R^{\underline{\lambda}}_t|\theta^{t-1},u^t\} = E\{R^{\underline{\lambda}}_t|\theta_{t-1},u_t\} = \Xi^{\underline{\lambda}}(\theta_{t-1},u_t)$.*

Proof: See appendix B. ■

Based on the above lemma we are now ready to state the first main result of this work.

Proposition 1. *The Tth inner bound region for the DM-MAC in the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$ is given by (8) with the quantity $C_T^{\underline{\lambda}}$ evaluated as*

$$C_T^{\underline{\lambda}} = \sup_{\bar{\mathcal{P}}} \left\{ \frac{1}{T}\sum_{t=1}^{T} \lambda_1 I(X_t;Z_t|Y_t,X^{t-1},\Theta_{t-1}) + \lambda_2 I(Y_t;Z_t|X_t,X^{t-1},\Theta_{t-1}) + \lambda_3 I(X_t,Y_t;Z_t|\Theta_{t-1}) \right\}. \tag{22}$$

All mutual information quantities are evaluated using the joint distribution

$$P(x^T,y^T,z^T,\theta^T) = \prod_{t=1}^{T} W(z_t|x_t,y_t)\overline{q}_1(x_t|x^{t-1},\theta_{t-1}) \overline{q}_2(y_t|x^{t-1},\theta_{t-1})\theta_{t-1}(x^{t-1})\delta_{\Psi(\theta_{t-2},\overline{q}_1,\overline{q}_2,z_{t-1})}(\theta_{t-1}), \tag{23}$$

and the supremum is over all input distributions of the form

$$P(x_t,y_t|x^{t-1},y^{t-1},z^{t-1}) = \overline{q}_1(x_t|x^{t-1},\theta_{t-1})\overline{q}_2(y_t|x^{t-1},\theta_{t-1}) \in \bar{\mathcal{P}}. \tag{24}$$

Proof: Due to Lemma 2 and using standard POMDP results, the optimal strategy g_t is a Markov policy (i.e., it is only a function of the state θ_{t-1}). This implies that the optimizing distributions can be of the form $\overline{q}_1(x_t|x^{t-1},\theta_{t-1})\overline{q}_2(y_t|y^{t-1},\theta_{t-1})$. Furthermore, using an inductive argument similar to the one used in the proof of Lemma 1 we can show the equivalence between the stochastic control problem and the optimization problem in (19), which concludes the proof. ■

Comparing (19) and (22) one can observe that although the dependence on the previous observations Z^t has been summarized in the "state" Θ_t, this does not result in a single-letter form for the quantity $C_T^{\underline{\lambda}}$. Furthermore, replacing Z^t with Θ_t can hardly be considered a simplification since the former belongs to the expanding alphabet $\mathcal{Z}^t$ while the latter belongs to an even larger expanding alphabet $\mathcal{P}(\mathcal{X}^t)$. The dependence of the new variables in this dynamical system is shown in Fig 1.

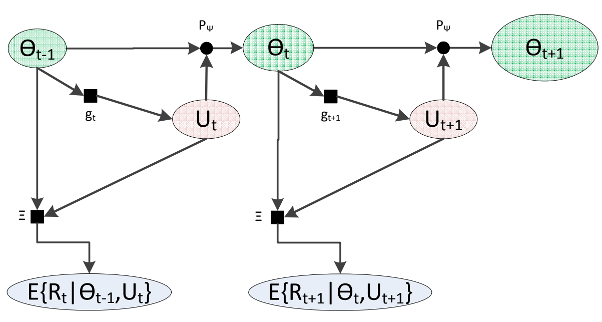

Fig. 1. Markov decision process evolution

Since the capacity expression for this class of channels is known to be in a single-letter form [4], a reasonable question to ask is how this expression comes about in the stochastic control framework developed thus far. In the following we show that there is additional structure in the problem that allows us further simplification.

Lemma 3. *For every action $u_t: \mathcal{X}^{t-1} \to \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ and every distribution $\theta_{t-1}\in\mathcal{P}(\mathcal{X}^{t-1})$, there exist a distribution $\phi_{t-1}\in\mathcal{P}(\mathcal{V})$, and an action $\hat{u}_t: \mathcal{V}\to\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$, such that the instantaneous reward $\Xi^{\underline{\lambda}}(\theta_{t-1},u_t)$ can be written as*

$$\begin{aligned}
&\Xi^{\underline{\lambda}}(\theta_{t-1},u_t) = \\
&\sum_{x_t,y_t,z_t} W(z_t|x_t,y_t) \\
&\sum_{v_{t-1}\in\mathcal{V}} \hat{u}_t^1(x_t|v_{t-1})\hat{u}_t^2(y_t|v_{t-1})\phi_{t-1}(v_{t-1}) \\
&\times\Bigg[\lambda_1 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{x}_t} W(z_t|\tilde{x}_t,y_t)\hat{u}_t^1(\tilde{x}_t|v_{t-1})} + \\
&\quad \lambda_2 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{y}_t} W(z_t|x_t,\tilde{y}_t)\hat{u}_t^2(\tilde{y}_t|v_{t-1})} + \\
&\quad \lambda_3 \log\Bigg(W(z_t|x_t,y_t)/[\sum_{\tilde{x}_t,\tilde{y}_t} W(z_t|\tilde{x}_t,\tilde{y}_t) \\
&\qquad \sum_{\tilde{v}_{t-1}} \hat{u}_t^1(\tilde{x}_t|\tilde{v}_{t-1})\hat{u}_t^2(\tilde{y}_t|\tilde{v}_{t-1})\phi_{t-1}(\tilde{v}_{t-1})]\Bigg) \qquad (25a)\\
&= \lambda_1 I(X_t;Z_t|Y_t,V_{t-1}) + \lambda_2 I(Y_t;Z_t|X_t,V_{t-1}) + \\
&\quad \lambda_3 I(X_t,Y_t;Z_t) \qquad (25b)\\
&= \hat{\Xi}^{\underline{\lambda}}(\phi_{t-1},\hat{u}_t), \qquad (25c)
\end{aligned}$$

where the mutual information quantities are evaluated using the distribution

$$P(x_t,y_t,z_t,v_{t-1}) = W(z_t|x_t,y_t)\hat{u}_t^1(x_t|v_{t-1})\hat{u}_t^2(y_t|v_{t-1})\phi_{t-1}(v_{t-1}). \tag{26}$$

Furthermore, the cardinality of $\mathcal{V}$ can be bounded by $|\mathcal{V}| \le |\mathcal{X}||\mathcal{Y}|$.

Proof: This is a consequence of Caratheodory's theorem (and its application by Ahlswede and Körner), as described in [16]. ■

Observe that the reward function in (25b) is exactly the reward relevant to the Cover and Leung region as shown in (15) in Fact 2. Thus, the significance of this result is that it establishes the connection between the stochastic control framework and the auxiliary random variable found in the single-letter expression from the information theoretic viewpoint. An important caveat of the above result is that it is an *existence* result, i.e., it does not construct the distribution ϕ but only guarantees its existence. The new quantities ϕ_{t-1} and $\hat{u}_t$ are interconnected with the remaining dynamical system as shown in Fig. 2.

One can think of the trajectory of this dynamical system as follows: given the optimal choice of the strategy g, for every realization of the state Θ_{t-1} an optimal action $U_t = g_t[\Theta_{t-1}]$ is generated (Markov policy). This action, together with the state Θ_{t-1} and the "system disturbance" will lead to a new state Θ_t and will also generate an optimal reward $\Xi^{\underline{\lambda}}(\Theta_{t-1},U_t)$. The last result implies that there exists a – yet

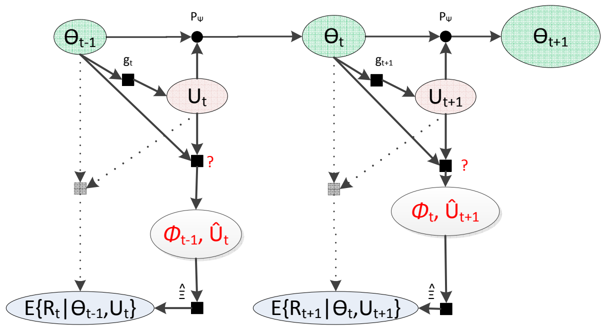

Fig. 2. Markov decision process evolution with reduced states and actions.

to be found – "quantization" of the alphabet $\mathcal{X}^{t-1}$ into a time-invariant alphabet $\mathcal{V}$ such that an equivalent action $\hat{U}_t$ need only be defined in this smaller space $\mathcal{V} \to \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ that results in the same exact reward.

Since we know that in the long run the expected reward per unit time has to be equal to C_{CL}^{Δ} in (15) and since (due to the last lemma) the reward expressions in (25b) and (15) are identical, one should expect that the optimal choice of actions will lead in the long run to states Θ_{t-1} which, when quantized, will result in "reduced" states Φ_{t-1} and actions $\hat{U}_t$ that are approaching asymptotically, the optimizing distributions P_V^* and $P_{X|V}^*, P_{Y|V}^*$ of (15), respectively. In other words, $\lim_{t\to\infty} \Phi_t = P_V^*$ and $\lim_{t\to\infty} \hat{U}_t = (P_{X|V}^*, P_{Y|V}^*)$ (where the starred quantities are the supremizing distributions in (15) and the limits are understood in some appropriate sense).

At this point it not clear how such a conclusion can be derived directly through the control theoretic framework without resorting to the known single-letter information theoretic result. The resolution of this question hinges on finding a "quantization" of the space $\mathcal{X}^{t-1}$ that together with Θ_{t-1} induces the distribution Φ_{t-1} and showing that with the right choice of "reduced" actions $\hat{U}_t$ this "quantized" distribution converges to P_V^*. This is a research direction we are currently pursuing.

Another research direction of interest is the extension of our methodology to the general DM-MAC with feedback to find single-letter expression of capacity region. The main difference between the general case and the case we discussed in the paper is that the general case results in a decentralized control problem.

Finally, a third interesting research direction is to investigate a simple sequential transmission scheme using the idea of the posterior matching scheme [17] that achieves any rate pair on the capacity region of the DM-MAC with feedback.

Appendix

A. Proof of Lemma 1

Note that the encoder 2 has a perfect knowledge of the encoder 1's past input history through the feedback information and its own history of input. Thus, the input distributions of interest are of form

$$P(x_t, y_t | x^{t-1}, y^{t-1}, z^{t-1}) = q_1(x_t | x^{t-1}, z^{t-1}) \cdot q_2(y_t | x^{t-1}, y^{t-1}, z^{t-1}). \tag{27}$$

We can similarly get

$$\begin{aligned}
&I(X_t; Z_t | Y_t, Y^{t-1}, Z^{t-1}) \\
&= I(X_t; Z_t | Y_t, X^{t-1}, Y^{t-1}, Z^{t-1}) && \text{(28a)} \\
&= E\left[\log \frac{W(Z_t|X_t, Y_t)}{\sum_{x_t} W(Z_t|x_t, Y_t) P(x_t|X^{t-1}, Y^{t-1}, Y_t, Z^{t-1})}\right] && \text{(28b)} \\
&= E\left[\log \frac{W(Z_t|X_t, Y_t)}{\sum_{x_t} W(Z_t|x_t, Y_t) q_1(x_t|X^{t-1}, Z^{t-1})}\right] && \text{(28c)} \\
&= I(X_t; Z_t | Y_t, X^{t-1}, Z^{t-1}) && \text{(28d)}
\end{aligned}$$

where (28c) is due to the conditional independence of X_t and Y^t given (X^{t-1}, Z^{t-1}). Note that the information theoretic quantities that appear in the bounds of (16) for each time t are evaluated based on the joint distribution $P(x^t, y_t, z^t)$. We now proceed by induction to show that for every sequence of input distributions $\{q_1(x_t|x^{t-1}, z^{t-1}),\ q_2(y_t|y^{t-1}, z^{t-1})\}_{t=1}^T$ inducing the sequence of measures $\{P_q(x^t, y_t, z^t)\}_{t=1}^T$, there exists a sequence of input distribution $\{q_1(x_t|x^{t-1}, z^{t-1}),\ \hat{q}_2(y_t|x^{t-1}, z^{t-1})\}$ which induces the same sequence of measures $\{\hat{P}(x^t, y_t, z^t)\}_{t=1}^T$.

For $t = 1$ we set $\hat{q}_2(y_1) = q_2(y_1)$ and have

$$\begin{aligned}
\hat{P}(x^1, y_1, z^1) &= W(z_1|x_1, y_1) q_1(x_1) \hat{q}_2(y_1) && \text{(29a)} \\
&= W(z_1|x_1, y_1) q_1(x_1) q_2(y_1) && \text{(29b)} \\
&= P(x^1, y_1, z^1). && \text{(29c)}
\end{aligned}$$

Now for $t+1$ we set $\hat{q}_2(y_{t+1}|x^t, z^t) = P_q(y_{t+1}|x^t, z^t) = \frac{\sum_{y^t} q_2(y_{t+1}|y^t, z^t) P_q(x^t, y^t, z^t)}{\sum_{y^t} P_q(x^t, y^t, z^t)}$ and have

$$\begin{aligned}
&\hat{P}(x^{t+1}, y_{t+1}, z^{t+1}) \\
&= W(z_{t+1}|x_{t+1}, y_{t+1}) q_1(x_{t+1}|x^t, z^t) \hat{q}_2(y_{t+1}|x^t, z^t) \sum_{y_t} \hat{P}(x^t, y_t, z^t) && \text{(30a)} \\
&= W(z_{t+1}|x_{t+1}, y_{t+1}) q_1(x_{t+1}|x^t, z^t) P_q(y_{t+1}|x^t, z^t) \sum_{y_t} P_q(x^t, y_t, z^t) && \text{(30b)} \\
&= P_q(x^{t+1}, y_{t+1}, z^{t+1}) && \text{(30c)}
\end{aligned}$$

where (30b) is due to the induction hypothesis and the construction of $\hat{q}_2(y_{t+1}|x^t, z^t)$.

The remaining part of the proof employs a result in [16] which utilized the convexity property of the capacity region of DM-MAC with feedback.

B. Proof of Lemma 2

For every $x^t \in \mathcal{X}^t$, we have

$$\begin{aligned}
&\theta_t(x^t) \\
&= P(x^t|z^t, u^t) && \text{(31a)}\\
&= \frac{P(z_t, x^t|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} && \text{(31b)}\\
&= \frac{P(z_t, x_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} && \text{(31c)}\\
&= \frac{\sum_{y_t} P(z_t, x_t, y_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} && \text{(31d)}\\
&= [\sum_{y_t} W(z_t|x_t, y_t)P(x_t|x^{t-1}, z^{t-1}, u^t) \\
&\quad P(y_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)]/P(z_t|z^{t-1}, u^t) && \text{(31e)}\\
&= [\sum_{y_t} W(z_t|x_t, y_t)u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1})]/ \\
&[\sum_{\tilde{x}_t, \tilde{y}_t, \tilde{x}^{t-1}} W(z_t|\tilde{x}_t, \tilde{y}_t)u_t^1(\tilde{x}_t|\tilde{x}^{t-1})u_t^2(\tilde{y}_t|\tilde{x}^{t-1})\theta_{t-1}(\tilde{x}^{t-1})] && \text{(31f)}
\end{aligned}$$

which establishes $\theta_t = \Psi(\theta_{t-1}, u_t^1, u_t^2, z_t)$. Furthermore,

$$\begin{aligned}
&P(\theta_t|\theta^{t-1}, u^t) && \text{(32a)}\\
&= \sum_{x^t, y_t, z^t} P(\theta_t|\theta^{t-1}, u^t, z^t, x^t, y_t)P(z_t|\theta^{t-1}, u^t, z^{t-1}, x^t, y_t) \\
&\quad P(x_t, y_t|\theta^{t-1}, u^t, z^{t-1}, x^{t-1})P(x^{t-1}, z^{t-1}|\theta^{t-1}, u^t) && \text{(32b)}\\
&= \sum_{z_t} \delta_{\Psi(\theta_{t-1}, u_t^1, u_t^2, z_t)}(\theta_t) \sum_{x_t, y_t} W(z_t|x_t, y_t) \\
&\qquad \sum_{x^{t-1}} u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1}) && \text{(32c)}\\
&= P_\Psi(\theta_t|\theta_{t-1}, u_t). && \text{(32d)}
\end{aligned}$$

Finally, the expected reward at time t conditioned on the information states θ^{t-1} and the control actions u^t is

$$\begin{aligned}
&E\left\{R_t^{\underline{\lambda}}|\theta^{t-1}, u^t\right\} \\
&= \sum_{x_t, y_t, z_t} W(z_t|x_t, y_t) \\
&\quad \sum_{x^{t-1}} u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1}) \\
&\quad \times \Bigg[\lambda_1 \log \frac{W(z_t|x_t, y_t)}{\sum_{\tilde{x}_t} W(z_t|\tilde{x}_t, y_t)u_t^1(\tilde{x}_t|x^{t-1})} + \\
&\quad \lambda_2 \log \frac{W(z_t|x_t, y_t)}{\sum_{\tilde{y}_t} W(z_t|x_t, \tilde{y}_t)u_t^2(\tilde{y}_t|x^{t-1})} + \\
&\quad \lambda_3 \log \Bigg(W(z_t|x_t, y_t)/[\sum_{\tilde{x}_t, \tilde{y}_t} W(z_t|\tilde{x}_t, \tilde{y}_t) \\
&\quad \sum_{\tilde{x}^{t-1}} u_t^1(\tilde{x}_t|\tilde{x}^{t-1})u_t^2(\tilde{y}_t|\tilde{x}^{t-1})\theta_{t-1}(\tilde{x}^{t-1})] \Bigg) \Bigg] && \text{(33a)}\\
&\triangleq \Xi^{\underline{\lambda}}(\theta_{t-1}, u_t), && \text{(33b)}
\end{aligned}$$

which is only a function of the information state θ_{t-1} and the action $u_t = (u_t^1, u_t^2)$.

REFERENCES

[1] C. Shannon, "The zero error capacity of a noisy channel," *Information Theory, IRE Transactions on*, vol. 2, no. 3, pp. 8–19, September 1956.

[2] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 100–102, Jan 1975.

[3] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Information Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[4] F. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Information Theory*, vol. 28, no. 1, pp. 93–95, Jan. 1982.

[5] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Information Theory*, vol. 30, no. 4, pp. 623–629, Jul 1984.

[6] S. I. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 811–833, March 2005.

[7] R. Venkataramanan and S. S. Pradhan, "A new achievable rate region for the discrete memoryless multiple-access channel with feedback," in *Proc. International Symposium on Information Theory*, June 2009.

[8] G. Kramer, "Directed information for channels with feedback," Ph.D. dissertation, ETH Series in Information Processing. Konstanz, Switzerland: Hartung-Gorre Verlag, 1998.

[9] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 780–798, March 2005.

[10] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Information Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[11] G. Como, S. Yuksel, and S. Tatikonda, "The error exponent of variable-length codes over markov channels with feedback," *IEEE Trans. Inform. Theory*, vol. 55, no. 5, pp. 2139–2160, May 2009.

[12] A. Kavcic, D. Mandic, X. Huang, and X. Ma, "Upper bounds on the capacities of non-controllable finite-state channels using dynamic programming methods," in *Proc. International Symposium on Information Theory*, June 28 2009–July 3 2009, pp. 2346–2350.

[13] J. H. Bae and A. Anastasopoulos, "A posterior matching scheme for finite-state channels with feedback," in *Proc. International Symposium on Information Theory*, Austin, TX, June 2010, pp. 2338 –2342.

[14] A. Anastasopoulos, "A sequential transmission scheme for the multiple access channel with noiseless feedback," in *Proc. Allerton Conf. Commun., Control, Comp.*, Monticello, Illinois, 2009.

[15] A. Mahajan, "Optimal sequential transmission over degraded broadcast channels with nested feedback," in *Proc. Allerton Conf. Commun., Control, Comp.*, Monticello, IL, Sept. 2009, pp. 795–802.

[16] M. Salehi, "Cardinality bounds on auxiliary variables in multiple-user theory via the method of ahlswede and korner," Stanford Univ., Stanford, CA, Tech. Rep. 33, Aug 1978.

[17] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Trans. Information Theory*, vol. 57, no. 3, pp. 1186–1222, Mar. 2011.