A stochastic control interpretation of the Cover and Leung region for the MAC with noiseless feedback

Kihyuk Sohn and Achilleas Anastasopoulos

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. Several schemes exist in the literature that provide achievable region expressions involving a number of auxiliary variables.

A series of recent results in point-to-point communications with noiseless feedback study the information theoretic problem of finding channel capacity in a stochastic control framework. This approach has not yet been fruitful in the study of multi-user communication scenaria.

In this paper 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 Cover and Leung region. This interpretation provides an understanding of the role of auxiliary random variables. In addition, through this interpretation a useful single-letter outer bound is derived. For the special case studied in this paper, this outer bound is tight.

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. When it comes to the multiple-access channels (MACs), the improvement becomes more dramatic since Gaarder and Wolf [2] showed that the capacity region can be expanded with output feedback. Subsequently, the capacity region for the MAC with feedback has been studied. 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. An outer bound was derived in [8],

[9] using a dependence balance bound which was shown to be tighter than the cut-set bound. The capacity region was determined by Kramer [10] 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 [11], [12], [12]–[15]. In particular, for finite-state channels (FSCs) with feedback The authors in [12], [15] provided a general stochastic control framework for evaluating the capacity of the FSC with feedback starting from the capacity expressions using directed information. Several multi-user channels have also been studied in a similar way; DM-MAC with feedback was considered in [16]; physically degraded broadcast channel with nested feedback was considered in [17]. 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 provide 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 using directed information expressions, and proceed through a two-step simplification process. This process allows us to interpret the role of the auxiliary variable in the capacity expression and provides a methodology for deriving single-letter outer bounds.

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

II. Preliminaries

A. 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.,

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

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

$$X_t = f_t^1(W_1, Z^{t-1}) \tag{2a}$$

$$Y_t = f_t^2(W_2, Z^{t-1}) \tag{2b}$$

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

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

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,

$$\mathcal{C}_{\mathcal{Y}\to\mathcal{X}} \triangleq \{W : H(X|Y,Z) = 0\}. \tag{4}$$

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].

Fact 1 ([3], [4], [18]). *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*

$$\mathcal{R}_{CL} = \bigcup_{P_{VXY}} \left\{ (R_1, R_2) : \begin{array}{ccccc} 0 & \leq & R_1 & \leq & I(X;Z|Y,V) \\ 0 & \leq & R_2 & \leq & I(Y;Z|X,V) \\ 0 & \leq & R_1 + R_2 & \leq & I(X,Y;Z) \end{array} \right\}, \tag{5}$$

where $co(A)$ denotes the convex hull of a set A, and all information quantities are evaluated using the joint distribution

$$P_{VXYZ}(v,x,y,z) = P_{VXY}(v,x,y)W(z|x,y) = P_V(v)P_{X|V}(x|v)P_{Y|V}(y|v)W(z|x,y), \tag{6}$$

where $|\mathcal{V}| \leq \min\{|\mathcal{X}||\mathcal{Y}|, |\mathcal{Z}|\}$. Furthermore, the capacity region $\mathcal{C}_{CL}$ can be expressed in the form

$$\mathcal{C}_{CL} = \{(R_1, R_2) \geq 0 : \forall (\lambda_1, \lambda_2, \lambda_3) \geq 0, \lambda_1 R_1 + \lambda_2 R_2 + \lambda_3 (R_1 + R_2) \leq C_{CL}(\underline{\lambda})\}, \tag{7}$$

where

$$C_{CL}(\underline{\lambda}) \triangleq \sup_{P_{VXY}} \{\lambda_1 I(X;Z|Y,V) + \lambda_2 I(Y;Z|X,V) + \lambda_3 I(X,Y;Z)\}. \tag{8}$$

B. Directed Information

There has been a dramatic improvement in computing the capacity of several communication channels by formulating the information theory problems into the stochastic control framework [12], [15]. The common 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. We follow a similar procedure in this paper. Our starting point is a multi-letter capacity expression for DM-MAC with feedback using directed information.

Fact 2 (Theorem 5.1 in [10], [4])**.** *The capacity region of the DM-MAC with feedback is*

$$\mathcal{C}_{FB} = \lim_{T \to \infty} \mathcal{R}_T \tag{9}$$

*where $\mathcal{R}_T$, the **directed information Tth inner bound region** (or Tth inner bound region), is defined as $\mathcal{R}_T = co\left(\mathcal{R}_T^1\right)$, with*

$$\mathcal{R}_T^1 = \bigcup_{\mathcal{P}_T^1} \left\{ (R_1, R_2) : \begin{array}{ccccc} 0 & \leq & R_1 & \leq & I_T(X \to Z||Y) \\ 0 & \leq & R_2 & \leq & I_T(Y \to Z||X) \\ 0 & \leq & R_1 + R_2 & \leq & I_T(X, Y \to Z) \end{array} \right\}, \tag{10}$$

where $I_T(A \to B||C) = \frac{1}{T}\sum_{t=1}^{T} I(A_t; B_t|C^t, B^{t-1})$. The union is over all input distributions

$$P(x_t, y_t|x^{t-1}, y^{t-1}, z^{t-1}) = q_1(x_t|x^{t-1}, z^{t-1}) \cdot q_2(y_t|y^{t-1}, z^{t-1}) \in \mathcal{P}_T^1 \tag{11}$$

for $t = 1, 2, ..., T$, and all information quantities are evaluated using the corresponding joint distribution

$$P(x^T, y^T, z^T) = \prod_{t=1}^{T} W(z_t|x_t, y_t) q_1(x_t|x^{t-1}, z^{t-1}) q_2(y_t|y^{t-1}, z^{t-1}). \tag{12}$$

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

$$\mathcal{R}_T = \left\{(R_1, R_2) \geq 0 : \forall(\lambda_1, \lambda_2, \lambda_3) \geq 0, \lambda_1 R_1 + \lambda_2 R_2 + \lambda_3 (R_1 + R_2) \leq C_T(\underline{\lambda})\right\}, \tag{13}$$

where

$$C_T(\underline{\lambda}) \triangleq \sup_{\mathcal{P}_T^1} \left\{\lambda_1 I_T(X \to Z||Y) + \lambda_2 I_T(Y \to Z||X) + \lambda_3 I_T(X, Y \to Z)\right\}. \tag{14}$$

C. Notation

We denote random variables with capital letters $(X, Y, Z, ...)$, their realizations with small letters $(x, y, z, ...)$, and alphabets with caligraphic letters $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, ...)$. A sequence of random variables is denoted with $X^t = (X_1, ..., X_t)$.

III. Stochastic Control Problem Formulation

Before going into the details, we first summarize and categorize the random variables of interest based on the accessibility to agents. The common information for all three agents (i.e., encoder 1, 2 and decoder) is the past output history Z^{t-1}. In addition to that, encoder 1 knows his private message W_1 and his input history X^{t-1}, and similarly for encoder 2. 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.

Lemma 1. *For the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$, the directed information Tth inner bound region is $\mathcal{R}_T = co(\mathcal{R}_T^2)$, where*

$$\mathcal{R}_T^2 = \bigcup_{\mathcal{P}_T^2} \left\{ (R_1, R_2) : \begin{array}{ccccc} 0 & \leq & R_1 & \leq & \frac{1}{T}\sum_{t=1}^T I(X_t; Z_t | Y_t, X^{t-1}, Z^{t-1}) \\ 0 & \leq & R_2 & \leq & \frac{1}{T}\sum_{t=1}^T I(Y_t; Z_t | X_t, X^{t-1}, Z^{t-1}) \\ 0 & \leq & R_1 + R_2 & \leq & \frac{1}{T}\sum_{t=1}^T I(X_t, Y_t; Z_t | Z^{t-1}) \end{array} \right\}. \tag{15}$$

The union is over all input distributions

$$P(x_t, y_t | x^{t-1}, y^{t-1}, z^{t-1}) = q_1(x_t | x^{t-1}, z^{t-1}) \cdot q_2(y_t | x^{t-1}, z^{t-1}) \in \mathcal{P}_T^2 \tag{16}$$

for $t = 1, 2, ..., T$, and all information quantities are evaluated using the corresponding joint distribution

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

Furthermore, the function $C_T(\underline{\lambda})$ in (14) *can be simplified as*

$$C_T(\underline{\lambda}) = \sup_{\mathcal{P}_T^2} \left\{ \frac{1}{T} \sum_{t=1}^T \lambda_1 I(X_t; Z_t | Y_t, X^{t-1}, Z^{t-1}) + \lambda_2 I(Y_t; Z_t | X_t, X^{t-1}, Z^{t-1}) + \lambda_3 I(X_t, Y_t; Z_t | Z^{t-1}) \right\}. \tag{18}$$

Proof: See appendix A. ■

The above lemma implies that we can restrict attention to channel input distributions of the form (16) without losing optimality. In addition the problem of finding the capacity is reduced to the maximization of a single quantity. We now formulate an equivalent stochastic control problem 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: $U_t = \left(U_t^1, U_t^2\right) : \mathcal{X}^{t-1} \to \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$. Actions at time t can depend on the observations up to time t and the interpretation is

$$u_t^1[z^{t-1}](x_t | x^{t-1}) = q_1(x_t | x^{t-1}, z^{t-1}), u_t^2[z^{t-1}](y_t | x^{t-1}) = q_2(y_t | x^{t-1}, z^{t-1}) \tag{19}$$

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

$$R_t(\underline{\lambda}) = \lambda_1 \log \frac{W(Z_t | X_t, Y_t)}{P(Z_t | Y_t, X^{t-1}, Z^{t-1})} + \lambda_2 \log \frac{W(Z_t | X_t, Y_t)}{P(Z_t | X_t, X^{t-1}, Z^{t-1})} + \lambda_3 \log \frac{W(Z_t | X_t, Y_t)}{P(Z_t | Z^{t-1})} \tag{20}$$

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 $\frac{1}{T}\sum_{t=1}^T E^g[R_t]$.

As a next step of simplification, we summarize Z^{t-1}, which is the common information available to both encoders and decoder, using Markov decision process (MDP) theory. By Lemma 1, we don't lose optimality in the sense of capacity region if we ignore the past input alphabets of encoder 2 while generating the channel inputs. Let us define a random variable $\Theta_t \in \mathcal{P}(\mathcal{X}^t)$ as the decoder's estimates about the channel inputs based on the information known to decoder, i.e., $\Theta_t(x^t) \triangleq P(x^t|Z^t, U^t), \forall x^t \in \mathcal{X}^t$.

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

Proof: See appendix B. ∎

The following simplification is now possible.

Proposition 1. *The Tth inner bound region for the DM-MAC in the class $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$ is given by* (13) *with the quantity $C_T(\underline{\lambda})$ evaluated as*

$$C_T(\underline{\lambda}) = \sup_{\bar{\mathcal{P}}} \left\{ \frac{1}{T} \sum_{t=1}^{T} \lambda_1 I(X_t; Z_t|Y_t, X^{t-1}, \Theta_{t-1}) + \lambda_2 I(Y_t; Z_t|X_t, X^{t-1}, \Theta_{t-1}) + \lambda_3 I(X_t, Y_t; Z_t|\Theta_{t-1}) \right\} \tag{21}$$

where the supremum is over all input distributions

$$P(x_t, y_t|x^{t-1}, y^{t-1}, z^{t-1}) = \bar{q}_1(x_t|x^{t-1}, \theta_{t-1})\bar{q}_2(y_t|x^{t-1}, \theta_{t-1}) \tag{22}$$

and the mutual information quantities evaluated using the joint distribution

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

Proof: See appendix C. ∎

As can be seen, the simplification above does not result in a single-letter form for the capacity region of DM-MAC with feedback. Furthermore, the domain of information state is not time-invariant and grows exponentially with time. 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 to reduce the action space.

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* $\overline{r}_t^{\lambda}(\theta_{t-1}, u_t)$ *can be written as*

$$\overline{r}_t^{\lambda}(\theta_{t-1}, u_t) = \sum_{x_t, y_t, z_t} W(z_t|x_t, y_t) \sum_{v_{t-1} \in \mathcal{V}} \hat{u}_t^1(x_t|v_{t-1})\hat{u}_t^2(y_t|v_{t-1})\phi_{t-1}(v_{t-1})$$
$$\times \Bigg[\lambda_1 \log \frac{W(z_t|x_t, y_t)}{\sum_{\tilde{x}_t} W(z_t|\tilde{x}_t, y_t)\hat{u}_t^1(\tilde{x}_t|v_{t-1})} + \lambda_2 \log \frac{W(z_t|x_t, y_t)}{\sum_{\tilde{y}_t} W(z_t|x_t, \tilde{y}_t)\hat{u}_t^2(\tilde{y}_t|v_{t-1})}$$
$$+\lambda_3 \log \frac{W(z_t|x_t, y_t)}{\sum_{\tilde{x}_t, \tilde{y}_t} W(z_t|\tilde{x}_t, \tilde{y}_t) \sum_{\tilde{v}_{t-1}} \hat{u}_t^1(\tilde{x}_t|\tilde{v}_{t-1})\hat{u}_t^2(\tilde{y}_t|\tilde{v}_{t-1})\phi_{t-1}(\tilde{v}_{t-1})}\Bigg] \quad (24)$$
$$= \lambda_1 I(X_t; Z_t|Y_t, V_{t-1}) + \lambda_2 I(Y_t; Z_t|X_t, V_{t-1}) + \lambda_3 I(X_t, Y_t; Z_t) \quad (25)$$
$$\triangleq \hat{r}^{\lambda}(\phi_{t-1}, \hat{u}_t), \quad (26)$$

where the mutual information quantities are evaluated using the distribution

$$P(x_t, y_t, z_t, v_{t-1}) = W(z_t|x_t, y_t)\hat{u}_t^1(x_t|v_{t-1})\hat{u}_t^2(y_t|v_{t-1})\phi_{t-1}(v_{t-1}). \quad (27)$$

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

Proof: This is a consequence of Caratheodory's theorem (and its application by Ahlswede and Körner), as described in [18]. ■

Observe that the reward function in (25) is exactly the reward relevant to the Cover and Leung region as shown in (8) in Fact 1. Consequently, if we maximize over both the actions $\hat{u}_t$ and the state ϕ_{t-1} we can obtain an outer bound on the capacity region of interest. This result is summarized following proposition.

Proposition 2. *The capacity region of the DM-MAC with feedback in the class* $\mathcal{C}_{\mathcal{Y}\to\mathcal{X}}$ *is outer bounded by*

$$C_{FB} \subset \{(R_1, R_2) \geq 0 : \forall(\lambda_1, \lambda_2, \lambda_3) \geq 0, \lambda_1 R_1 + \lambda_2 R_2 + \lambda_3(R_1 + R_2) \leq C(\underline{\lambda})\}, \quad (28)$$

where $C(\underline{\lambda})$ *is defined in* (8)*.*

Proof: Due to Proposition 1, and Lemma 3 the Tth inner bound $\mathcal{R}_T^3$ is outer bounded by the left-hand side of (28). Since the latter is independent of T it is also an outer bound of C_{FB} due to Fact 2, Lemma 1, and Fact 2. ■

Since it is known that this outer bound is tight it is clear that the described dynamical system follows a trajectory where $\lim_{t\to\infty} \Phi_t = P_V^*$ and $\lim_{t\to\infty} \hat{U}_t = (P_{X|V}^*, P_{Y|V}^*)$ (where the starred quantities are the supremizing distributions in Proposition 2). 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^*.

Another direction is a direct 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 the type of information pattern.

Another interesting research direction is to investigate a simple sequential transmission scheme using the idea of the posterior matching scheme [19], [20] 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

$$P(x_t, y_t|x^{t-1}, y^{t-1}, z^{t-1}) = q_1(x_t|x^{t-1}, z^{t-1}) \cdot q_2(y_t|x^{t-1}, y^{t-1}, z^{t-1}) \tag{29}$$

If we reformulate the bound for R_1 at time t in the same way, we get

$$\begin{aligned} I(X_t; Z_t|Y_t, Y^{t-1}, Z^{t-1}) &= I(X_t; Z_t|Y_t, X^{t-1}, Y^{t-1}, Z^{t-1}) &(30)\\ &= E\left[\log \frac{W(Z_t|X_t, Y_t)}{\sum_{x_t} W(Z_t|x_t, Y_t) P(x_t|X^{t-1}, Y^{t-1}, Y_t, Z^{t-1})}\right] &(31)\\ &= E\left[\log \frac{W(Z_t|X_t, Y_t)}{\sum_{x_t} W(Z_t|x_t, Y_t) q(x_t|X^{t-1}, Z^{t-1})}\right] &(32)\\ &= I(X_t; Z_t|Y_t, X^{t-1}, Z^{t-1}) &(33)\end{aligned}$$

where (32) is due to the conditional independence of X_t and Y^t given (X^{t-1}, Z^{t-1}).

Let's define the bounds at time t as $I_{1,t}, I_{2,t}$ and $I_{3,t}$,

$$I_{1,t} = I(X_t; Z_t|Y_t, X^{t-1}, Z^{t-1}) \tag{34}$$

$$I_{2,t} = I(Y_t; Z_t|X_t, X^{t-1}, Z^{t-1}) \tag{35}$$

$$I_{3,t} = I(X_t, Y_t; Z_t|Z^{t-1}). \tag{36}$$

Note that (34), (35), (36) 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 $\left\{q_1(x_t|x^{t-1}, z^{t-1}) q_2(y_t|y^{t-1}, z^{t-1})\right\}_{t=1}^T$ inducing the sequence of measures $\{P_q(x^t, y_t, z^t)\}_{t=1}^T$, there exists a sequence of input distribution $\left\{q_1(x_t|x^{t-1}, z^{t-1})\hat{q}_2(y_t|x^{t-1}, z^{t-1})\right\}$ 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

$$\hat{P}(x^1, y_1, z^1) = W(z_1|x_1, y_1)q_1(x_1)\hat{q}_2(y_1) = W(z_1|x_1, y_1)q_1(x_1)q_2(y_1) = P(x^1, y_1, z^1) \tag{37}$$

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

$$\hat{P}(x^{t+1}, y_{t+1}, z^{t+1}) = W(z_{t+1}|x_{t+1}, y_{t+1})q_1(x_{t+1}|x^t, z^t)\hat{q}_2(y_{t+1}|x^t, z^t)\sum_{y_t}\hat{P}(x^t, y_t, z^t) \tag{38}$$

$$= W(z_{t+1}|x_{t+1}, y_{t+1})q_1(x_{t+1}|x^t, z^t)P_q(y_{t+1}|x^t, z^t)\sum_{y_t}P_q(x^t, y_t, z^t) \tag{39}$$

$$= P_q(x^{t+1}, y_{t+1}, z^{t+1}) \tag{40}$$

where (39) 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 of [18] 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

$$\theta_t(x^t) = P(x^t|z^t, u^t) = \frac{P(z_t, x^t|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} \tag{41}$$

$$= \frac{P(z_t, x_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} \tag{42}$$

$$= \frac{\sum_{y_t} P(z_t, x_t, y_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} \tag{43}$$

$$= \frac{\sum_{y_t} W(z_t|x_t, y_t)P(x_t|x^{t-1}, z^{t-1}, u^t)P(y_t|x^{t-1}, z^{t-1}, u^t)P(x^{t-1}|z^{t-1}, u^t)}{P(z_t|z^{t-1}, u^t)} \tag{44}$$

$$= \frac{\sum_{y_t} W(z_t|x_t, y_t)u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1})}{\sum_{\tilde{x}_t, \tilde{y}_t, \tilde{x}^{t-1}} W(z_t|\tilde{x}_t, \tilde{y}_t)u_t^1(\tilde{x}_t|\tilde{x}^{t-1})u_t^2(\tilde{y}_t|\tilde{x}^{t-1})\theta_{t-1}(\tilde{x}^{t-1})} \tag{45}$$

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

$$P(\theta_t|\theta^{t-1},u^{t,1},u^{t,2}) \tag{46}$$

$$= \sum_{x^t,y_t,z^t} P(\theta_t|\theta^{t-1},u^{t,1},u^{t,2},z^t,x^t,y_t)P(z_t|\theta^{t-1},u^{t,1},u^{t,2},z^{t-1},x^t,y_t) \tag{47}$$

$$P(x_t,y_t|\theta^{t-1},u^{t,1},u^{t,2},z^{t-1},x^{t-1})P(x^{t-1},z^{t-1}|\theta^{t-1},u^{t,1},u^{t,2})$$

$$= \sum_{z_t} \delta_{\Phi(\theta_{t-1},u_t^1,u_t^2,z_t)}(\theta_t) \sum_{x_t,y_t} W(z_t|x_t,y_t) \sum_{x^{t-1}} u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1}) \tag{48}$$

$$= P(\theta_t|\theta_{t-1},u_t^1,u_t^2) \tag{49}$$

C. Proof of Proposition 1

Define $\Theta_t \in \mathcal{P}(\mathcal{X}^t)$ as an information state. By Lemma 2, θ_t is recursively updated as a function of u_t^1, u_t^2 and θ_{t-1} and therefore the process $(\Theta_t)_t$ is a controlled Markov chain with an action (u_t^1, u_t^2).

Let $r_t(\underline{\lambda})$ be the instantaneous reward at time t which is defined as follows:

$$r_t(\underline{\lambda}) = \lambda_1 \log\frac{P(z_t|x_t,y_t,x^{t-1},z^{t-1})}{P(z_t|y_t,x^{t-1},z^{t-1})} + \lambda_2 \log\frac{P(z_t|x_t,y_t,x^{t-1},z^{t-1})}{P(z_t|x_t,x^{t-1},z^{t-1})} + \lambda_3 \log\frac{P(z_t|x_t,y_t,z^{t-1})}{P(z_t|z^{t-1})} \tag{50}$$

$$= \lambda_1 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{x}_t} W(z_t|\tilde{x}_t,y_t)u_t^1(\tilde{x}_t|x^{t-1})} + \lambda_2 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{y}_t} W(z_t|x_t,\tilde{y}_t)u_t^2(\tilde{y}_t|x^{t-1})}$$

$$+ \lambda_3 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{x}_t,\tilde{y}_t} W(z_t|\tilde{x}_t,\tilde{y}_t)\sum_{\tilde{x}^{t-1}} u_t^1(\tilde{x}_t|\tilde{x}^{t-1})u_t^2(\tilde{y}_t|\tilde{x}^{t-1})\theta_{t-1}(\tilde{x}^{t-1})} \tag{51}$$

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

$$E\left[R_t(\underline{\lambda})|\theta^{t-1},u^{t,1},u^{t,2}\right] \tag{52}$$

$$= \sum_{x_t,y_t,z_t} W(z_t|x_t,y_t) \sum_{x^{t-1}} u_t^1(x_t|x^{t-1})u_t^2(y_t|x^{t-1})\theta_{t-1}(x^{t-1})$$

$$\times\Bigg[\lambda_1 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{x}_t} W(z_t|\tilde{x}_t,y_t)u_t^1(\tilde{x}_t|x^{t-1})} + \lambda_2 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{y}_t} W(z_t|x_t,\tilde{y}_t)u_t^2(\tilde{y}_t|x^{t-1})}$$

$$+\lambda_3 \log\frac{W(z_t|x_t,y_t)}{\sum_{\tilde{x}_t,\tilde{y}_t} W(z_t|\tilde{x}_t,\tilde{y}_t)\sum_{\tilde{x}^{t-1}} u_t^1(\tilde{x}_t|\tilde{x}^{t-1})u_t^2(\tilde{y}_t|\tilde{x}^{t-1})\theta_{t-1}(\tilde{x}^{t-1})}\Bigg] \tag{53}$$

$$\triangleq \overline{r}_t^{\lambda}(\theta_{t-1},u_t^1,u_t^2), \tag{54}$$

which is a function of the information state θ_{t-1} and the action $u_t = (u_t^1, u_t^2)$ only. By standard MDP results, the optimal action is a Markov policy such that at time t it can be determined by looking at the information state θ_{t-1} only, i.e., the optimizing distribution can be a form of $\overline{q}_1(x_t|x^{t-1},\theta_{t-1})\overline{q}_2(y_t|y^{t-1},\theta_{t-1})$.

REFERENCES

[1] C. Shannon. The zero error capacity of a noisy channel. *Information Theory, IRE Transactions on*, 2(3):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.). 21(1):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*, 27(3):292–298, May 1981.

[4] F. Willems. The feedback capacity region of a class of discrete memoryless multiple access channels. *IEEE Trans. Information Theory*, 28(1):93–95, Jan. 1982.

[5] L. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Information Theory*, 30(4):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*, 51(3):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] A. P. Hekstra and F. M. J. Willems. Dependence balance bounds for single-output two-way channels. *IEEE Trans. Information Theory*, 35:287–290, Jan 1989.

[9] R. Tandon and S. Ulukus. Outer bounds for multiple access channels with feedback using dependence balance. *IEEE Trans. Information Theory*, 55(10):4494–4507, Oct. 2009.

[10] G. Kramer. *Directed Information for Channels with Feedback*. PhD thesis, ETH Series in Information Processing. Konstanz, Switzerland: Hartung-Gorre Verlag, 1998.

[11] J. Chen and T. Berger. The capacity of finite-state Markov channels with feedback. *Information Theory, IEEE Transactions on*, 51(3):780–798, March 2005.

[12] S. Tatikonda and S. Mitter. The capacity of channels with feedback. *IEEE Trans. Information Theory*, 55(1):323–349, Jan. 2009.

[13] G. Como, S. Yuksel, and S. Tatikonda. The error exponent of variable-length codes over markov channels with feedback. 55(5):2139–2160, May 2009.

[14] 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. IEEE International Symposium on Information Theory ISIT 2009*, pages 2346–2350, June 28 2009–July 3 2009.

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

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

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

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

[19] O. Shayevitz and M. Feder. The posterior matching feedback scheme: Capacity achieving and error analysis. In *Proc. International Symposium on Information Theory*, pages 900–904, 6–11 July 2008.

[20] O. Shayevitz and M. Feder. Optimal feedback communication via posterior matching. *IEEE Trans. Information Theory*, Mar 2011. (to appear).