Capacity Region of Finite State Multiple-Access Channels With Delayed State Information at the Transmitters

Uria Basher, Avihay Shirazi, and Haim H. Permuter, *Member, IEEE*

***Abstract*—A single-letter characterization is provided for the capacity region of finite-state multiple access channels. The channel state is a Markov process, the transmitters have access to delayed state information, and channel state information is available at the receiver. The delays of the channel state information are assumed to be asymmetric at the transmitters. We apply the result to obtain the capacity region for a finite-state Gaussian MAC, and for a finite-state multiple-access fading channel. We derive power control strategies that maximize the capacity region for these channels.**

***Index Terms*—Capacity region, delayed feedback, directed information, finite-state channel, Gaussian multiple-access channel, multiple-access channel, multiplexing coding scheme, successive decoding.**

I. Introduction

WIRELESS communication is an example of channels where the channel characteristics are time-varying. In a wireless setting, the user's motion and the changes in the environment, as well as the interference, may lead to temporal changes in the channel quality. Such channel variation models can include fast fading due to multi-path and slow fading due to shadowing. In fast fading, the channel state is assumed to be changing for every channel use, while in slow fading, the channel is assumed to be constant for each finite block length.

In such communication problems, the channel state information (CSI) can be transmitted to the transmitters either explicitly, or through output CSI feedback. Frequently, the CSI feedback is not instantaneous; the transmitters have only delayed information regarding the state of the channel. The availability of the delayed CSI at the transmitters will possibly increase the capacity region. The increase in the capacity region due to CSI depends on the CSI delays relative to the rate at which the channel is time-varying. When a channel is slowly time-varying and the delays are small, CSI may significantly increase the capacity region. However, if the channel is changing rapidly relative to the CSI delays, the transmitters can no longer adapt to the channel variations. Hence, availability of delayed CSI may not result in any significant capacity region improvement. Therefore, we are motivated to study the effect of channel memory and delays on the multiple access channel (MAC) capacity region.

Let us now present a brief literature review. We are modeling a time-varying channel as a finite-state Markov channel (FSMC) [1], [2]. The FSMC is a channel with a finite number of states. During each symbol transmission, the channel's state is fixed. The channel transition probability function is determined by the channel state. The time variation in the channel characteristics is modeled by the statistics of the underlying state process.

Capacity of memoryless channels, with different cases of state information being available in a causal or non causal manner at the transmitter and at the receiver, has been studied by Shannon [3] and by Gelfand and Pinsker [4]. In [5], Goldsmith and Varaiya consider the fading channels with perfect CSI at the transmitter and at the receiver. They proved that with instantaneous and perfect state information, the transmitter can adapt the data rates for each channel state to maximize the average transmission rate. Viswanathan [6] loosened this assumption of perfect instantaneous CSI, and gave a single letter characterization of the capacity of Markov channels with delayed CSI. Caire and Shamai [7] consider the case that the channel state is independent identically distributed (i.i.d.), and the CSI at the transmitter is a deterministic function of the CSI at the receiver. They showed that optimal coding is particularly simple. Chen and Berger in [8] found the capacity of an FSC with inter-symbol interference (ISI), where current CSI is available at the transmitter and the receiver. For a comprehensive survey on channel coding with state information see [9].

The MAC with state has received much attention in recent years due to its importance in wireless communication systems. On the one hand, complete knowledge of the CSI at the transmitters is an unrealistic assumption in wireless communications. On the other hand, it is reasonable to assume that the receiver does possess full knowledge of the CSI. This practical consideration has motivated the investigation of a MAC where each transmitter is informed with its own CSI, while the receiver is informed with the full CSI.

Our work is also related to [10], [11], and [12]. In [10] the authors found the capacity region of FS-MAC, where the channel state process is i.i.d., the transmitters have access to partial (quantized) CSI, and complete CSI is available at the receiver. In [11] the capacity of general FS-MAC with varying degrees of causal CSI at the transmitters is characterized in non-single-letter formulas. In [12] the capacity region of the

Manuscript received January 11, 2011; revised July 24, 2011; accepted August 28, 2011. Date of publication January 18, 2012; date of current version May 15, 2012. This work was supported in part by the Marie Curie Reintegration Fellowship within the European Community Framework Program and by the Ministry of Defense, Israel.

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel (e-mail: basher@bgu.ac.il; avihays@bgu.ac.il; haimp@bgu.ac.il).

Communicated by T. Uyematsu, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2172726

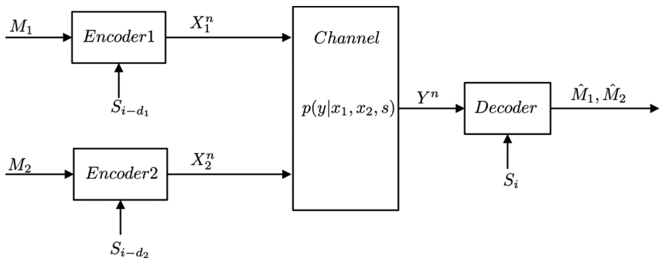

Fig. 1. FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2 - The state process has memory and is assumed to be FSM. The CSI is fed back to the encoders through a noiseless feedback channel. CSI from the decoder is received at Encoder 1 and Encoder 2 after time delays of d_1 and d_2 symbol durations, respectively. We are considering the above problem setting in the cases where $d_1 > d_2$, $d_1 = d_2$, and $d_2 < d_1 = \infty$.

FS-MAC with feedback that may be an arbitrary time-invariant function of the channel output has been derived. Recent related works also include [13], [14], and [15]. In [13], the state-dependent MAC with causal and strictly causal side information at the transmitters has been studied. In [14], [15] the authors considered a MISO broadcast channels with delayed feedback. They established the optimal sum-degrees of freedom (DoF), which shows that even when the state process is i.i.d., or in the presence of arbitrary large delay, the CSI can still significantly increase the DoF.

In this work, we consider the capacity region of a finite state Markov Multiple-access channel (FSM-MAC) with CSI at the decoder (receiver) and delayed CSI at the encoders (transmitters) with delays d_1 and d_2 as illustrated in Fig. 1. The channel probability function at each time instant depends on the state of an underlying finite-state Markov process. The decoder, in addition to the channel output, also receives the channel state at each time instant (perfect CSI). The channel state is fed back to the encoders through a noiseless feedback channel. CSI from the decoder is received at Encoder 1 and Encoder 2 after time delays of d_1 and d_2 symbol durations, respectively. The time delays d_1 and d_2 are assumed to be known at both Encoder 1 and Encoder 2. Each encoder, at each time instant, chooses the channel input based on the message to be transmitted and the CSI that he possesses. A formal description of the system model is presented in Section II. The main result of this paper is a computable characterization of the capacity region for this channel model.

The remainder of the paper is organized as follows: In Section II, we concretely describe the communication model. In Section III, we state our main results, which are the capacity regions for different cases of time delays. Section IV provides the outer bound on the capacity region of FSM-MAC with CSI at the decoder and asymmetrical delayed CSI at the encoders. In Section V, we complete the proof of the capacity region, by providing the proof of the achievability. In Section VI, we provide alternative proof for capacity region. The alternative proof is based on a multi-letter expression for the capacity region of FS-MAC with time-invariant feedback [12]. In Section VII, we apply the general results of Section III to obtain the capacity region for a finite-state Gaussian MAC, and for a finite-state multiple-access fading channel. We derive optimization problems on the power allocation that maximize the capacity region for these channels. This power allocation would be the optimal power control policy for maximizing throughput in the presence of delayed CSI. We conclude in Section VIII with a summary of this work.

II. Channel Model and Notation

A. Channel Model

In this paper, we consider the communication system of FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2, respectively, as illustrated in Fig. 1. The MAC setting consists of two senders and one receiver. Each sender $j \in \{1,2\}$ chooses an index m_j uniformly from the set $\{1,\ldots,2^{nR_j}\}$ and independent of the other sender. The input to the channel from encoder $j \in \{1,2\}$ is denoted by $\{X_{j,1}, X_{j,2}, X_{j,3}, \ldots\}$, and the output of the channel is denoted by $\{Y_1, Y_2, Y_3, \ldots\}$. We use the notation V^n to denote the sequence $(V_1,\ldots,V_n)$, therefore, X_j^n, Y^n denote the sequences $\{X_{j,1},\ldots,X_{j,n}\}$, $\{Y_1,\ldots,Y_n\}$, respectively. A finite-state Markov channel is, at each time instant, in one of a finite number of states $\mathcal{S} = \{s_1, s_2, \ldots, s_k\}$. In each state, the channel is a DMC with inputs alphabet $\mathcal{X}_1$, $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. Let the random variables S_i, S_{i-d} denote the channel state at times i and $i-d$, respectively. Similarly, denote by $X_{1,i}$, $X_{2,i}$, and Y_i the inputs and the output of the channel at time i. The channel transition probability function at time i depends on the state S_i, and the inputs $X_{1,i}$, $X_{2,i}$ at time i, and is given by $P(y_i|x_{1,i},x_{2,i},s_i)$. The channel output at any time i is assumed to depend only on the channel inputs and state at time i. Hence

$$P(y_i|x_1^i, x_2^i, s^i) = P(y_i|x_{1,i}, x_{2,i}, s_i). \tag{1}$$

The state process $\{S_i\}$ is assumed to be an irreducible, aperiodic, finite-state homogeneous Markov chain and hence is ergodic. The state process is independent of the channel inputs and output when conditioned on the previous states, i.e.,

$$P(s_i|s^{i-1}, x_1^{i-1}, x_2^{i-1}, y^{i-1}) = P(s_i|s_{i-1}). \tag{2}$$

Furthermore, we assume that the state process is independent of M_1 and M_2,

$$\begin{aligned} P(s^n, m_1, m_2) &= P(s^n)P(m_1)P(m_2) \\ &= \prod_{i=1}^{n} P(s_i|s_{i-1})P(m_1)P(m_2). \end{aligned} \tag{3}$$

Now, let K be the one step state transition probability matrix of the Markov process, and let π be the steady state probability distribution of the Markov process. The (S_i, S_{i-d}) joint distribution is stationary and is given by

$$\pi_d(S_i = s_l, S_{i-d} = s_j) = \pi(s_j)K^d(s_l, s_j), \tag{4}$$

where $K^d(s_l, s_j)$ is the $(l,j)\text{th}$ element of the d-step transition probability matrix K^d of the Markov state process. Without loss of generality, let us assume that $d_1 \geq d_2$. Furthermore, for simplicity, let us define the joint distribution of the variables $(S, \tilde{S}_1, \tilde{S}_2)$ as the joint distribution of the variables $(S_i, S_{i-d_1}, S_{i-d_2})$, i.e.,

$$\begin{aligned} &P(S = s_l, \tilde{S}_1 = s_j, \tilde{S}_2 = s_v) \\ &= P(S_i = s_l, S_{i-d_1} = s_j, S_{i-d_2} = s_v) \\ &= \pi(s_j) K^{d_1-d_2}(s_v, s_j) K^{d_2}(s_l, s_v), \end{aligned} \tag{5}$$

where $s_j, s_l, s_v \in \mathcal{S}$.

B. Code Description

An $(n, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code for FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delay d_1 and d_2 consists of

1) Two sets of integers $\mathcal{M}_1 = \{1, 2, \ldots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, \ldots, 2^{nR_2}\}$, called the *message sets*.
2) For each encoder, an encoding function $f_j,\ j \in \{1, 2\}$, maps the set of messages to channel input words of block length n. Each f_j works through a sequence of functions $f_{j,i}$ that depend only on the message M_j and the channel states up to time $i - d_j$. For encoder 1 $(j = 1)$:

$$X_{1,i} = \left\{ \begin{array}{cc} f_{1,i}(M_1), & 1 \le i \le d_1 \\ f_{1,i}(M_1, S^{i-d_1}), & d_1 + 1 \le i \le n \end{array} \right\}. \tag{6}$$

Similarly for encoder 2 $(j = 2)$:

$$X_{2,i} = \left\{ \begin{array}{cc} f_{2,i}(M_2), & 1 \le i \le d_2 \\ f_{2,i}(M_2, S^{i-d_2}), & d_2 + 1 \le i \le n \end{array} \right\} \tag{7}$$

3) A decoding function ψ that maps a received sequence of n channel outputs and channel states to the messages set

$$\psi : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{M}_1 \times \mathcal{M}_2. \tag{8}$$

We define the average probability of error for the $(n, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code as follows:

$$\begin{aligned} P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{m_1,m_2} \sum_{s_1^n} P_{S^n}(s^n) \Pr\{\psi(y^n, s^n) \\ \neq (m_1, m_2) | (m_1, m_2) \text{was sent}\}. \end{aligned} \tag{9}$$

We use standard definitions [16] of achievability and capacity region, namely, a pair rate (R_1, R_2) is *achievable* for FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2, if there exists a sequence of $(n, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes with $P_e^{(n)} \rightarrow 0$ as n goes to infinity. The *capacity region* is the closure of the set of achievable (R_1, R_2) rate pairs.

III. Main Results

Here we present the main results of this paper. Recall, that the joint distributions of $(S, \tilde{S}_1)$, and $(S, \tilde{S}_2)$ is given in (5). Without loss of generality, let us assume that $d_1 \ge d_2$.

Theorem 1 (Capacity Region of FSM-MAC With Delayed CSI $d_1 \ge d_2$): The capacity region of FSM-MAC with CSI at the decoder and asymmetrical delayed CSI at the encoders with delays d_1 and d_2 as shown in Fig. 1 is given in (10) at the bottom of the page. where U is an auxiliary random variable with cardinality $|\mathcal{U}| \le 3$.

The proof of Theorem 1 is presented in Sections IV, and V. In Section IV we prove the outer bound of the capacity region, and Section V is devoted to the proof of the achievability. The proof of the achievability is based on a multiplexing coding scheme, and successive decoding. In addition, we provide alternative proof of Theorem 1 in Section VI. The proof for the cardinality bound of U is presented in Appendix A.

Now, directly from Theorem 1 we can derive the capacity region in the case of $d_1 = d_2$. Since $d_1 = d_2$ we have $\tilde{S}_1 = \tilde{S}_2$, hence we denote $\tilde{S} = \tilde{S}_1 = \tilde{S}_2$. Using Theorem 1 we get,

Theorem 2 (Capacity Region of FSM-MAC With Symmetrical Delayed CSI $d_1 = d_2$): The capacity region of FSM-MAC with CSI at the decoder and symmetrical delayed CSI at the encoders with delay d is given in (11) at the bottom of the page. where U is an auxiliary random variable with cardinality $|\mathcal{U}| \le 3$.

Now we consider the case that encoder 1 does not have state information at all, i.e., $d_1 = \infty$.

Theorem 3 (Capacity Region of FSM-MAC With Delayed CSI Only to one Encoder): The capacity region of FSM-MAC with CSI at the decoder and delayed CSI only to one encoder is given by (12) at the bottom of the next page. Where Q is an auxiliary random variable with cardinality $|\mathcal{Q}| \le 3$.

The proof of Theorem 3 is quite similar to the proof of Theorem 1; the details are presented in Appendix B. In Appendix C, we present the capacity region for the case where there are three encoders. In addition, a sketch of the proof is provided.

$$\mathcal{R} = \bigcup_{P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)} \left(\begin{array}{r} R_1 < I(X_1; Y | X_2, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_2 < I(X_2; Y | X_1, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_1 + R_2 < I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2, U), \end{array} \right) \tag{10}$$

$$\mathcal{R} = \bigcup_{P(u|\tilde{s})P(x_1|\tilde{s},u)P(x_2|\tilde{s},u)} \left(\begin{array}{r} R_1 < I(X_1; Y | X_2, S, \tilde{S}, U), \\ R_2 < I(X_2; Y | X_1, S, \tilde{S}, U), \\ R_1 + R_2 < I(X_1, X_2; Y | S, \tilde{S}, U), \end{array} \right) \tag{11}$$

IV. CONVERSE

In this section, we provide the outer bound on the capacity region of MAC with receiver CSI and asymmetrical delayed CSI feedback, i.e., we give the converse proof for Theorem 1. Without loss of generality let us assume that $d_1 \geq d_2$.

Proof: Given an achievable rate (R_1, R_2) we need to show that there exists joint distribution of the form $P(s, \tilde{s}_1, \tilde{s}_2)P(u|\tilde{s}_1)P(x_1|\tilde{s}_1, u)P(x_2|\tilde{s}_1, \tilde{s}_2, u)P(y|x_1, x_2, s)$ such that,

$$\begin{aligned} R_1 &< I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_2 &< I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_1 + R_2 &< I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U) \end{aligned}$$

where U is an auxiliary random variable with cardinality $|\mathcal{U}| \leq 3$. The proof for the cardinality bound is presented in Appendix A. Since (R_1, R_2) is an achievable pair-rate, there exists an $(n, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code with a probability of error $P_e^{(n)}$ arbitrarily small. By Fano's inequality

$$\begin{aligned} H(M_1, M_2|Y^n, S^n) &\leq n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}) \\ &\triangleq n\varepsilon_n \end{aligned} \tag{13}$$

and it is clear that $\varepsilon_n \rightarrow 0$ as $P_e^{(n)} \rightarrow \infty$. Then we have

$$H(M_1|Y^n, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\varepsilon_n, \tag{14}$$
$$H(M_2|Y^n, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\varepsilon_n. \tag{15}$$

We can now bound the rate R_1 as

$$\begin{aligned} nR_1 &= H(M_1) \\ &= H(M_1) + H(M_1|Y^n, S^n) - H(M_1|Y^n, S^n) \\ &\overset{(a)}{\leq} I(M_1; Y^n, S^n) + n\varepsilon_n \\ &\overset{(b)}{=} I(M_1; Y^n|S^n) + I(M_1; S^n) + n\varepsilon_n \\ &\overset{(c)}{=} I(M_1; Y^n|S^n) + n\varepsilon_n \\ &\overset{(d)}{=} I(X_1^n; Y^n|S^n) + n\varepsilon_n \\ &= H(X_1^n|S^n) - H(X_1^n|Y^n, S^n) + n\varepsilon_n \\ &\overset{(e)}{=} H(X_1^n|X_2^n, S^n) - H(X_1^n|Y^n, S^n) + n\varepsilon_n \\ &\overset{(f)}{\leq} H(X_1^n|X_2^n, S^n) - H(X_1^n|Y^n, X_2^n, S^n) + n\varepsilon_n \\ &= I(X_1^n; Y^n|X_2^n, S^n) + n\varepsilon_n \\ &= H(Y^n|X_2^n, S^n) - H(Y^n|X_1^n, X_2^n, S^n) + n\varepsilon_n \\ &= \sum_{i=1}^{n} \{H(Y_i|Y^{i-1}, X_2^n, S^n) \\ &\quad - H(Y_i|Y^{i-1}, X_1^n, X_2^n, S^n)\} + n\varepsilon_n \\ &\overset{(g)}{\leq} \sum_{i=1}^{n} \{H(Y_i|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1}) \\ &\quad - H(Y_i|Y^{i-1}, X_1^n, X_2^n, S^n)\} + n\varepsilon_n \\ &\overset{(h)}{=} \sum_{i=1}^{n} \{H(Y_i|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1}) \\ &\quad - H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S_1^{i-d_1-1})\} \\ &\quad + n\varepsilon_n \\ &= \sum_{i=1}^{n} \left\{I(Y_i; X_{1,i}|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S_1^{i-d_1-1})\right\} \\ &\quad + n\varepsilon_n \end{aligned}$$

where

a) follows from Fano's inequality;
b) follows from chain rule;
c) follows from the fact that M_1 and S^n are independent;
d) follows from the fact that X_1^n is a deterministic function of (M_1, S^n) and the Markov chain $(M_1, S^n) - (X_1^n, S^n) - Y^n$;
e) follows from the fact that X_1^n and M_2 are independent, and the fact that X_2^n is a deterministic function of (M_2, S^n). Therefore, X_1^n and X_2^n are independent given S^n;
f) (f) and (g) follow from the fact that conditioning reduces entropy;
g) follows from the fact that the channel output at time i depends only on the state S_i and the the inputs $X_{1,i}$ and $X_{2,i}$.

Hence, we have

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(Y_i; X_{1,i}|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S_1^{i-d_1-1}) + \varepsilon_n. \tag{16}$$

Similarly, we have

$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(Y_i; X_{2,i}|X_{1,i}, S_i, S_{i-d_2}, S_{i-d_1}, S_1^{i-d_1-1}) + \varepsilon_n. \tag{17}$$

To bound the sum of the rates, consider

$$\begin{aligned} &n(R_1 + R_2) \\ &= H(M_1, M_2) \\ &= H(M_1, M_2) + H(M_1, M_2|Y^n, S^n) \\ &\quad - H(M_1, M_2|Y^n, S^n) \\ &\overset{(a)}{\leq} I(M_1, M_2; Y^n, S^n) + n\varepsilon_n \\ &\overset{(b)}{=} I(M_1, M_2; Y^n|S^n) + I(M_1, M_2; S^n) \\ &\quad + n\varepsilon_n \\ &\overset{(c)}{=} I(M_1, M_2; Y^n|S^n) + n\varepsilon_n \\ &\overset{(d)}{=} I(X_1^n, X_2^n; Y^n|S^n) + n\varepsilon_n \\ &= H(Y^n|S^n) - H(Y^n|X_1^n, X_2^n, S^n) + n\varepsilon_n \end{aligned}$$

$$\mathcal{R} = \bigcup_{P(q)P(x_1|q)P(x_2|\tilde{s},q)} \begin{pmatrix} R_1 < I(X_1; Y|X_2, S, \tilde{S}, Q), \\ R_2 < I(X_2; Y|X_1, S, \tilde{S}, Q), \\ R_1 + R_2 < I(X_1, X_2; Y|S, \tilde{S}, Q), \end{pmatrix} \tag{12}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \{H(Y_i|Y^{i-1}, S^n) \\
&\quad - H(Y_i|Y^{i-1}, X_1^n, X_2^n, S^n)\} + n\varepsilon_n \\
&\overset{(e)}{=} \sum_{i=1}^{n} \{H(Y_i|Y^{i-1}, S^n) \\
&\quad - H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1})\} \\
&\quad + n\varepsilon_n \\
&\overset{(f)}{\le} \sum_{i=1}^{n} \{H(Y_i|S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1}) \\
&\quad - H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1})\} \\
&\quad + n\varepsilon_n \\
&= \sum_{i=1}^{n} \{I(Y_i; X_{1,i}, X_{2,i}|S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1})\} \\
&\quad + n\varepsilon_n
\end{aligned}
$$

where

a) follows from Fano's inequality;
b) follows from chain rule;
c) follows from the fact that M_1, M_2, and S^n are independent;
d) follows from the fact that X_1^n, X_2^n is a deterministic function of (M_1, M_2, S^n) and the Markov chain $(M_1, M_2, S^n) - (X_1^n, X_2^n, S^n) - Y^n$;
e) follows from the fact that the channel output at time i depends only on the state S_i, and the inputs $X_{1,i}$, and $X_{2,i}$;
f) follows from the fact that conditioning reduces entropy.

Hence, we have

$$
\begin{aligned}
R_1 + R_2 \le \frac{1}{n} \sum_{i=1}^{n} I(Y_i; X_{1,i}, X_{2,i}|S_i,& \\
S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1}) + \varepsilon_n.& \qquad (18)
\end{aligned}
$$

The expressions in (16)–(18) are the average of the mutual informations calculated at the empirical distribution in column i of the codebook. We can rewrite these equations with the new variable Q, where $Q = i \in \{1, 2, \ldots, n\}$ with probability $\frac{1}{n}$. The equations become

$$
\begin{aligned}
R_1 &\le \frac{1}{n} \sum_{i=1}^{n} I(Y_i; X_{1,i}|X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, S^{i-d_1-1}) \\
&\quad + \varepsilon_n \\
&= \frac{1}{n} \sum_{i=1}^{n} I(Y_Q; X_{1,Q}|X_{2,Q}, S_Q, S_{Q-d_2}, \\
&\qquad\qquad S_{Q-d_1}, S^{Q-d_1-1}, Q = i) + \varepsilon_n \\
&= I(Y_Q; X_{1,Q}|X_{2,Q}, S_Q, S_{Q-d_2}, S_{Q-d_1}, \\
&\qquad\qquad S^{Q-d_1-1}, Q) + \varepsilon_n. \qquad (19)
\end{aligned}
$$

Now let us denote $X_1 \triangleq X_{1,Q}, X_2 \triangleq X_{2,Q}, Y \triangleq Y_Q, S \triangleq S_Q$, $\tilde{S}_1 \triangleq S_{Q-d_1}$, $\tilde{S}_2 \triangleq S_{Q-d_2}$, and $U \triangleq (S^{Q-d_1-1}, Q)$.

We have

$$
\begin{aligned}
R_1 &\le I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U) + \varepsilon_n, \\
R_2 &\le I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U) + \varepsilon_n, \\
R_1 + R_2 &\le I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U) + \varepsilon_n.
\end{aligned}
$$

To complete the converse proof we need to show the following Markov relations hold:

1) $P(u|s, \tilde{s}_1, \tilde{s}_2) = P(u|\tilde{s}_1)$;
2) $P(x_1|s, \tilde{s}_1, \tilde{s}_2, u) = P(x_1|\tilde{s}_1, u)$;
3) $P(x_2|x_1, s, \tilde{s}_1, \tilde{s}_2, u) = P(x_2|\tilde{s}_1, \tilde{s}_2, u)$;
4) $P(y|x_1, x_2, s, \tilde{s}_1, \tilde{s}_2, u) = P(y|x_1, x_2, s)$.

We prove the above using the following claims:

1) follows from the fact that $S^{i-d_1-1} - S_{i-d_1} - S_{i-d_2} - S_i$ and so is $(S_1^{Q-d_1-1}, Q) - S_{Q-d_1} - S_{Q-d_2} - S_Q$;
2) follows from the fact that $X_{1,i} = f_{1,i}(M_1, S^{i-d_1})$ and that M_1 and S^n are independent. Hence

$$
\begin{aligned}
P(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, s_1^{q-d_1-1}, q = i) \\
= P(x_{1,q}|s_{q-d_1}, s_1^{q-d_1-1}, q = i).
\end{aligned}
$$

Since this is true for all i,

$$
\begin{aligned}
P(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, s_1^{q-d_1-1}, q) \\
= P(x_{1,q}|s_{q-d_1}, s_1^{q-d_1-1}, q).
\end{aligned}
$$

Therefore we have

$$P(x_1|s, \tilde{s}_1, \tilde{s}_2, u) = P(x_1|\tilde{s}_1, u);$$

3) we assume that $d_1 \ge d_2$, since M_2 and (M_1, S^n) are independent, and the state process is Markov chain, we have

$$
\begin{aligned}
P(m_2, s^{i-d_2}|s_i, s_{i-d_1}, s_{i-d_2}, s^{i-d_1}, m_1) \\
= P(m_2, s^{i-d_2}|s_{i-d_1}, s_{i-d_2}, s^{i-d_1}).
\end{aligned}
$$

Therefore, we have the Markov chain $(M_2, S^{i-d_2}) - (S_{i-d_1}, S_{i-d_2}, S^{i-d_1}) - (M_1, S_i, S^{i-d_1})$. Since $X_{1,i} = f_{1,i}(M_1, S^{i-d_1})$ and $X_{2,i} = f_{2,i}(M_2, S^{i-d_2})$ where $f_{1,i}, f_{2,i}$ are deterministic functions, we obtain the following Markov chain:

$$
\begin{aligned}
X_{2,i} - (M_2, S^{i-d_2}) - (S_{i-d_1}, S_{i-d_2}, S^{i-d_1}) \\
-(M_1, S_i, S^{i-d_1}) - X_{1,i} \qquad (20)
\end{aligned}
$$

which implies

$$
\begin{aligned}
P(x_{2,i}|x_{1,i}, s_i, s_{i-d_1}, s_{i-d_2}, s^{i-d_1-1}) \\
= P(x_{2,i}|s_{i-d_1}, s_{i-d_2}, s^{i-d_1-1}).
\end{aligned}
$$

Since this is true for all i:

$$
\begin{aligned}
P(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, s_1^{q-d_1-1}, q) \\
= P(x_{2,q}|s_{q-d_1}, s_{q-d_2}, s_1^{q-d_1-1}, q).
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
P(x_2|x_1, s, \tilde{s}_1, \tilde{s}_2, u) \\
= P(x_2|\tilde{s}_1, \tilde{s}_2, u); \qquad (21)
\end{aligned}
$$

4) follows from the fact that the channel output at any time i is assumed to depend only on the channel inputs and state at time i.

Hence, taking the limit as $n \to \infty$, $P_e^{(n)} \to 0$, we have the following converse:

$$
\begin{aligned}
R_1 &\le I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U), \\
R_2 &\le I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U), \\
R_1 + R_2 &\le I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U)
\end{aligned}
$$

for some choice of joint distribution $P(s,\tilde{s}_1,\tilde{s}_2)P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)P(y|x_1,x_2,s)$ and for some choice of auxiliary random variable U defined on $|\mathcal{U}| \leq 3$. This completes the proof of the converse. ■

V. Proof of the Achievability of Theorem 1

In the previous section we proved the converse of the capacity region of Theorem 1. In this section we prove the achievability part. The main idea of the proof is using multiplexing coding (e.g., [5] and [6]), i.e., multiplexing the input of the channel at each encoder (the multiplexer is controlled by the delayed CSI), then, using the CSI known at the decoder, demultiplexing the output at the decoder.

Proof: For simplicity, we present first the proof without the auxiliary random variable U. Then, we complete the achievability proof of Theorem 1 by using time sharing, where the time sharing is a function of the delayed CSI. To prove the achievability of the capacity region, we need to show that for a fix $P(x_1|\tilde{s}_1)P(x_2|\tilde{s}_1,\tilde{s}_2)$ and (R_1,R_2) that satisfy

$$R_1 \leq I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2),$$
$$R_2 \leq I(X_2;Y|X_1,S,\tilde{S}_1,\tilde{S}_2),$$
$$R_1+R_2 \leq I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2)$$

there exists a sequence of $(n,2^{nR_1},2^{nR_2},d_1,d_2)$ codes where $P_e^{(n)} \longrightarrow 0$ as $n \longrightarrow \infty$. Without loss of generality, we assume that the finite-state space $\mathcal{S}=\{1,2,\ldots,k\}$, and that the steady state probability $\pi(l)>0$ for all $l \in \mathcal{S}$.

Encoder 1: Construct k codebooks $\mathcal{C}_1^{\tilde{s}_1}$ (where the subscript is for Encoder 1) for all $\tilde{S}_1 \in \mathcal{S}$, when in each codebook $\mathcal{C}_1^{\tilde{s}_1}$ there are $2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}$ codewords, where $n_1(\tilde{s}_1)=(P(\tilde{S}_1=\tilde{s}_1)-\epsilon')n$, for $\epsilon'>0$. Every codeword $\mathcal{C}_1^{\tilde{s}_1}(i)$ when $i \in \{1,2,\ldots,2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}\}$ has a length of $n_1(\tilde{s}_1)$ symbols. Each codeword from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook is built $X_1^{\tilde{s}_1} \sim$ i.i.d. $P(x_1^{\tilde{s}_1}|\tilde{S}_1=\tilde{s}_1)$ (where the subscript is for Encoder 1). A message M_1 is chosen according to a uniform distribution $\Pr(M_1=m_1)=2^{-nR_1}$, $m_1 \in \{1,2,\ldots,2^{nR_1}\}$. Every message m_1 is mapped into k sub messages $\mathcal{V}_1(m_1)=\{V_1^1(m_1),V_1^2(m_1),\ldots,V_1^k(m_1)\}$ (one message from each codecook). Hence, every message m_1 is specified by a k dimensional vector. For a fix block length n, let $N_{\tilde{s}_1}$ be the number of times during the n symbols for which the feedback information at encoder 1 regarding the channel state is $\tilde{S}_1=\tilde{s}_1$. Every time that the delayed CSI is $\tilde{S}_1=\tilde{s}_1$, encoder 1 sends the next symbol from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook. Since $N_{\tilde{s}_1}$ is not necessarily equivalent to $n_1(\tilde{s}_1)$, an error is declared if $N_{\tilde{s}_1}<n_1(\tilde{s}_1)$, and the code is zero-filled if $N_{\tilde{s}_1}>n_1(\tilde{s}_1)$. Therefore, we can send a total of $2^{nR_1}=2^{\sum_{\tilde{s}_1\in\mathcal{S}} n_1(\tilde{s}_1)R_1(\tilde{s}_1)}$ messages. The codebook construction of encoder 1 is illustrated in Fig. 2.

Encoder 2: Construct $k\times k$ codebooks $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ (where the subscript is for Encoder 2) for all $(\tilde{s}_1,\tilde{s}_2) \in \{\mathcal{S}\times\mathcal{S}\}$, when in each codebook $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ there are $2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}$ codewords, where $n_2(\tilde{s}_1,\tilde{s}_2)=(P(\tilde{S}_1,\tilde{S}_2=\tilde{s}_1,\tilde{s}_2)-\epsilon')n$, for $\epsilon'>0$. Every codeword $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}(i)$ when $i \in \{1,2,\ldots,2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}\}$ has a length of $n_2(\tilde{s}_1,\tilde{s}_2)$ symbols. Each codeword from the $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ codebook is built $X_2^{\tilde{s}_1,\tilde{s}_2} \sim$ i.i.d. $P(x_2^{\tilde{s}_1,\tilde{s}_2}|(\tilde{S}_1,\tilde{S}_2)=$

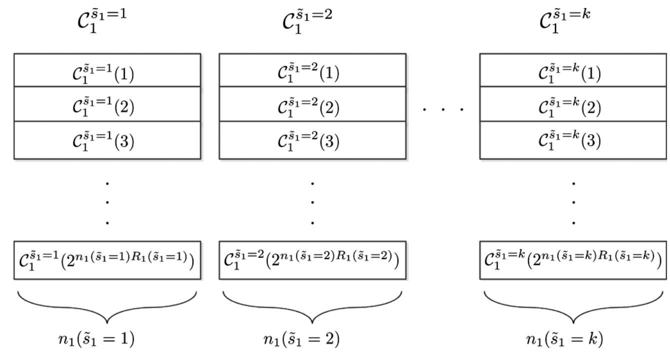

Fig. 2. Multiplexing coding: Encoder 1's codebook is assembled from k sub codebooks $\mathcal{C}_1^{\tilde{s}_1}$, when in each codebook $\mathcal{C}_1^{\tilde{s}_1}$ there are $2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}$ codewords. Each codeword from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook is built $X_1^{\tilde{s}_1} \sim$ i.i.d. $P(x_1^{\tilde{s}_1}|\tilde{S}_1=\tilde{s}_1)$. In a similar way, we use multiplexing coding to assembled the codebook of Encoder 2, where the multiplexer is controlled by both $\tilde{s}_1$ and $\tilde{s}_2$.

$(\tilde{s}_1,\tilde{s}_2))$ (where the subscript is for Encoder 2). A message M_2 is chosen according to a uniform distribution $\Pr(M_2=m_2)=2^{-nR_2}$, $m_2 \in \{1,2,\ldots,2^{nR_2}\}$. Every message m_2 is mapped into $k\times k$ sub messages $\mathcal{V}_2(m_2)=\left\{V_2^{1,1}(m_2),V_2^{1,2}(m_2),\ldots,V_2^{k,k}(m_2)\right\}$ (one message from each codecook). Hence, every message m_2 is specified by a $k\times k$ dimensional vector. For a fix block length n, let $N_{\tilde{s}_1,\tilde{s}_2}$ be the number of times during the n symbols for which the feedback information at encoder 2 regarding the channel state is $(\tilde{S}_1,\tilde{S}_2)=(\tilde{s}_1,\tilde{s}_2)$. Every time that the delayed CSI is $(\tilde{S}_1,\tilde{S}_2)=(\tilde{s}_1,\tilde{s}_2)$, encoder 2 sends the next symbol from the $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ codebook. Since $N_{\tilde{s}_1,\tilde{s}_2}$ is not necessarily equivalent to $n_2(\tilde{s}_1,\tilde{s}_2)$, an error is declared if $N_{\tilde{s}_1,\tilde{s}_2}<n_2(\tilde{s}_1,\tilde{s}_2)$, and the code is zero-filled if $N_{\tilde{s}_1,\tilde{s}_2}>n_2(\tilde{s}_1,\tilde{s}_2)$. Therefore, we can send a total of $2^{nR_2}=2^{\sum_{\tilde{s}_1,\tilde{s}_2\in\mathcal{S}\times\mathcal{S}} n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}$ messages.

Decoding: We use successive decoding; in this method, instead of decoding the two messages simultaneously, the decoder first decodes one of the messages by itself, where the other user's message is considered as noise. After decoding the first user's message, the decoder turns to decode the second message. When decoding the second message, the decoder uses the information about the first message as side information. This decoding rule aims to achieve the two corner points of the rate region, i.e., $(R_1=I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2)-\epsilon, R_2=I(X_2;Y|S,\tilde{S}_1,\tilde{S}_2)-\epsilon)$, and $(R_1=I(X_1;Y|S,\tilde{S}_1,\tilde{S}_2)-\epsilon, R_2=I(X_2;Y|X_1,S,\tilde{S}_1,\tilde{S}_2)-\epsilon)$. The rate region is illustrated in Fig. 3.

To achieve the first point, let us analyze the case where the decoder first decodes X_2^n. The information $\tilde{S}_1$, $\tilde{S}_2$ used to multiplex the codewords at the encoder is also available at the decoder. Hence, upon receiving a block of channel outputs and states (Y^n,S^n), the decoder first demultiplexes it into outputs corresponding to the component codebooks of encoder 2. Then, the decoder separately decodes each component codeword $V_2^{\tilde{s}_1,\tilde{s}_2}$ where $(\tilde{s}_1,\tilde{s}_2) \in \mathcal{S}\times\mathcal{S}$. For each codebook $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$, the decoder has $(Y^{n_2(\tilde{s}_1,\tilde{s}_2)},S^{n_2(\tilde{s}_1,\tilde{s}_2)})$ and searches $(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)})$ such that $(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)},Y^{n_2(\tilde{s}_1,\tilde{s}_2)},S^{n_2(\tilde{s}_1,\tilde{s}_2)})$ are strongly jointly typical sequences [16], i.e., $(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)},Y^{n_2(\tilde{s}_1,\tilde{s}_2)},S^{n_2(\tilde{s}_1,\tilde{s}_2)}) \in$

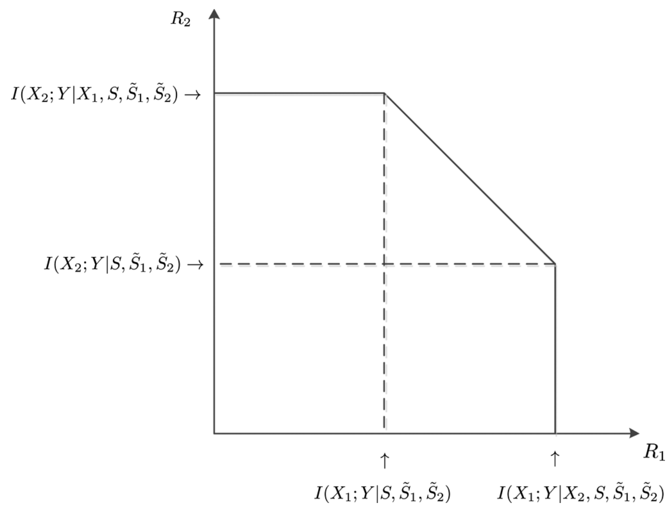

Fig. 3. Rate region.

$A_\varepsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)$ given $(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)$. The decoder declares that $\hat{m}_2$ is sent if it is a unique message such that $(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(\hat{m}_2), Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)}) \in A_\varepsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)$ given $(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)$ for all $\tilde{s}_1,\tilde{s}_2 \in \mathcal{S} \times \mathcal{S}$, otherwise it declares an error. If such $\hat{m}_2$ is found, the decoder has $X_2^n(\hat{m}_2)$, but now the decoder is using the information $\tilde{S}_1$ to demultiplex (Y^n, S^n) into outputs corresponding to the component codebooks of encoder 1 (which have k codebooks). The decoder declares that $\hat{m}_1$ is sent if it is a unique message such that $(X_1^{n_1(\tilde{s}_1)}(\hat{m}_1), X_2^{n_1(\tilde{s}_1)}(\hat{m}_2), Y^{n_1(\tilde{s}_1)}, S^{n_1(\tilde{s}_1)}) \in A_\varepsilon^{*(n_1(\tilde{s}_1))}(X_1,X_2,Y,S)$ given $\tilde{S}_1 = \tilde{s}_1$ for all $\tilde{s}_1 \in \mathcal{S}$, otherwise it declares error.

Analysis of the probability of error: First, we analyze the probability of error for the component codeword $V_2^{\tilde{s}_1,\tilde{s}_2}$ at encoder 2, i.e., $\Pr(N_{\tilde{s}_1,\tilde{s}_2} < n_2(\tilde{s}_1,\tilde{s}_2))$. Since that the state process is stationary and ergodic $\lim_{n\to\infty} \frac{N(\tilde{s}_1,\tilde{s}_2)}{n} = P(\tilde{s}_1,\tilde{s}_2)$ in probability. Therefore, $\Pr(N_{\tilde{s}_1,\tilde{s}_2} < n_2(\tilde{s}_1,\tilde{s}_2)) \to 0$ as $n\to\infty$. Now, we analyze the probability to decode incorrectly the component codeword $V_2^{\tilde{s}_1,\tilde{s}_2}$ that was sent from the $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ codebook of encoder 2. Without loss of generality, we can assume that the first codeword was sent from the $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ codebook of encoder 2, which we denote by $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}(1)$. Since $S^{n_2(\tilde{s}_1,\tilde{s}_2)}$ is ergodic and by using the Law of Large Numbers (L.L.N.) as $n_2(\tilde{s}_1,\tilde{s}_2)\to\infty$ we have $\Pr\left\{S^{n_2(\tilde{s}_1,\tilde{s}_2)} \in A_\varepsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(S)\right\} \to 1$. By the construction of the codebook $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}(1)$, X_2 and S are independent given $(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)$. Hence $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(1)$ and $S^{n_2(\tilde{s}_1,\tilde{s}_2)}$ are strongly jointly typical sequences with probability 1. Finally from the codebooks construction and the channel transition probability we have that

$$
\begin{aligned}
&p(y_i|x_2^i, s^i, \tilde{s}_1, \tilde{s}_2) \\
&= \sum_{x_{1,i}\in X_{1,i}} p(x_{1,i}|x_2^i, s^i, \tilde{s}_1, \tilde{s}_2) p(y_i|x_{1,i}, x_2^i, s^i, \tilde{s}_1, \tilde{s}_2) \\
&= \sum_{x_{1,i}\in X_{1,i}} p(x_{1,i}|\tilde{s}_1, \tilde{s}_2) p(y_i|x_{1,i}, x_{2,i}, s_i, \tilde{s}_1, \tilde{s}_2) \\
&= p(y_i|x_{2,i}, s_i, \tilde{s}_1, \tilde{s}_2).
\end{aligned}
$$

Now using the fact that $p(y_i|x_2^i, s^i, \tilde{s}_1, \tilde{s}_2) = p(y_i|x_{2,i}, s_i, \tilde{s}_1, \tilde{s}_2)$, and the L.L.N. we have

$$
\Pr\left\{\left(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(1), S^{n_2(\tilde{s}_1,\tilde{s}_2)}, Y^{n_2(\tilde{s}_1,\tilde{s}_2)}\right) \in A_\epsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)|(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)\right\} \to 1,
$$

as $n_2(\tilde{s}_1,\tilde{s}_2)\to\infty$. A decoding error occurs only if either the correct codeword $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(1)$, is not strongly jointly typical with $(Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)})$, or there is an incorrect codeword $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i)$, where $i \neq 1$, that is strongly jointly typical with $(Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)})$. Let us define these events:

$$
E_1 = \left\{\left(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(1), Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)}\right) \notin A_\epsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)|(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)\right\}, \quad (22)
$$

$$
E_2 = \left\{\exists i \neq 1 : \left(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i), Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)}\right) \in A_\epsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)|(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)\right\}. \quad (23)
$$

In addition, we define the following event:

$$
E_{2,i} = \left\{\left(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i), Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)}\right) \in A_\epsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}(X_2,Y,S)|(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)\right\}. \quad (24)
$$

Then by the union of events bound:

$$
\begin{aligned}
P_e^{(n_2(\tilde{s}_1,\tilde{s}_2))} &= \Pr(E_1 \cup E_2) \\
&\leq P(E_1) + P(E_2).
\end{aligned} \quad (25)
$$

Now let us find the probability of each event,

1) $P(E_1)$- As mentioned above as $n_2(\tilde{s}_1,\tilde{s}_2)\to\infty$ we have,

$$
P(E_1) \to 0.
$$

2) $P(E_2)$- for $i \neq 1$ the probability of error,

$$
\begin{aligned}
&P(E_2) \\
&= \Pr\left\{\left(X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i), Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)}\right) \in A_\epsilon^{*(n_2(\tilde{s}_1,\tilde{s}_2))}|(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)\right\} \\
&\leq \sum_{i=2}^{2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)}} P(E_{2,i}) \\
&\overset{(a)}{\leq} 2^{n_2(\tilde{s}_1,\tilde{s}_2)R_2(\tilde{s}_1,\tilde{s}_2)} \\
&\quad \times 2^{-n_2(\tilde{s}_1,\tilde{s}_2)\left(I\left(X_2;Y,S|\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2\right)-\epsilon\right)},
\end{aligned} \quad (26)
$$

where (a) follows from the fact that $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(1)$ and $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i)$ are independent given $(\tilde{S}_1,\tilde{S}_2) = (\tilde{s}_1,\tilde{s}_2)$ for $i \neq 1$, so are $(Y^{n_2(\tilde{s}_1,\tilde{s}_2)}, S^{n_2(\tilde{s}_1,\tilde{s}_2)})$ and $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i)$. Hence, using [16, Lemma 10.6.2] the probability that $X_2^{n_2(\tilde{s}_1,\tilde{s}_2)}(i)$ and $(Y^{n_2(\tilde{s}_1,\tilde{s}_2}, S^{n_2(\tilde{s}_1,\tilde{s}_2)})$ are strongly jointly typical is

$2^{-n_2(\tilde{s}_1,\tilde{s}_2)\left(I\left(X_2;Y,S|\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2\right)-\epsilon\right)}$. For $P(E_2)\to 0$ as $n_2(\tilde{s}_1,\tilde{s}_2)\to\infty$, we need to choose

$$\begin{aligned} R_2(\tilde{s}_1,\tilde{s}_2) &\leq I(X_2;Y,S|\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2)-\epsilon \\ &= I(X_2;Y|S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2) \\ &\quad + I(X_2;S|\tilde{S}_1=\tilde{s}_2,\tilde{S}_2=\tilde{s}_2)-\epsilon \\ &\overset{(a)}{=} I(X_2;Y|S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2)-\epsilon \end{aligned}$$

where (a) follows from the independence of X_2 and S given $(\tilde{S}_1,\tilde{S}_2)=(\tilde{s}_1,\tilde{s}_2)$.

Similarly, we can analyze the probability of error to the rest of the codebooks of encoder 2, i.e., $\mathcal{C}_2^{\tilde{s}_1,\tilde{s}_2}$ for every $(\tilde{s}_1,\tilde{s}_2)\in\{\mathcal{S}\times\mathcal{S}\}$. Therefore, as $n\to\infty$

$$\begin{aligned} &R_2 \\ &\leq \sum_{\tilde{s}_1,\tilde{s}_2}\frac{n_2(\tilde{s}_1,\tilde{s}_2)}{n}R_2(\tilde{s}_1,\tilde{s}_2) \\ &\leq \sum_{\tilde{s}_1,\tilde{s}_2}\frac{n_2(\tilde{s}_1,\tilde{s}_2)}{n}(I(X_2;Y|S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2)-\epsilon) \\ &= \sum_{\tilde{s}_1,\tilde{s}_2}(P(\tilde{s}_1,\tilde{s}_2)-\epsilon')(I(X_2;Y|S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2)-\epsilon) \\ &= I(X_2;Y|S,\tilde{S}_1,\tilde{S}_2)-\epsilon'' \end{aligned} \tag{27}$$

where $\epsilon''=\epsilon+\epsilon'\sum_{\tilde{s}_1,\tilde{s}_2}I(X_2;Y|S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2=\tilde{s}_2)-\epsilon\epsilon'$.

Let us analyze the probability of error for the component codeword $V_1^{\tilde{s}_1}$. As mentioned above, since that the state process is stationary and ergodic $\lim_{n\to\infty}\frac{N(\tilde{s}_1)}{n}=P(\tilde{s}_1)$ in probability. Therefore, the probability that an error is declared at encoder 1, $\Pr(N_{\tilde{s}_1}<n_1(\tilde{s}_1))\to 0$ as $n\to\infty$. Now, we analyze the probability to decode incorrectly the component codeword $V_1^{\tilde{s}_1}$, that was sent from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook of encoder 1 after $\hat{M}_2$ was decoded correctly. Without loss of generality, we can assume that the first codeword was sent from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook of encoder 1, i.e., $\mathcal{C}_2^{\tilde{s}_1}(1)$ was sent. Again from the ergodicity of $S^{n_1(\tilde{s}_1)}$, the construction of the codebooks, and channel transition probability we have that

$$\Pr\left\{(X_1^{n_1(\tilde{s}_1)}(1),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)})\in A_\epsilon^{*(n_1(\tilde{s}_1))}(X_1,X_2,Y,S)|\tilde{S}_1=\tilde{s}_1\right\}\to 1$$

as $n_1(\tilde{s}_1)\to\infty$.

A decoding error occurs only if either

$$\left(X_1^{n_1(\tilde{s}_1)}(1),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)}\right)$$

are not strongly jointly typical, or there is an incorrect codeword $X_1^{n_1(\tilde{s}_1)}(i)$, where $i\neq 1$, that is strongly jointly typical with $(X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)})$. Let us define these events:

$$E_3=\left\{\left(X_1^{n_1(\tilde{s}_1)}(1),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)}\right)\notin A_\epsilon^{*(n_1(\tilde{s}_1))}(X_1,X_2,Y,S)|\tilde{S}_1=\tilde{s}_1\right\}, \tag{28}$$

$$E_4=\left\{\exists i\neq 1:\left(X_1^{n_1(\tilde{s}_1)}(i),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)}\right)\in A_\epsilon^{*(n_1(\tilde{s}_1))}(X_1,X_2,Y,S)|\tilde{S}_1=\tilde{s}_1\right\}. \tag{29}$$

In addition, we define the following event:

$$E_{4,i}=\left\{\left(X_1^{n_1(\tilde{s}_1)}(i),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)}\right)\in A_\epsilon^{*(n_1(\tilde{s}_1))}(X_1,X_2,Y,S)|\tilde{S}_1=\tilde{s}_1\right\}. \tag{30}$$

Then by the union of events bound,

$$\begin{aligned} P_e^{(n_1(\tilde{s}_1))} &= \Pr(E_3\cup E_4) \\ &\leq P(E_3)+P(E_4). \end{aligned} \tag{31}$$

Now let us find the probability of each event:

1) $P(E_3)$- As mentioned above as $n_1(\tilde{s}_1)\to\infty$ we have,

$$P(E_3)\to 0.$$

2) $P(E_4)$- for $i\neq 1$ the probability of error,

$$\begin{aligned} &P(E_4) \\ &= \Pr\left(\left(X_1^{n_1(\tilde{s}_1)}(i),X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)}\right)\in A_\epsilon^{(n_1(\tilde{s}_1))}|\tilde{S}_1=\tilde{s}_1\right) \\ &\leq \sum_{i=2}^{2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}}P(E_{4,i}) \\ &\overset{(a)}{\leq} 2^{n_1(\tilde{s}_1)R_1(\tilde{s}_1)}\cdot 2^{-n_1(\tilde{s}_1)\left(I\left(X_1;X_2,Y,S|\tilde{S}_1=\tilde{s}_1\right)-\epsilon\right)}, \end{aligned}$$

where (a) follows from the fact that $X_1^{n_1(\tilde{s}_1)}(i)$ and $(X_2^{n_1(\tilde{s}_1)}(\hat{M}_2),Y^{n_1(\tilde{s}_1)},S^{n_1(\tilde{s}_1)})$ are independent given $\tilde{S}_1=\tilde{s}_1$. For $P(E_4)\to 0$ as $n_1(\tilde{s}_1)\to\infty$, we need to choose,

$$\begin{aligned} R_1(\tilde{s}_1) &\leq I(X_1;X_2,Y,S|\tilde{S}_1=\tilde{s}_1)-\epsilon \\ &= I(X_1;Y|X_2,S,\tilde{S}_1=\tilde{s}_1) \\ &\quad + I(X_1;X_2,S|\tilde{S}_1=\tilde{s}_1)-\epsilon \\ &\overset{(a)}{=} I(X_1;Y|X_2,S,\tilde{S}_1=\tilde{s}_1)-\epsilon, \\ &= H(Y|X_2,S,\tilde{S}_1=\tilde{s}_1) \\ &\quad - H(Y|X_1,X_2,S,\tilde{S}_1=\tilde{s}_1)-\epsilon, \\ &\overset{(b)}{=} H(Y|X_2,S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2) \\ &\quad - H(Y|X_1,X_2,S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2)-\epsilon, \\ &= I(X_1;Y|X_2,S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2)-\epsilon, \end{aligned}$$

where (a) follows from the independence of X_1 and (X_2,S) given $\tilde{S}_1=\tilde{s}_1$, and (b) follows from the independence of Y and $\tilde{S}_2$ given $(X_2,S,\tilde{S}_1=\tilde{s}_1)$.

Similarly, we can analyze the probability of error to the rest of the codbooks of encoder 1, i.e., $\mathcal{C}_1^{\tilde{s}_1}$ for every $\tilde{s}_1\in\{\mathcal{S}\}$. Therefore, as $n\to\infty$

$$\begin{aligned} R_1 &\leq \sum_{\tilde{s}_1}\frac{n_1(\tilde{s}_1)}{n}R_1(\tilde{s}_1) \\ &\leq \sum_{\tilde{s}_1}\frac{n_1(\tilde{s}_1)}{n}(I(X_1;Y|X_2,S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2)-\epsilon) \\ &= \sum_{\tilde{s}_1}(P(\tilde{s}_1)-\epsilon')(I(X_1;Y|X_2,S,\tilde{S}_1=\tilde{s}_1,\tilde{S}_2)-\epsilon) \\ &= I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2)-\epsilon'', \end{aligned} \tag{32}$$

where $\epsilon'' = \epsilon + \epsilon' \sum_{\tilde{s}_1} I(X_1; Y|X_2, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2) - \epsilon\epsilon'$.

Thus the total average probability of decoding error $P_e^{(n)} \to 0$ as $n \to \infty$ if $R_1 < I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2), R_2 < I(X_2; Y|S, \tilde{S}_1, \tilde{S}_2)$. The achievability of the other corner point follows by changing the decoding order. To show achievability of other points in $\mathcal{R}(X_1, X_2)$, we use time sharing between corner points and points on the axes. Thus, the probability of error, conditioned on a particular codeword being sent, goes to zero if the conditions of the following are met:

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2), \\ R_2 &\leq I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2). \end{aligned} \tag{33}$$

To show that the region in Theorem 1 is achievable, we use time sharing, where the time sharing is a function of $\tilde{s}_1$. In the analysis of the probability of error, we analyze the probability to decode incorrectly the component codeword $V_2^{\tilde{s}_1, \tilde{s}_2}$ that was sent from the $\mathcal{C}_2^{\tilde{s}_1, \tilde{s}_2}$ codebook of encoder 2. We derived that in order that the probability of error will be arbitrarily small, one have to choose,

$$R_2(\tilde{s}_1, \tilde{s}_2) \leq I(X_2; X_1, Y|S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2 = \tilde{s}_2) - \epsilon.$$

In addition, we analyze the probability to decode incorrectly the component codeword $V_1^{\tilde{s}_1}$ that was sent from the $\mathcal{C}_1^{\tilde{s}_1}$ codebook of encoder 1. We obtained,

$$R_1(\tilde{s}_1) \leq I(X_1; Y|X_2, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2) - \epsilon.$$

Now, since $\tilde{s}_1$ is fixed, and the fact that $\tilde{s}_1$ is known at the decoder and in both of the encoders, we can use time sharing, where the time sharing is a function of $\tilde{s}_1$,

$$\begin{aligned} &R_2(\tilde{s}_1, \tilde{s}_2) \\ &\leq \sum_u p(u|\tilde{s}_1) I(X_2; X_1, Y|S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2 = \tilde{s}_2, U = u) - \epsilon, \\ &= I(X_2; X_1, Y|S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2 = \tilde{s}_2, U) - \epsilon. \end{aligned}$$

Similarly, for encoder 1

$$\begin{aligned} &R_1(\tilde{s}_1) \\ &\leq \sum_u p(u|\tilde{s}_1) I(X_1; Y|X_2, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2, U = u) - \epsilon, \\ &= I(X_1; Y|X_2, S, \tilde{S}_1 = \tilde{s}_1, \tilde{S}_2, U) - \epsilon. \end{aligned}$$

Similarly, we can analyze the probability of error to the rest of the codebooks of encoder 1 and encoder 2. Hence, the probability of error, conditioned on a particular codeword being sent, goes to zero if the following conditions are met:

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_2 &\leq I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U). \end{aligned} \tag{34}$$

The above bound shows that the average probability of error, which by symmetry is equal to the probability for an individual pair of codewords (m_1, m_2), averaged over all choices of codebooks in the random code construction, is arbitrarily small. Hence, there exists at least one $(n, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code with arbitrarily small probability of error. This completes the proof of achivability proof. ■

VI. Alternative Proof

In this section, we provide an alternative proof for Theorem 1. The alternative proof is based on a multi-letter expression for the capacity region of FS-MAC with time-invariant feedback [12]. In order to use the capacity region of FS-MAC with time-invariant feedback, we treat the knowledge of the state at the encoders as being part of the feedback from the decoder to the encoders.

Throughout this section we use the causal conditioning notation $(\cdot\|\cdot)$. We denote the probability mass function (pmf) of Y^n causally conditioned on X^{n-d}, for some integer $d \geq 0$, as $P(y^n\|x^{n-d})$ which is defined as

$$P(y^n\|x^{n-d}) = \prod_{i=1}^{n} P(y_i|y^{i-1}, x^{i-d}) \tag{35}$$

(if $i - d \leq 0$ then x^{i-d} is set to null). The directed information $I(X^n \to Y^n)$ was defined by Massey in [17] as

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i|Y^{i-1}). \tag{36}$$

Directed information has been widely used in the characterization of capacity of point-to-point channels [8], [18]–[22], compound channels [23], network capacity [24], rate distortion [25], [26], and broadcast channel [27]. Directed information can also be expressed in terms of causal conditioning as

$$\begin{aligned} I(X^n \to Y^n) &= \sum_{i=1}^{n} I(X^i; Y_i|Y^{i-1}) \\ &= \mathbf{E}\left[\log \frac{P(Y^n\|X^n)}{P(Y^n)}\right] \end{aligned} \tag{37}$$

where $\mathbf{E}$ denotes expectation. Directed information between X_1^n to Y^n causally conditioned on X_2^n is defined as

$$\begin{aligned} I(X_1^n \to Y^n\|X_2^n) &\triangleq \sum_{i=1}^{n} I(X_1^i; Y_i|Y^{i-1}, X_2^i) \\ &= \mathbf{E}\left[\log \frac{P(Y^n\|X_1^n, X_2^n)}{P(Y^n\|X_2^n)}\right] \end{aligned} \tag{38}$$

where $P(y^n\|x_1^n, x_2^n) = \prod_{i=1}^{n} P(y_i|y^{i-1}, x_1^i, x_2^i)$.

Now let us present a result from [12] that we need for the proof. Consider the FS-MAC with time-invariant feedback as illustrated in Fig. 4. The channel is characterized by a conditional probability $P(y_i, s_{i+1}|x_{1,i}, x_{2,i}, s_i)$ that satisfies

$$P(y_i, s_{i+1}|x_1^i, x_2^i, s^i, y^{i-1}) = P(y_i, s_{i+1}|x_{1,i}, x_{2,i}, s_i). \tag{39}$$

In addition, we assume that the channel is stationary, indecomposable, and without ISI, i.e.,

$$P(y_i, s_{i+1}|x_{1,i}, x_{2,i}, s_i) = p(s_{i+1}|s_i) p(y_i|x_{1,i}, x_{2,i}, s_i) \tag{40}$$

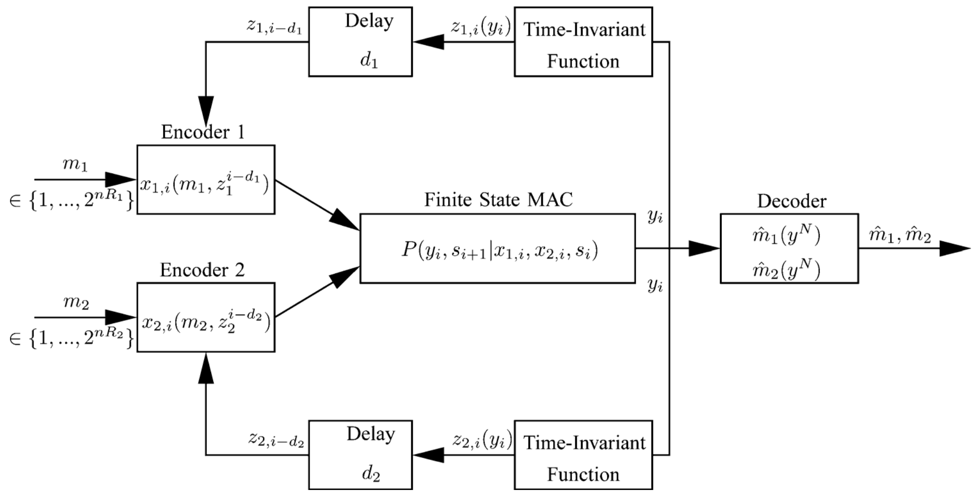

Fig. 4. Channel with feedback, where the channel feedbacks, Z_1 and Z_2, are time-invariant deterministic functions of the output, Y.

and

$$P(s_0) = \pi(s_0) \tag{41}$$

where $\pi(s_0)$ is the unique stationary distribution, i.e., $\lim_{n\to\infty} \Pr(S_n = s|s_0) = \pi(s_0)$, $\forall s_0 \in \mathcal{S}$.

Lemma 4 [12, Theorem 13]: The capacity of a stationary, indecomposable FS-MAC without ISI and with time-invariant feedback, as illustrated in Fig. 4, is $\hat{\mathcal{R}} = \lim_{n\to\infty} \hat{\mathcal{R}}_n$, where $\hat{\mathcal{R}}_n$ is given by

$$\hat{\mathcal{R}}_n = \bigcup_{\mathcal{L}} \begin{pmatrix} R_1 \le \frac{1}{n} I(X_1^n \to Y^n \| X_2^n), \\ R_1 \le \frac{1}{n} I(X_2^n \to Y^n \| X_1^n), \\ R_1 + R_2 \le \frac{1}{n} I((X_1, X_2)^n \to Y^n). \end{pmatrix} \tag{42}$$

where $\mathcal{L}$ is the set of joint distributions of the form $P(x_1^n \| z^{n-d_1}) P(x_2^n \| z^{n-d_2})$.

In [12], Theorem 13, only the case where $d_1 = d_2 = 1$ was considered, but the result extends straightforwardly to any delay d_1 and d_2. The following theorem provides an alternative proof for Theorem 1 based on Lemma 4.

Theorem 5: Let us denote $\mathcal{R}_n$ and $\mathcal{R}$ to be the following regions:

$$\mathcal{R}_n = \bigcup_{\mathcal{P}} \begin{pmatrix} R_1 \le \frac{1}{n} I(X_1^n \to Y^n, S^n \| X_2^n), \\ R_1 \le \frac{1}{n} I(X_2^n \to Y^n, S^n \| X_1^n), \\ R_1 + R_2 \le \frac{1}{n} I((X_1, X_2)^n \to Y^n, S^n), \end{pmatrix}, \tag{43}$$

$$\mathcal{R} = \bigcup_{\mathcal{M}} \begin{pmatrix} R_1 < I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_2 < I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U), \\ R_1 + R_2 < I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U), \end{pmatrix} \tag{44}$$

where $\mathcal{P}$ and $\mathcal{M}$ are the sets of joint distributions of the form $P(x_1^n \| s^{n-d_1}) P(x_2^n \| s^{n-d_2})$ and $P(u|\tilde{s}_1) P(x_1|\tilde{s}_1, u) P(x_2|\tilde{s}_1, \tilde{s}_2, u)$, respectively. The capacity region for the FSM-MAC with CSI at the decoder and asymmetrical delayed CSI at the encoders with delays d_1 and d_2, as illustrated in Fig. 1, is $\lim_{n\to\infty} \mathcal{R}_n = \mathcal{R}$.

Proof: In order to adapt the model in Fig. 4 to our model, we can consider the state information at the decoder as a part of the channel's output. Therefore, the capacity region is

$$\mathcal{R}_{feedback} = \lim_{n\to\infty} \bigcup_{\mathcal{L}} \begin{cases} R_1 \le \frac{1}{n} I(X_1^n \to Y^n, S^n \| X_2^n), \\ R_1 \le \frac{1}{n} I(X_2^n \to Y^n, S^n \| X_1^n), \\ R_1 + R_2 \le \frac{1}{n} I((X_1, X_2)^n \to Y^n, S^n) \end{cases} \tag{45}$$

where $\mathcal{L} \triangleq \{P(x_1^n \| z^{n-d_1}) P(x_2^n \| z^{n-d_2})\}$. Now, by choosing the deterministic function of the output $z_{1,i}(y_i, s_i) = z_{2,i}(y_i, s_i) = s_i$, (45) yields the capacity region for the FSM-MAC with CSI at the decoder and asymmetrical delayed CSI at the encoders as shown in Fig. 1. Note that $\mathcal{R}_{feedback} = \lim_{n\to\infty} \mathcal{R}_n$, hence the capacity region is $\lim_{n\to\infty} \mathcal{R}_n$. In order to complete the proof we need to show that $\lim_{n\to\infty} \mathcal{R}_n = \mathcal{R}$. First let us show that $\lim_{n\to\infty} \mathcal{R}_n \supseteq \mathcal{R}$:

$$\mathcal{R}_n = \bigcup_{\mathcal{P}} \begin{cases} R_1 \le \frac{1}{n} I(X_1^n \to Y^n, S^n \| X_2^n), \\ R_1 \le \frac{1}{n} I(X_2^n \to Y^n, S^n \| X_1^n), \\ R_1 + R_2 \le \frac{1}{n} I((X_1, X_2)^n \to Y^n, S^n). \end{cases}$$

$$= \bigcup_{\mathcal{P}} \begin{cases} R_1 \le \frac{1}{n} \sum_{i=1}^n I(X_1^i; Y_i, S_i | X_2^i, Y^{i-1}, S^{i-1}), \\ R_1 \le \frac{1}{n} \sum_{i=1}^n I(X_2^i; Y_i, S_i | X_1^i, Y^{i-1}, S^{i-1}), \\ R_1 + R_2 \le \frac{1}{n} \sum_{i=1}^n I(X_1^i, X_2^i; Y_i, S_i | Y^{i-1}, S^{i-1}). \end{cases}$$

To bound R_1, consider

$$\begin{aligned} R_1 &\le \frac{1}{n} \sum_{i=1}^n \{I(X_1^i; Y_i, S_i | X_2^i, Y^{i-1}, S^{i-1})\} \\ &= \frac{1}{n} \sum_{i=1}^n \{H(Y_i, S_i | X_2^i, Y^{i-1}, S^{i-1}) \\ &\qquad - H(Y_i, S_i | X_1^i, X_2^i, Y^{i-1}, S^{i-1})\} \\ &= \frac{1}{n} \sum_{i=1}^n \{H(S_i | X_2^i, Y^{i-1}, S^{i-1}) \\ &\qquad + H(Y_i | X_2^i, Y^{i-1}, S^i)\} \\ &\quad - \frac{1}{n} \sum_{i=1}^n \{H(S_i | X_1^i, X_2^i, Y^{i-1}, S^{i-1}) \\ &\qquad + H(Y_i | X_1^i, X_2^i, Y^{i-1}, S^i)\} \end{aligned}$$

$$\overset{(a)}{=}\frac{1}{n}\sum_{i=1}^{n}\left\{H(S_i|S^{i-1})+H(Y_i|X_2^i,Y^{i-1},S^i)\right.$$
$$\left.-H(S_i|S^{i-1})-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}$$
$$=\frac{1}{n}\sum_{i=1}^{n}\left\{H(Y_i|X_2^i,Y^{i-1},S^i)\right.$$
$$\left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}.$$

where (a) follows from the fact that the channel is without ISI, and from the fact that the channel's output at time i depends only on the state S_i, and the inputs $X_{1,i}$, $X_{2,i}$. We can bound R_2 and R_1+R_2 in a similar way. Hence we obtain

$$\mathcal{R}_n=\bigcup_{\mathcal{P}}\left\{\begin{array}{rcl} R_1 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_2^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_1^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_1+R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}. \end{array}\right..$$

The fact that the sequence

$$\left\{P(x_{1,i}|x_1^{i-1},s^{i-d_1})\,P(x_{2,i}|x_2^{i-1},s^{i-d_2})\right\}_{i=1}^n$$

determines uniquely the term $P(x_1^n\|s^{n-d_1})P(x_2^n\|s^{n-d_2})$ follows immediately from the definition of the later. Now using [21, Lemma 3], we have that $P(x_1^n\|s^{n-d_1})P(x_2^n\|s^{n-d_2})$ determines uniquely $\left\{P(x_{1,i}|x_1^{i-1},s^{i-d_1})\,P(x_{2,i}|x_2^{i-1},s^{i-d_2})\right\}_{i=1}^n$. Hence

$$\mathcal{R}_n=\bigcup_{\mathcal{T}}\left\{\begin{array}{rcl} R_1 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_2^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_1^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_1+R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}. \end{array}\right.$$

where $\mathcal{T}\triangleq\{P(x_{1,i}|x_1^{i-1},s^{i-d_1})P(x_{2,i}|x_2^{i-1},s^{i-d_2})\}_{i=1}^n$. Let us assume that $d_1\ge d_2$, furthermore, we restrict the inputs of the channel by assuming that $P(x_{1,i}|x_1^{i-1},s^{i-d_1})=P(x_{1,i}|s_{i-d_1})$, $P(x_{2,i}|x_2^{i-1},s^{i-d_2})=P(x_{2,i}|s_{i-d_1},s_{i-d_2})$. Therefore

$$\mathcal{R}_n\supseteq\bigcup_{\mathcal{W}}\left\{\begin{array}{rcl} R_1 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_2^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_1^i,Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_1+R_2 & \le & \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|Y^{i-1},S^i)\right. \\ & & \left.-H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \end{array}\right.$$

where $\mathcal{W}\triangleq\{P(x_{1,i}|s_{i-d_1})P(x_{2,i}|s_{i-d_1},s_{i-d_2})\}_{i=1}^n$. Since we assumed that $P(x_{1,i}|x_1^{i-1},s^{i-d_1})=P(x_{1,i}|s_{i-d_1})$, we have the following equalities:

$$P(y_i|x_2^i,y^{i-1},s^i)$$
$$=\sum_{x_{1,i}}P(x_{1,i}|x_2^i,y^{i-1},s^i)P(y_i|x_{1,i},x_2^i,y^{i-1},s^i)$$
$$\overset{(a)}{=}\sum_{x_{1,i}}P(x_{1,i}|s_i,s_{i-d_1},s_{i-d_2})P(y_i|x_{1,i},x_{2,i},s_i,s_{i-d_1},s_{i-d_2})$$
$$=P(y_i|x_{2,i},s_i,s_{i-d_1},s_{i-d_2}) \tag{46}$$

where (a) follows from the fact that the channel's output at time i depends only on the state S_i, and the inputs $X_{1,i}$, $X_{2,i}$, and from the fact that $P(x_{1,i}|x_2^i,y^{i-1},s^i)=P(x_{1,i}|s^{i-d_1})=P(x_{1,i}|s_{i-d_1})$. From (46) we get

$$H(Y_i|X_2^i,Y^{i-1},S^i)=H(Y_i|X_{2,i},S_i,S_{i-d_1},S_{i-d_2}).$$

Similarly

$$H(Y_i|X_1^i,Y^{i-1},S^i)=H(Y_i|X_{1,i},S_i,S_{i-d_1},S_{i-d_2}).$$
$$H(Y_i|Y^{i-1},S^i)=H(Y_i|S_i,S_{i-d_1},S_{i-d_2}).$$

Therefore

$$\mathcal{R}_n\supseteq\bigcup_{\mathcal{W}}\left\{\begin{array}{l} R_1\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1},S_{i-d_2}), \\ R_2\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_1},S_{i-d_2}), \\ R_1+R_2\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_1},S_{i-d_2}). \end{array}\right.$$

Now, in order to obtain that $\lim_{n\to\infty}\mathcal{R}_n\supseteq\mathcal{R}$, we need to show that

$$\mathcal{R}\subseteq\lim_{n\to\infty}\bigcup_{\mathcal{W}}\left\{\begin{array}{l} R_1\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1},S_{i-d_2}), \\ R_2\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_1},S_{i-d_2}), \\ R_1+R_2\le\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_1},S_{i-d_2}). \end{array}\right.$$

Consider the region $\mathcal{R}$, an achievable region is uniquely determined for every fixed joint distribution $P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)$. The rate R_1 is given by

$$R_1\le I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2,U)$$
$$=\sum_{\tilde{s}_1}P(\tilde{s}_1)\sum_{u}P(u|\tilde{s}_1)I(X_1;Y|X_2,S,\tilde{s}_1,\tilde{S}_2,U=u). \tag{47}$$

In addition, we have

$$\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1},S_{i-d_2})$$
$$=\frac{1}{n}\sum_{i=1}^n\sum_{s_{i-d_1}}P(s_{i-d_1})I(Y_i;X_{1,i}|X_{2,i},S_i,s_{i-d_1},S_{i-d_2})$$
$$\overset{(a)}{=}\sum_{\tilde{s}_1}P(\tilde{s}_1)\sum_{i=1}^n\frac{1}{n}I(Y_i;X_{1,i}|X_{2,i},S_i,\tilde{s}_1,S_{i-d_2}) \tag{48}$$

where (a) follows from the fact that the distribution $P(s_{i-d_1})$ is stationary, therefore $P(s_{i-d_1})=P(\tilde{s}_1)$. For every $U=u$ and $\tilde{S}_1=\tilde{s}_1$, if $P(U=u|\tilde{S}_1=\tilde{s}_1)$ is rational of the form $k(u,\tilde{s}_1)/n$, where $k(u,\tilde{s}_1)\in\mathbb{N}$, then we can choose $k(u,\tilde{s}_1)$ terms from $\{P(x_{1,i}|s_{i-d_1})P(x_{2,i}|s_{i-d_1},s_{i-d_2})\}_{i=1}^n$ such that $P(x_{1,i}|s_{i-d_1})P(x_{2,i}|s_{i-d_1},s_{i-d_2})=P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)$. If

$P(U=u|\tilde{S}_1=\tilde{s}_1)$ is irrational or rational but not of the form $k(u,\tilde{s}_1)/n$, we can get arbitrarily close to $P(U=u|\tilde{S}_1=\tilde{s}_1)$ by using longer and longer block lengths. Therefore, using (47) and (48) we have that when $n\to\infty$, for every given joint distribution $P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)$, we can choose $\{P(x_{1,i}|s_{i-d_1})P(x_{2,i}|s_{i-d_1},s_{i-d_2})\}_{i=1}^n$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1},S_{i-d_2}) = I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2,U).$$

By using the same argument for R_2 and for R_1+R_2, we get that for every given joint distribution $P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)$, we can chose $\{P(x_{1,i}|s_{i-d_1})P(x_{2,i}|s_{i-d_1},s_{i-d_2})\}_{i=1}^n$ such that the following equalities hold simultaneously:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1},S_{i-d_2}) = I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2,U), \tag{49}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_1},S_{i-d_2}) = I(X_2;Y|X_1,S,\tilde{S}_1,\tilde{S}_2,U), \tag{50}$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_1},S_{i-d_2}) = I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2,U). \tag{51}$$

Using (49)–(51), we obtain

$$\lim_{n\to\infty}\mathcal{R}_n \supseteq \mathcal{R}. \tag{52}$$

In order to complete the proof, we need to show that $\lim_{n\to\infty}\mathcal{R}_n\subseteq\mathcal{R}$. We have that

$$\mathcal{R}_n = \bigcup_{\mathcal{T}}\left\{\begin{array}{r} R_1 \le \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_2^i,Y^{i-1},S^i) \right.\\ \left. -H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_2 \le \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_1^i,Y^{i-1},S^i)\right. \\ \left. -H(Y_i|X_{1,i},X_{2,i},S_i)\right\}, \\ R_1+R_2 \le \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|Y^{i-1},S^i)\right. \\ \left. -H(Y_i|X_{1,i},X_{2,i},S_i)\right\}. \end{array}\right. .$$

Consider the rate R_1,

$$\begin{aligned} R_1 &\le \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_2^i,Y^{i-1},S^i)-H(Y_i|X_{1,i},X_{2,i},S_i)\right\} \\ &\le \frac{1}{n}\sum_{i=1}^n\left\{H(Y_i|X_{2,i},S_i,S_{i-d_2},S^{i-d_1}) \right.\\ &\qquad \left. -H(Y_i|X_{1,i},X_{2,i},S_i)\right\} \\ &= \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_2},S^{i-d_1}). \end{aligned}$$

We can bound R_2 and R_1+R_2 in a similar way. Hence we get

$$\mathcal{R}_n \subseteq \bigcup_{\mathcal{T}}\left\{\begin{array}{r} R_1 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_2},S^{i-d_1}), \\ R_2 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_2},S^{i-d_1}), \\ R_1+R_2 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_2},S^{i-d_1}). \end{array}\right. \tag{53}$$

Now, consider the joint distribution

$$\begin{aligned} &P(s_i,s_{i-d_2},s^{i-d_1},x_{1,i},x_{2,i},y_i) \\ &= P(s_i,s_{i-d_2},s^{i-d_1})P(x_{1,i}|s^{i-d_1})P(x_{2,i}|x_{1,i},s^{i-d_1},s_{i-d_2}) \\ &\quad\times P(y_i|x_{1,i},x_{2,i},s_i) \\ &\overset{(a)}{=} P(s_i,s_{i-d_2},s^{i-d_1})P(x_{1,i}|s^{i-d_1})P(x_{2,i}|s^{i-d_1},s_{i-d_2})) \\ &\quad\times P(y_i|x_{1,i},x_{2,i},s_i) \end{aligned}$$

where (a) follows from the fact that

$$\begin{aligned} &P(x_{2,i}|x_{1,i},s^{i-d_1},s_{i-d_2}) \\ &= \sum_{M_2,s_{i-d_1+1}^{i-d_2-1}}\Big\{P(M_2,s_{i-d_1+1}^{i-d_2-1}|x_{1,i},s^{i-d_1},s_{i-d_2}) \\ &\qquad\qquad \times P(x_{2,i}|x_{1,i},s^{i-d_2},M_2)\Big\} \\ &= \sum_{M_2,s_{i-d_1+1}^{i-d_2-1}} P(M_2,s_{i-d_1+1}^{i-d_2-1}|s^{i-d_1},s_{i-d_2})P(x_{2,i}|s^{i-d_2},M_2) \\ &= P(x_{2,i}|s^{i-d_1},s_{i-d_2}). \end{aligned}$$

Note that the elements

$$I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_2},S^{i-d_1}),$$
$$I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_2},S^{i-d_1}),$$

and

$$I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_2},S^{i-d_1}),$$

are uniquely determined by the joint distribution $P(s_i,s_{i-d_2},s^{i-d_1},x_{1,i},x_{2,i},y_i)$. Hence, R_1, R_2, and R_1+R_2 in (53) are uniquely determined by the joint distribution $\left\{P(s_i,s_{i-d_2},s^{i-d_1},x_{1,i},x_{2,i},y_i)\right\}_{i=1}^n$. In the joint distribution $P(s_i,s_{i-d_2},s^{i-d_1},x_{1,i},x_{2,i},y_i)$, we control only $P(x_{1,i}|s^{i-d_1})P(x_{2,i}|s^{i-d_1},s_{i-d_2})$, since the distributions $P(s_i,s_{i-d_2},s^{i-d_1})$ and $P(y_i|x_{1,i},x_{2,i},s_i)$ are determined by the channel transition probability. Hence

$$\mathcal{R}_n \subseteq \bigcup_{\mathcal{V}}\left\{\begin{array}{r} R_1 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d_1}, \\ S_{i-d_2},S^{i-d_1-1}), \\ R_2 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d_1}, \\ S_{i-d_2},S^{i-d_1-1}), \\ R_1+R_2 \le \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d_1}, \\ S_{i-d_2},S^{i-d_1-1}). \end{array}\right.$$

where $\mathcal{V}\triangleq\left\{P(x_{1,i}|s^{i-d_1})P(x_{2,i}|s^{i-d_1},s_{i-d_2})\right\}_{i=1}^n$. In the same way as we did in the proof of the converse (Section IV, (19)), we can rewrite these equations with the new variable Q, where $Q=i\in\{1,2,\ldots,n\}$ with probability $\frac{1}{n}$. Furthermore, we denote $X_1\triangleq X_{1,Q}$, $X_2\triangleq X_{2,Q}$, $Y\triangleq Y_Q$, $S\triangleq S_Q$, $\tilde{S}_1\triangleq S_{Q-d_1}$, $\tilde{S}_2\triangleq S_{Q-d_2}$, and $U\triangleq(S^{Q-d_1-1},Q)$. Hence we derive that

$$\mathcal{R}_n \subseteq \bigcup_{\mathcal{M}}\left\{\begin{array}{r} R_1 < I(X_1;Y|X_2,S,\tilde{S}_1,\tilde{S}_2,U), \\ R_2 < I(X_2;Y|X_1,S,\tilde{S}_1,\tilde{S}_2,U), \\ R_1+R_2 < I(X_1,X_2;Y|S,\tilde{S}_1,\tilde{S}_2,U) \end{array}\right.$$

where $\mathcal{M}\triangleq\{P(u|\tilde{s}_1)P(x_1|\tilde{s}_1,u)P(x_2|\tilde{s}_1,\tilde{s}_2,u)\}$, which completes the alternative proof of Theorem 1. ∎

VII. EXAMPLES

In this section, we apply the general results of Section III to obtain the capacity region for a finite-state Gaussian MAC, and for the finite-state multiple-access fading channel. We derive optimization problems on the power allocation that maximizes the capacity region for these channels. This power allocation would be the optimal power control policy for maximizing throughput in the presence of delayed channel state information.

A. Capacity Region for a Finite State Additive Gaussian MAC

We now apply Theorem 1 to compute the capacity region of a power-constrained FS additive Gaussian noise (AGN) MAC, and illustrate the effect of the delayed CSI on the capacity region. For a finite state AGN MAC the channel output Y_i at time i, given the channel inputs $X_{1,i}, X_{2,i}$, is given by

$$Y_i = X_{1,i} + X_{2,i} + N_{S_i} \tag{54}$$

where N_{S_i} is a zero-mean Gaussian random variable with variance depending on the state S_i of the channel at time i. In addition to the channel output Y_i the receiver has accesses to the state S_i. The receiver feeds back the CSI to the transmitters through a noiseless feedback channel. The CSI from the receiver is received at transmitter 1 and transmitter 2 after time delays of d_1, d_2 symbol durations, respectively. The state process is assumed to be Markov with steady state distribution $\pi(s)$ and one step transition matrix K. It is clear that the finite state AGN is an FSMC. While the capacity region formula derived in Section III (Theorem 1) was for finite inputs and output alphabets, the result can be generalized to continuous alphabets with inputs constraints. First, we apply only the sum rate formula to explicitly determine the sum rate of the finite state Markov AGN MAC with transmitters power constraints $\mathcal{P}_1$ and $\mathcal{P}_2$.

$$R_1 + R_2 < \max_{p(u|\tilde{s}_1)p(x_1|\tilde{s}_1,u)p(x_2|\tilde{s}_1,\tilde{s}_2,u)} I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U) \tag{55}$$

subject to the power constraints

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_u P(u|\tilde{s}_1) E[X_1^2|\tilde{s}_1, u] \le \mathcal{P}_1,$$
$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_u P(u|\tilde{s}_1) E[X_2^2|\tilde{s}_1, \tilde{s}_2, u] \le \mathcal{P}_2.$$

To compute the maximum sum rate explicitly, we have to first determine the distributions $P(x_1|\tilde{s}_1, u)$ and $P(x_2|\tilde{s}_1, \tilde{s}_2, u)$ for each $\tilde{S}_1$, $\tilde{S}_2$, and U. Suppose $\mathcal{P}_1(\tilde{s}_1, u)$, $\mathcal{P}_2(\tilde{s}_1, \tilde{s}_2, u)$ is the power allocated to states $(\tilde{s}_1, \tilde{s}_2)$ and u. In addition, we denote $h(Z)$ to be the differential entropy of the continuous random variable Z. We can bound the sum rate:

$$\begin{aligned}
&I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U)\\
&= \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2) \sum_u P(u|\tilde{s}_1)\\
&\times I(X_1, X_2; Y|s, \tilde{s}_1, \tilde{s}_2, u)\\
&\overset{(a)}{=} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2) \sum_u P(u|\tilde{s}_1)\\
&\times (h(X_1 + X_2 + N_s|s, \tilde{s}_1, \tilde{s}_2, u) - h(N_s|s))\\
&\overset{(b)}{\le} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2) \sum_u P(u|\tilde{s}_1)\\
&\times \frac{1}{2} \log \left(\frac{E[(X_1 + X_2 + N_s)^2|s, \tilde{s}_1, \tilde{s}_2, u]}{E[N_s^2|s]} \right)\\
&\overset{(c)}{=} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2) \sum_u P(u|\tilde{s}_1)\\
&\times \frac{1}{2} \log \left(1 + \frac{\mathcal{P}_1(\tilde{s}_1, u) + \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2, u)}{\sigma_s^2} \right)\\
&\overset{(d)}{\le} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2)\\
&\times \log \left(1 + \frac{\mathcal{P}_1(\tilde{s}_1) + \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)}{\sigma_s^2} \right)
\end{aligned} \tag{56}$$

where

a) follows from the fact that N_s is independent of $\tilde{S}_1, \tilde{S}_2, U$ given S;
b) follows from the fact that Gaussian distribution has the largest entropy for a given variance;
c) follows from the fact that X_1, X_2 are independent of N_s and independent of each other given $S, \tilde{S}_1, \tilde{S}_2$, and U. Furthermore, we denote $\mathcal{P}_1(\tilde{s}_1) = E[X_1^2|\tilde{s}_1]$, and $\mathcal{P}_2(\tilde{s}_1, \tilde{s}_2, u) = E[X_2^2|\tilde{s}_1, \tilde{s}_2, u]$;
d) follows from Jensen's inequality.

Furthermore, we can achieve (56) if we choose $X_1(\tilde{s}_1, u)$, to be zero-mean Gaussian with variance $\mathcal{P}_1(\tilde{s}_1)$, and $X_2(\tilde{s}_1, \tilde{s}_2, u)$ to be zero-mean Gaussian with variance $\mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)$, both independent of N_s and independent of each other. We now have the following result, for an FSM AGN MAC with average power constraints $\mathcal{P}_1$ and $\mathcal{P}_2$ and CSI at the transmitters with delays d_1 and d_2:

$$\begin{aligned}
&R_1 + R_2\\
&\le \max_{\mathcal{P}_1(\tilde{s}_1), \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \sum_s P(s|\tilde{s}_2)\\
&\times \log \left(1 + \frac{\mathcal{P}_1(\tilde{s}_1) + \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)}{\sigma_s^2} \right)\\
&= \max_{\mathcal{P}_1(\tilde{s}_1), \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1 - d_2}(\tilde{s}_2, \tilde{s}_1)\\
&\times \sum_s K^{d_2}(s, \tilde{s}_2) \log \left(1 + \frac{\mathcal{P}_1(\tilde{s}_1) + \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)}{\sigma_s^2} \right)
\end{aligned} \tag{57}$$

subject to the power constraints:

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \mathcal{P}_1(\tilde{s}_1) \le \mathcal{P}_1, \tag{58}$$

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2) \le \mathcal{P}_2. \tag{59}$$

Similarly, we can derive maximization on R_1 and R_2, for R_1:

$$\begin{aligned}
R_1 \le \max_{\mathcal{P}_1(\tilde{s}_1)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1 - d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2)\\
\times \log \left(1 + \frac{\mathcal{P}_1(\tilde{s}_1)}{\sigma_s^2} \right)
\end{aligned} \tag{60}$$

subject to the power constraint:

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \mathcal{P}_1(\tilde{s}_1) \le \mathcal{P}_1, \tag{61}$$

and for R_2:

$$R_2 \le \max_{\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)} \frac{1}{2}\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2) \times \log\left(1+\frac{\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)}{\sigma_s^2}\right) \quad (62)$$

subject to the power constraint:

$$\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}P(\tilde{s}_2|\tilde{s}_1)\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2) \le \mathcal{P}_2. \quad (63)$$

It is important to mention that in the general case the three equations (57), (60), and (62) do not achieve their maximum in the same distribution, i.e., not in the same power allocation. In the same way we can derive the maximization problem for two special cases. The first case is $d = d_1 = d_2$, since the delays are the same we denote $\tilde{S} = \tilde{S}_1 = \tilde{S}_2$, hence we have,

$$R_1 \le \max_{\mathcal{P}_1(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s})\log\left(1+\frac{\mathcal{P}_1(\tilde{s})}{\sigma_s^2}\right), \quad (64)$$

$$R_2 \le \max_{\mathcal{P}_2(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s})\log\left(1+\frac{\mathcal{P}_2(\tilde{s})}{\sigma_s^2}\right), \quad (65)$$

$$R_1+R_2 \le \max_{\mathcal{P}_1(\tilde{s}),\mathcal{P}_2(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s}) \times\log\left(1+\frac{\mathcal{P}_1(\tilde{s})+\mathcal{P}_2(\tilde{s})}{\sigma_s^2}\right) \quad (66)$$

subject to the power constraints:

$$\sum_{\tilde{s}}\pi(\tilde{s})\mathcal{P}_1(\tilde{s}) \le \mathcal{P}_1, \quad (67)$$

$$\sum_{\tilde{s}}\pi(\tilde{s})\mathcal{P}_2(\tilde{s}) \le \mathcal{P}_2. \quad (68)$$

The second case is $d_2 \le d_1 = \infty$, let us denote $d = d_2$ and $\tilde{S} = \tilde{S}_2$, therefore we have

$$R_1 \le \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s})\log\left(1+\frac{\mathcal{P}_1}{\sigma_s^2}\right) \quad (69)$$

$$R_2 \le \max_{\mathcal{P}_2(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s})\log\left(1+\frac{\mathcal{P}_2(\tilde{s})}{\sigma_s^2}\right) \quad (70)$$

$$R_1+R_2 \le \max_{\mathcal{P}_2(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}}\pi(\tilde{s})\sum_{s}K^{d}(s,\tilde{s}) \times\log\left(1+\frac{\mathcal{P}_1+\mathcal{P}_2(\tilde{s})}{\sigma_s^2}\right) \quad (71)$$

subject to the power constraints:

$$\sum_{\tilde{s}}\pi(\tilde{s})\mathcal{P}_2(\tilde{s}) \le \mathcal{P}_2. \quad (72)$$

Now to gain some intuition on the capacity region, we consider the case when there are only two states. At any given time i the channel is in one of two possible states G or B. In the good state G, the channel is "good" and the noise variance is σ_G^2, and in the bad state B, the channel is "bad" and the noise variance is σ_B^2,

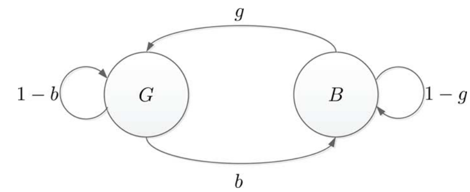

Fig. 5. Two-state AGN channel.

where $\sigma_B^2 > \sigma_G^2$. The state process is specified by the transition probabilities given by

$$P(G|B) = g,$$
$$P(B|G) = b.$$

The state process is illustrated in Fig. 5, the steady state distribution of the Markov chain is given by

$$\pi(G) = \frac{g}{g+b},$$
$$\pi(B) = \frac{b}{b+g}.$$

By solving the optimization problems (57), (66), and (71) for the two state example, we present the maximum sum rate versus delay plot in Figs. 6, 7, and 8 which shows the effect of the CSI delay on the sum rate for $\mathcal{P}_1 = 10, \mathcal{P}_2 = 10, \sigma_G^2 = 1, \sigma_B^2 = 100, g = 0.1, b = 0.1$. The details on solving the optimization problem for the two state example are presented in Appendix D.

Perhaps it seems that the improvement in the sum rate due to CSI is small, however, we should remember that when we encode large blocks, this small improvement in the sum rate can be of great importance. In addition, this improvement in the sum rate due to CSI is for the specific example of two states AGN-MAC. In Figs. 9, 10, and 11 Now, we present the capacity rate region for the two states AGN-MAC in the asymmetrical case $d_1 \ge d_2$ by solving numerically the following optimization problem for different values of α,

$$\max_{R_1,R_2} \alpha R_1 + R_2 \quad (73)$$

subject to the constraints:

$$R_1 \le \frac{1}{2}\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2) \times \log\left(1+\frac{\mathcal{P}_1(\tilde{s}_1)}{\sigma_s^2}\right) \quad (74)$$

$$R_2 \le \frac{1}{2}\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2) \times \log\left(1+\frac{\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)}{\sigma_s^2}\right) \quad (75)$$

$$R_1+R_2 \le \frac{1}{2}\sum_{\tilde{s}_1}\pi(\tilde{s}_1)\sum_{\tilde{s}_2}K^{d_1-d_2}(\tilde{s}_2,\tilde{s}_1)\sum_{s}K^{d_2}(s,\tilde{s}_2) \times \log\left(1+\frac{\mathcal{P}_1(\tilde{s}_1)+\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)}{\sigma_s^2}\right) \quad (76)$$

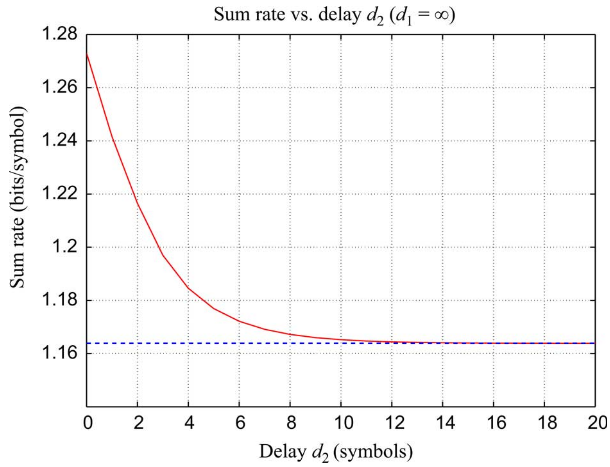

Fig. 6. Sum rate versus delay for the two state channel where transmitter 1 does not have the CSI $d_2 \leq d_1 = \infty$. The dashed line corresponds with the case where CSI is not available at the encoders.

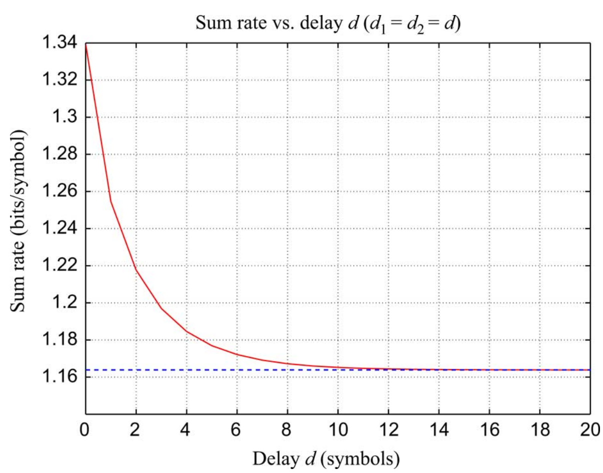

Fig. 7. Sum rate versus delay for the two state channel where $d_1 = d_2$ (symmetrical delay). The dashed line corresponds with the case where CSI is not available at the encoders.

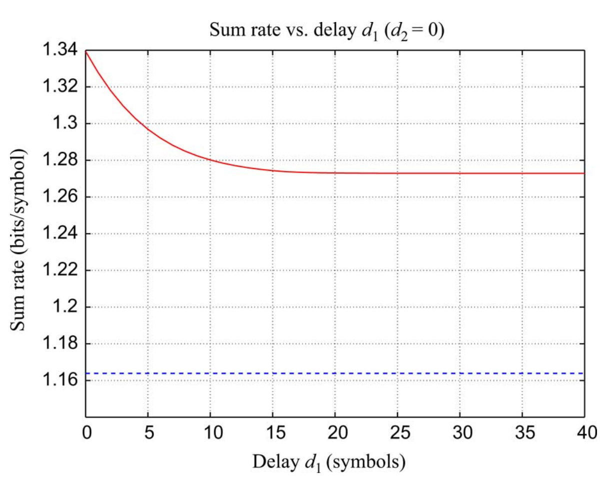

Fig. 8. Sum rate versus delay for the two state channel where $0 = d_2 \leq d_1$ (asymmetrical delay). The dashed line corresponds with the case where CSI is not available at the encoders.

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1)\mathcal{P}_1(\tilde{s}_1) \leq \mathcal{P}_1 \tag{77}$$

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1)\mathcal{P}_2(\tilde{s}_1, \tilde{s}_2) \leq \mathcal{P}_2. \tag{78}$$

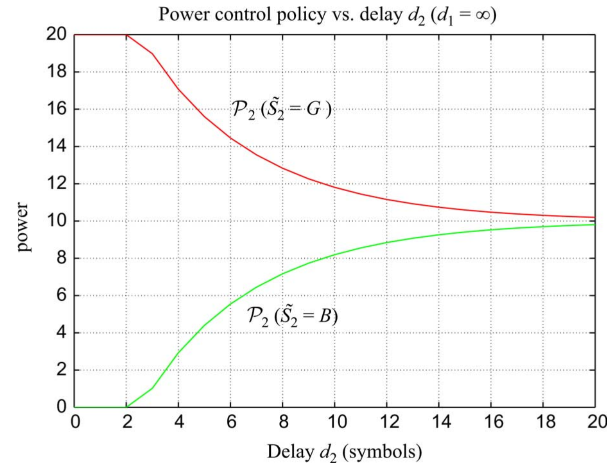

Fig. 9. Power control policy versus delay that achieves the maximum sum rate where $d_2 \leq d_1 = \infty$.

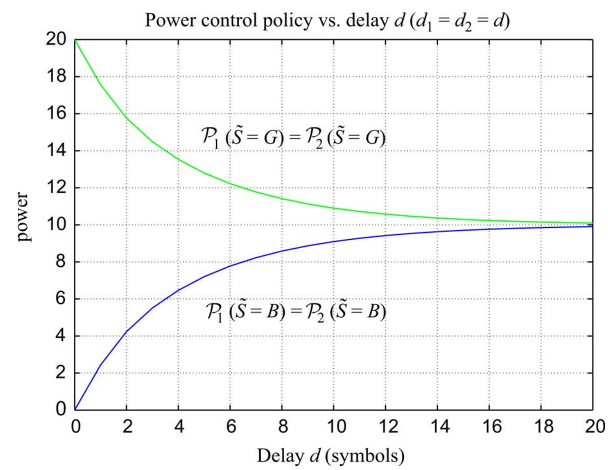

Fig. 10. Power control policy versus delay that achieves the maximum sum rate where $d_1 = d_2$ (symmetrical delay).

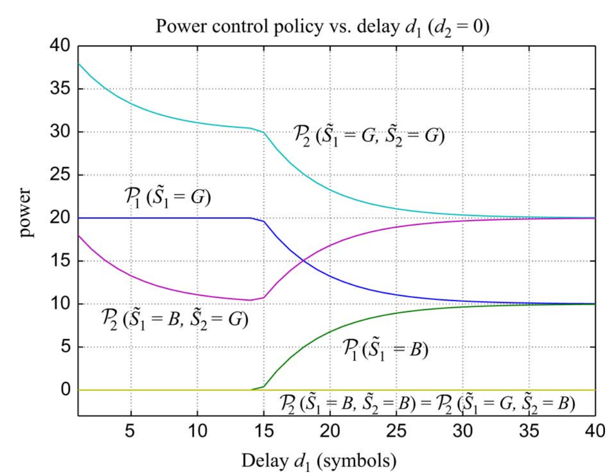

Fig. 11. Power control policy versus delay that achieves the maximum sum rate where $0 = d_2 \leq d_1$ (asymmetrical).

In order to solve the optimization problem (73) we used CVX, a package for specifying and solving convex optimization problems [28]. The capacity rate region for $d_2 = 0$ and different values of d_1 are presented in Fig. 12.

Similarly, we solve the optimization problem for the symmetrical case $d_1 = d_2$, and for the case that transmitter 1 does not

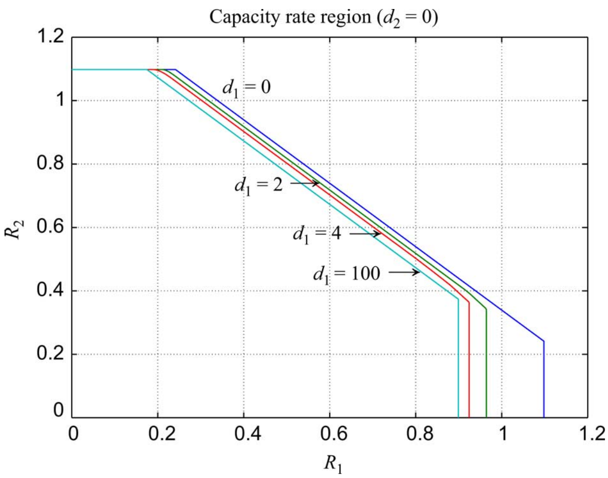

Fig. 12. Capacity rate region for the two states AGN-MAC—asymmetrical case $d_2 = 0$.

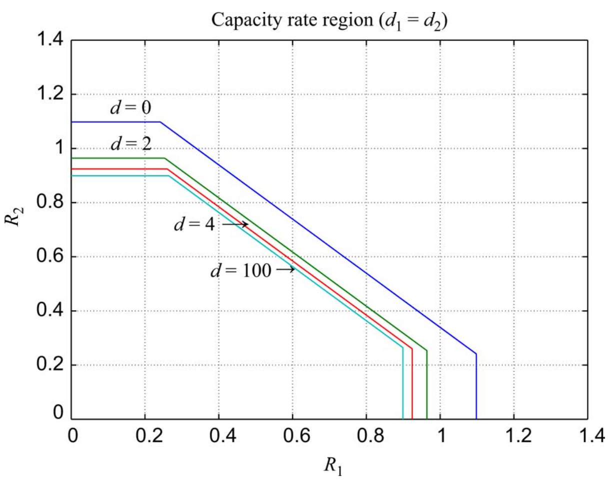

Fig. 13. Capacity rate region for the two states AGN-MAC—symmetrical case $d = d_1 = d_2$.

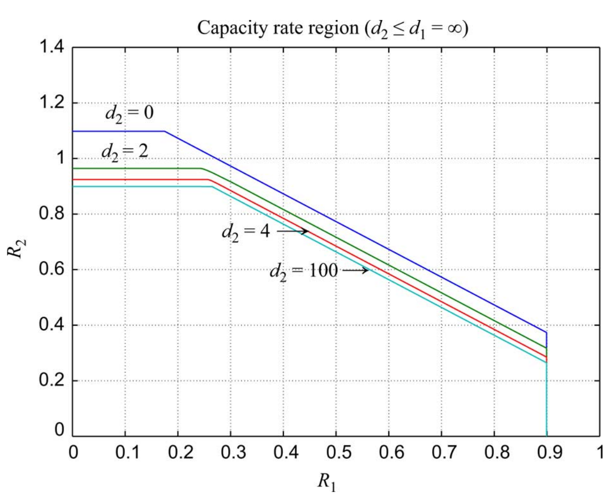

Fig. 14. Capacity rate region for the two states AGN-MAC—Transmitter 1 does not have the CSI $d_2 \leq d_1 = \infty$.

have any CSI, i.e., $d_2 < d_1 = \infty$. The rate regions are illustrated in Figs. 13 and 14, respectively.

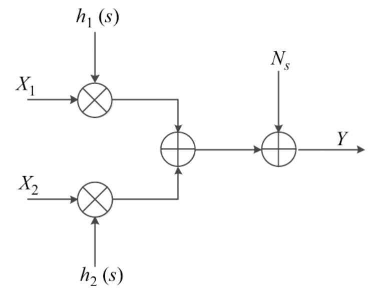

Fig. 15. Fading channel.

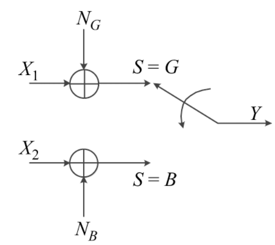

Fig. 16. Channel behaves like a switch, at any given time i the channel is in one of two possible states G or B, where $\sigma_B^2 > \sigma_G^2$. The state process is illustrated in Fig. 5.

B. Capacity Region for a Finite State Multiple-Access Fading Channel

We apply Theorem 1 to compute the capacity region of a power constrained FS Multiple-Access fading channel, and illustrate the effect of the delayed CSI on the capacity region. Consider the discrete-time multiple-access Gaussian channel:

$$Y_i = h_1(s_i)X_{1,i} + h_2(s_i)X_{2,i} + N_{S_i} \tag{79}$$

where $X_{1,i}, X_{2,i}$ are the transmitted waveform, and $h_1(s_i), h_2(s_i)$ are the fading process of the users. The terms $h_1(s_i), h_2(s_i)$ are deterministic functions of s_i. The noise N_{S_i} is a zero-mean Gaussian random variable with variance depending on the state of the channel at time i. Furthermore, the users are subject to the average transmitter power constraints of $\mathcal{P}_1$, and $\mathcal{P}_2$. The state process is assumed to be Markov with steady state distribution $\pi(s)$ and one step transition matrix K, as described in Section II. The FS Multiple-Access fading channel is illustrated in Fig. 15. We apply the capacity region formula to explicitly determine the capacity region of the multiple-access Gaussian fading channel with transmitters power constraints $\mathcal{P}_1$ and $\mathcal{P}_2$. In a similar way to the FSM Additive Gaussian MAC, it can be shown that the capacity achieving distributions are $X_1(\tilde{s}_1, u)$ zero-mean Gaussian with variance $\mathcal{P}_1(\tilde{s}_1)$, and $X_2(\tilde{s}_1, \tilde{s}_2, u)$ zero-mean Gaussian with variance $\mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)$, both independent of N_s and independent of each other. We derive the optimization problem given in (80)–(82) subject to the power constraints (83) and (84), shown at the bottom of the next page. In the same way, we can derive the optimization problem for the symmetrical case $d_1 = d_2$, and for the case that transmitter 1 does not have any CSI, i.e., $d_2 < d_1 = \infty$. Let us solve the optimization problems for the following FSM multiple-access fading channel examples:

Example 1 (AGN Switch Channel): Consider the discrete-time multiple-access Gaussian two state switch channel as described in Fig. 16. We solve the optimization problem:

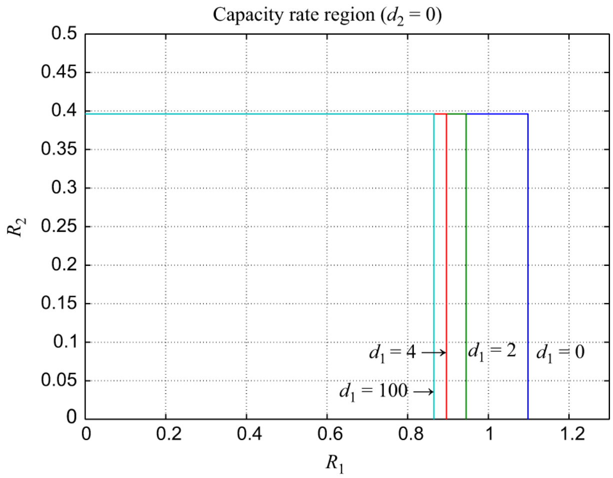

Fig. 17. Capacity rate region for the two states switch channel—asymmetrical case $d_2 = 0$.

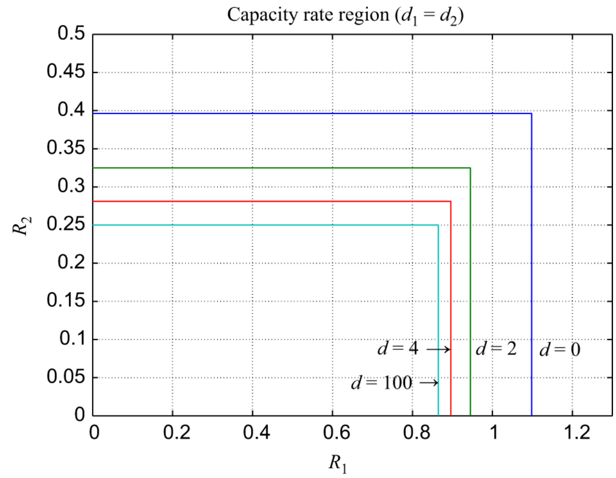

Fig. 18. Capacity rate region for the two states switch channel—symmetrical case $d = d_1 = d_2$.

$\max(\alpha R_1 + R_2)$, for different values of α in the same way we did in the FS additive Gaussian noise (AGN) MAC example. In Figs. 17, 18, and 19 we present the capacity rate region for $\mathcal{P}_1 = 10$, $\mathcal{P}_2 = 10$, $\sigma_G^2 = 1$, $\sigma_B^2 = 10$, $g = 0.1$, $b = 0.1$, $h_1(G) = 1$, $h_1(B) = 0$, $h_2(G) = 0$, $h_2(B) = 1$, in the following cases: asymmetrical, symmetrical, and the case that transmitter 1 does not have any CSI.

As one can see from Figs. 17, 18, and 19 the capacity rate region shape indicates that the users do not interrupt each other, so each of them can transmit at its own maximal rate independently of the other user. This makes perfect sense, since the transmission of each one of them is dependent only on the switch and not on the other's transmission.

Example 2 (Multiple-Access Fading Channel): Consider the power constrained FS Multiple-Access fading channel as illustrated in Fig. 15 with only two states: $S = 1$, $S = 2$. The state process is Markov and illustrated in Fig. 5, with a slight change, instead of denoting the states "good" and "bad" we use $S = 1$, $S = 2$. We solve the optimization problem: $\max(\alpha R_1 + R_2)$, for different values of α in the same way we did before. In Figs. 20, 21, and 22 we present the capacity rate region for $\mathcal{P}_1 = 10$, $\mathcal{P}_2 = 10$, $\sigma_{s=1}^2 = \sigma_{s=2}^2 = 1$, $g = 0.1$, $b = 0.1$, $h_1(s = 1) = 1$, $h_1(s = 2) = 0.5$, $h_2(s = 1) = 0.5$, $h_2(s = 2) = 1$.

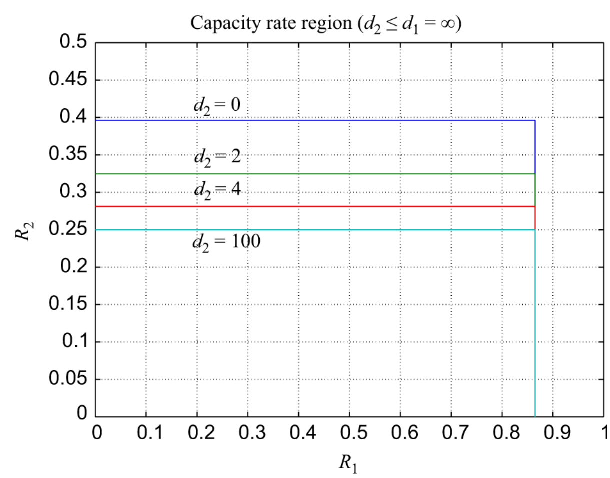

Fig. 19. Capacity rate region for the two states switch channel—Transmitter 1 does not have the CSI $d_2 \leq d_1 = \infty$.

$$R_1 \leq \max_{\mathcal{P}_1(\tilde{s}_1)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \times \log\left(1 + \frac{h_1(s)^2 \mathcal{P}_1(\tilde{s}_1)}{\sigma_s^2}\right) \tag{80}$$

$$R_2 \leq \max_{\mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \log\left(1 + \frac{h_2(s)^2 \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)}{\sigma_s^2}\right) \tag{81}$$

$$R_1 + R_2 \leq \max_{\mathcal{P}_1(\tilde{s}_1), \mathcal{P}_2(\tilde{s}_1,\tilde{s}_2)} \frac{1}{2} \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{s} K^{d_2}(s, \tilde{s}_2) \log\left(1 + \frac{h_1(s)^2 \mathcal{P}_1(\tilde{s}_1) + h_2(s)^2 \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2)}{\sigma_s^2}\right) \tag{82}$$

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \mathcal{P}_1(\tilde{s}_1) \leq \mathcal{P}_1 \tag{83}$$

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} P(\tilde{s}_2|\tilde{s}_1) \mathcal{P}_2(\tilde{s}_1, \tilde{s}_2) \leq \mathcal{P}_2. \tag{84}$$

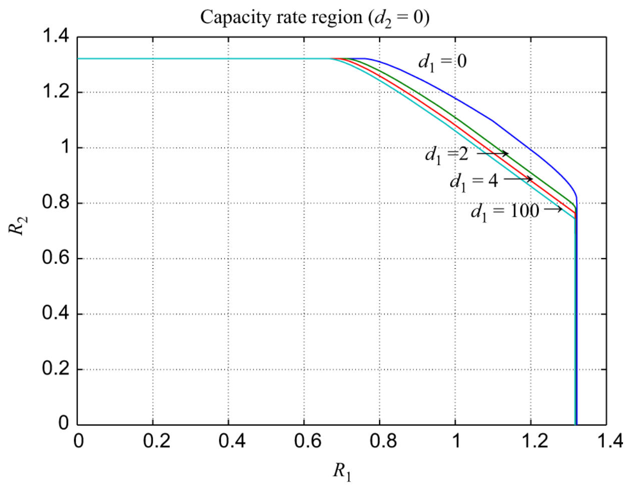

Fig. 20. Capacity rate region for the two states fading channel—asymmetrical case $d_2 = 0$.

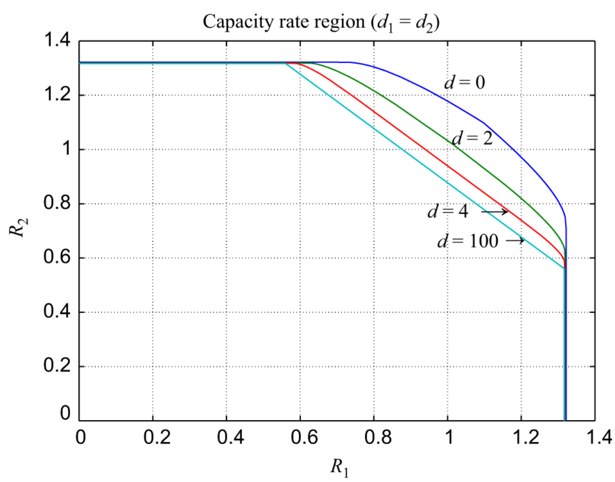

Fig. 21. Capacity rate region for the two states fading channel—symmetrical case $d = d_1 = d_2$.

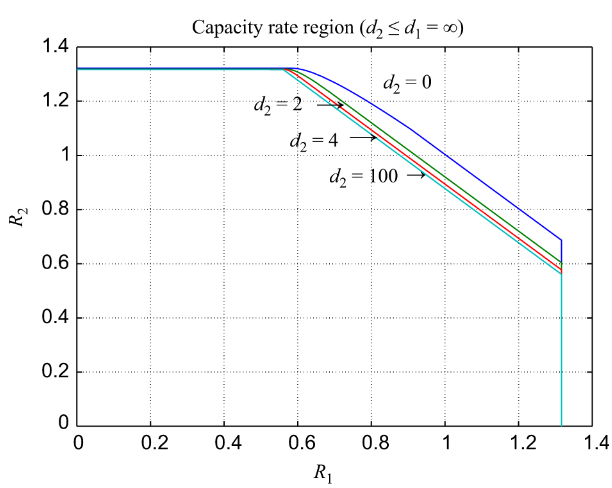

Fig. 22. Capacity rate region for the two states fading channel—Transmitter 1 does not have the CSI $d_2 \leq d_1 = \infty$.

VIII. Summary

The requirement for high rates multi-user communications systems is constantly increasing, so it becomes essential to achieve capacity by deriving the benefit from the channel structure. Motivated by this we studied the problem of finite-state MAC, where the channel state is a Markov process, the transmitters have access to delayed state information, and channel state information is available at the receiver. The delays of the channel state information is assumed to be asymmetric at the transmitters. We obtained a computable characterization of the capacity region for this channel. We provide the outer bound on the capacity region and the proof of the achievability, which is based on multiplexing coding. In addition, we provide alternative proof for the capacity region. The alternative proof is based on a multi-letter expression for the capacity region of FS-MAC with time-invariant feedback. Then we apply the result to derive power control strategies to maximize the capacity region for finite-state additive Gaussian MAC, and for the multiple-access fading channel. The results and the insight in this paper are an intermediate step toward understanding network communication with delayed state information.

Appendix A
Cardinality Bound of the Auxiliary Random Variable U

Let us prove now the cardinality bound for Theorem 1, which is derived directly from the Fenchel–Eggleston–Carathéodry theory [29]. Let us denote the set $\mathcal{Z}$ to be $\mathcal{Z} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{S} \times \tilde{\mathcal{S}}_1 \times \tilde{\mathcal{S}}_2$, let $\mathcal{P}(\mathcal{Z})$ be the set of PMFs on $\mathcal{Z}$, and let $\mathcal{P}(\mathcal{Z}|\mathcal{U}) \subseteq \mathcal{P}(\mathcal{Z})$ be a collection of PMFs $p(z|u)$ on $\mathcal{Z}$ indexed by $u \in \mathcal{U}$. Let $g_j,\ j = 1, \ldots, k$ be continues functions on $\mathcal{P}(\mathcal{Z}|\mathcal{U})$. Then, for any $U \sim F_U(u)$, there exists a finite random variable $U' \sim p(u')$ taking at most k values in $\mathcal{U}$ such that

$$\mathbb{E}\Big[g_j(p_{Z|U}(z|U))\Big] = \int_{\mathcal{U}} g_j(p_{Z|U}(z|u))\mathrm{d}F(u) \tag{85}$$

$$= \sum_{u'} g_j(p_{Z|U}(z|u'))p(u'). \tag{86}$$

Let us denote

$$g_1\big(p(z|u)\big) = I(X_1; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, U = u) \tag{87}$$

$$g_2\big(p(z|u)\big) = I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, U = u) \tag{88}$$

$$g_3\big(p(z|u)\big) = I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2, U = u) \tag{89}$$

then, by using the given technique, we can see that $|\mathcal{U}| \leq 3$. By utilizing the same technique, and similar considerations, we can bound the cardinality of the auxiliary variable in Theorem 2 to be $|\mathcal{U}| \leq 3$ and the cardinality of the auxiliary variable in Theorem 3 to be $|\mathcal{Q}| \leq 3$.

Appendix B
Proof of Theorem 3

The proof of Theorem 3 is similar to the case where the CSI is available at the decoder and asymmetrical delayed CSI is available at the encoders with delays d_1 and d_2 ($d_1 \geq d_2$), only now $d_1 \longrightarrow \infty$. We give here the proof of the converse, and only a brief outline of the achievability proof. Since only encoder 2 has the CSI we denote $d = d_2$ and $\tilde{S} = \tilde{S}_2$.

A) Converse Theorem 3: Given an achievable rate (R_1, R_2) we need to show that there exists joint distribution of the form $P(s,\tilde{s})P(q)P(x_1|q)P(x_2|\tilde{s},q)P(y|x_1,x_2,s)$ such that,

$$\begin{aligned} R_1 &< I(X_1;Y|X_2,S,\tilde{S},Q), \\ R_2 &< I(X_2;Y|X_1,S,\tilde{S},Q), \\ R_1+R_2 &< I(X_1,X_2;Y|S,\tilde{S},Q) \end{aligned}$$

where Q is an random variable with a cardinality bound $|\mathcal{Q}| \leq 3$. The proof of the cardinality bound is similar to the proof in Appendix A. Since (R_1, R_2) is an achievable pair-rate, there exists an $(n, 2^{nR_1}, 2^{nR_2}, d)$ code with a probability of error $P_e^{(n)}$ arbitrarily small. By Fano's inequality:

$$\begin{aligned} H(M_1,M_2|Y^n,S^n) &\leq n(R_1+R_2)P_e^{(n)} + H(P_e^{(n)}) \\ &\triangleq n\varepsilon_n \end{aligned} \tag{90}$$

and it is clear that $\varepsilon_n \to 0$ as $P_e^{(n)} \to \infty$. Then we have

$$H(M_1|Y^n,S^n) \leq H(M_1,M_2|Y^n,S^n) \leq n\varepsilon_n, \tag{91}$$

$$H(M_2|Y^n,S^n) \leq H(M_1,M_2|Y^n,S^n) \leq n\varepsilon_n. \tag{92}$$

We can now bound the rate R_1 as

$$\begin{aligned} nR_1 &= H(M_1) \\ &= H(M_1) + H(M_1|Y^n,S^n) - H(M_1|Y^n,S^n) \\ &\overset{(a)}{\leq} I(M_1;Y^n,S^n) + n\varepsilon_n \\ &\overset{(b)}{=} I(M_1;Y^n|S^n) + I(M_1;S^n) + n\varepsilon_n \\ &\overset{(c)}{=} I(M_1;Y^n|S^n) + n\varepsilon_n \\ &\overset{(d)}{=} I(X_1^n;Y^n|S^n) + n\varepsilon_n \\ &= H(X_1^n|S^n) - H(X_1^n|Y^n,S^n) + n\varepsilon_n \\ &\overset{(e)}{=} H(X_1^n|X_2^n,S^n) - H(X_1^n|Y^n,S^n) + n\varepsilon_n \\ &\overset{(f)}{\leq} H(X_1^n|X_2^n,S^n) - H(X_1^n|Y^n,X_2^n,S^n) + n\varepsilon_n \\ &= I(X_1^n;Y^n|X_2^n,S^n) + n\varepsilon_n \\ &= H(Y^n|X_2^n,S^n) - H(Y^n|X_1^n,X_2^n,S^n) + n\varepsilon_n \\ &= \sum_{i=1}^n \Big\{ H(Y_i|Y^{i-1},X_2^n,S^n) \\ &\qquad - H(Y_i|Y^{i-1},X_1^n,X_2^n,S^n) \Big\} + n\varepsilon_n \\ &\overset{(g)}{\leq} \sum_{i=1}^n \Big\{ H(Y_i|X_{2,i},S_i,S_{i-d}) \\ &\qquad - H(Y_i|Y^{i-1},X_1^n,X_2^n,S^n) \Big\} + n\varepsilon_n \\ &\overset{(h)}{=} \sum_{i=1}^n \Big\{ H(Y_i|X_{2,i},S_i,S_{i-d}) \\ &\qquad - H(Y_i|X_{1,i},X_{2,i},S_i,S_{i-d}) \Big\} + n\varepsilon_n \\ &= \sum_{i=1}^n \{ I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d}) \} + n\varepsilon_n, \end{aligned}$$

where

a) follows from Fano's inequality;
b) follows from chain rule;
c) follows from the fact that M_1 and S^n are independent;
d) follows from the fact that X_1^n is a deterministic function of (M_1, S^n) and the Markov chain $(M_1,S^n) - (X_1^n,S^n) - Y^n$;
e) follows from the fact that X_1^n and M_2 are independent, and the fact that X_2^n is a deterministic function of (M_2, S^n). Therefore, X_1^n and X_2^n are independent given S^n;
f) (f) and (g) follow from the fact that conditioning reduces entropy;
g) follows from the fact that the channel output at time i depends only on the state S_i and the the inputs $X_{1,i}$ and $X_{2,i}$.

Hence, we have

$$R_1 \leq \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d}) + \varepsilon_n. \tag{93}$$

Similarly, we have

$$R_2 \leq \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{2,i}|X_{1,i},S_i,S_{i-d}) + \varepsilon_n \tag{94}$$

and the sum rate:

$$R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i},X_{2,i}|S_i,S_{i-d}) + \varepsilon_n. \tag{95}$$

The expressions in (93)–(95) are the average of the mutual informations calculated at the empirical distribution in column i of the codebook. We can rewrite these equations with the new variable Q, where $Q = i \in \{1,2,\ldots,n\}$ with probability $\frac{1}{n}$. The equations become

$$\begin{aligned} R_1 &\leq \frac{1}{n}\sum_{i=1}^n I(Y_i;X_{1,i}|X_{2,i},S_i,S_{i-d}) + \varepsilon_n \\ &= \frac{1}{n}\sum_{i=1}^n I(Y_Q;X_{1,Q}|X_{2,Q},S_Q,S_{Q-d},Q=i) + \varepsilon_n \\ &= I(Y_Q;X_{1,Q}|X_{2,Q},S_Q,S_{Q-d},Q) + \varepsilon_n. \end{aligned} \tag{96}$$

Now let us denote $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, $Y \triangleq Y_Q$, $S \triangleq S_Q$, and $\tilde{S} \triangleq S_{Q-d}$.

We have

$$\begin{aligned} R_1 &\leq I(X_1;Y|X_2,S,\tilde{S},Q) + \varepsilon_n, \\ R_2 &\leq I(X_2;Y|X_1,S,\tilde{S},Q) + \varepsilon_n, \\ R_1+R_2 &\leq I(X_1,X_2;Y|S,\tilde{S},Q) + \varepsilon_n. \end{aligned}$$

Now we need to show the following Markov relations hold:

1) $P(q|s,\tilde{s}) = P(q)$.
2) $P(x_1|s,\tilde{s},q) = P(x_1|q)$.
3) $P(x_2|x_1,s,\tilde{s},q) = P(x_2|\tilde{s},q)$
4) $P(y|x_1,x_2,s,\tilde{s},q) = P(y|x_1,x_2,s)$.

We prove the above using the following claims:

1) follows from the fact that Q and the state process S^n are independent.
2) follows from the fact that $X_{1,i} = f_{1,i}(M_1)$ and that M_1 and S^n are independent.
3) follows from the fact that M_2 and (M_1, S^n) are independent, and the fact that state process is a Markov chain, hence

$$P(m_2, s^{i-d}|s_i, s_{i-d}, m_1) = P(m_2, s^{i-d}|s_{i-d}).$$

Therefore, we have the Markov chain $(M_2, S^{i-d}) - S_{i-d} - (M_1, S_i)$. Since $X_{1,i} = f_{1,i}(M_1)$ and $X_{2,i} = f_{2,i}(M_2, S^{i-d})$, where $f_{1,i}, f_{2,i}$ are deterministic functions, we get the following Markov chain:

$$X_{2,i} - (M_2, S^{i-d}) - S_{i-d} - (M_1, S_i) - X_{1,i}. \tag{97}$$

Therefore

$$P(x_{2,i}|x_{1,i}, s_i, s_{i-d}) = P(x_{2,i}|s_{i-d}).$$

Since this is true for all i,

$$P(x_{2,q}|x_{1,q}, s_q, s_{q-d}, q) = P(x_{2,q}|s_{q-d}, q).$$

We have $P(x_2|x_1, s, \tilde{s}, q) = P(x_2|\tilde{s}, q)$.

4) follows from the fact that the channel output at time i depends only on the state S_i and the the inputs $X_{1,i}$ and $X_{2,i}$.

Hence, taking the limit as $n \to \infty$, $P_e^{(n)} \to 0$, we have the following converse:

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, S, \tilde{S}, Q), \\ R_2 &\le I(X_2; Y|X_1, S, \tilde{S}, Q), \\ R_1 + R_2 &\le I(X_1, X_2; Y|S, \tilde{S}, Q) \end{aligned}$$

for some choice of joint distribution

$$P(s, \tilde{s})P(q)P(x_1|q)P(x_2|\tilde{s}, q)P(y|x_1, x_2, s)$$

and for some choice of random variable Q defined on $|\mathcal{Q}| \le 3$. This completes the proof of the converse.

A. Achievability Theorem 3

To prove the achievability of the capacity region, we need to show that for a fixed $P(x_1)P(x_2|\tilde{s})$ and (R_1, R_2) that satisfy,

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, S, \tilde{S}), \\ R_2 &\le I(X_2; Y|X_1, S, \tilde{S}), \\ R_1 + R_2 &\le I(X_1, X_2; Y|S, \tilde{S}) \end{aligned}$$

there exists a sequence of $(n, 2^{nR_1}, 2^{nR_2}, d)$ codes where $P_e^{(n)} \to 0$ as $n \to \infty$. Without loss of generality we assume that the finite-state space $\mathcal{S} = \{1, 2, \dots, k\}$, and that the steady state probability $\pi(l) > 0$ for all $l \in \mathcal{S}$.

Encoder 1: construct 2^{nR_1} independent codewords $X_1^n(i)$ where $i \in \{1, 2, .., 2^{nR_1}\}$ of length n, generate each symbol i.i.d., $X_1^n(i) \sim \prod_{l=1}^n P(X_{1,l})$.

Encoder 2: construct k codebooks $\mathcal{C}_2^{\tilde{s}}$ (where the subscript is for Encoder 2) for all $\tilde{S} \in \mathcal{S}$, when in each codebook $\mathcal{C}_2^{\tilde{s}}$ there are $2^{n_2(\tilde{s})R_2(\tilde{s})}$ codewords, where $n_2(\tilde{s}) = (P(\tilde{S} = \tilde{s}) - \epsilon')n$, for $\epsilon' > 0$. Every codeword $\mathcal{C}_2^{\tilde{s}}(i)$ where $i \in \{1, 2, \dots, 2^{n_2(\tilde{s})R_2(\tilde{s})}\}$ has a length of $n_2(\tilde{s})$ symbols. Each codeword from the $\mathcal{C}_2^{\tilde{s}}$ codebook is built $X_2^{\tilde{s}} \sim$ i.i.d. $P(x_2^{\tilde{s}}|\tilde{S} = \tilde{s})$ (where the subscript is for Encoder 2). A message M_2 is chosen according to a uniform distribution $\Pr(M_2 = m_2) = 2^{-nR_2}$, $m_2 \in \{1, 2, \dots, 2^{nR_2}\}$. Every message m_2 is mapped into k sub messages $\mathcal{V}_2(m_2) = \{V_2^1(m_2), V_2^2(m_2), \dots, V_2^k(m_2)\}$ (one message from each codebook). Hence, every message m_2 is specified by a k dimensional vector. For a fix block length n, let $N_{\tilde{s}}$ be the number of times during the n symbols for which the feedback information at encoder 2 regarding the channel state is $\tilde{S} = \tilde{s}$. Every time that the delayed CSI is $\tilde{S} = \tilde{s}$, encoder 2 sends the next symbol from $\mathcal{C}_2^{\tilde{s}}$ codebook. Since $N_{\tilde{s}}$ is not necessarily equivalent to $n_2(\tilde{s})$, an error is declared if $N_{\tilde{s}} < n_2(\tilde{s})$, and the code is zero-filled if $N_{\tilde{s}} > n_2(\tilde{s})$. Therefore we can send total of $2^{nR_2} = 2^{\sum_{\tilde{s} \in \mathcal{S}} n_2(\tilde{s})R_2(\tilde{s})}$ messages.

Decoding: we use successive decoding, similar to the decoding in Section V. It can be shown that the probability of error, conditioned on a particular codeword being sent, goes to zero if the conditions of the following are met:

$$\begin{aligned} R_1 &\le I(X_1; Y|X_2, S, \tilde{S}), \\ R_2 &\le I(X_2; Y|X_1, S, \tilde{S}), \\ R_1 + R_2 &\le I(X_1, X_2; Y|S, \tilde{S}). \end{aligned}$$

The above bound shows that the average probability of error, which by symmetry is equal to the probability for an individual pair of codewords (m_1, m_2), averaged over all choices of codebooks in the random code construction, is arbitrarily small. Hence there exists at least one $(n, 2^{nR_1}, 2^{nR_2}, d)$ code with an arbitrarily small probability of error. To complete the proof we use time-sharing to allow any (R_1, R_2) in the convex hull to be achieved.

Appendix C
Three Users Multiple-Access Channels With Delayed State Information

In this section, we generalize the result derived for two senders to three senders. The FSM-MAC with Delayed CSI in this case is shown in Fig. 23. We send independent messages M_1, M_2 and M_3 over the channel from the senders. The codes, rates, and achievability are all defined in exactly the same way as in the two-sender case. In addition, without loss of generality, we assume that $d_1 \ge d_2 \ge d_3$.

Theorem 6 (Capacity Region of 3-User FSM-MAC With Delayed CSI $d_1 \ge D_2 \ge D_3$): The capacity region of 3-user FSM-MAC with CSI at the decoder and asymmetrical delayed CSI at the encoders with delays d_1, d_2 and d_3, as shown in

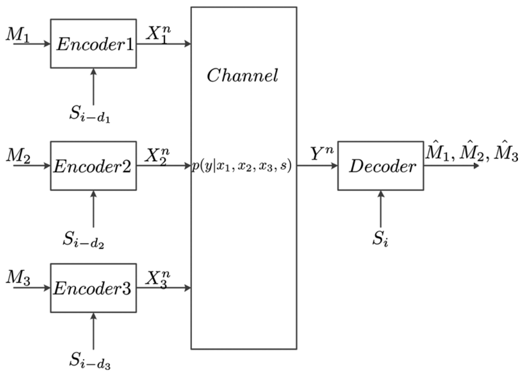

Fig. 23. Three users FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1, d_2, d_3.

Fig. 23 is given in (98), shown at the bottom of the page, where

$$W = P(u_1|\tilde{s}_1)P(u_2|\tilde{s}_1,\tilde{s}_2)P(x_1|\tilde{s}_1,u_1)P(x_2|\tilde{s}_1,\tilde{s}_2,u_1,u_2) \times P(x_3|\tilde{s}_1,\tilde{s}_2,\tilde{s}_3,u_1,u_2) \quad (99)$$

and U_1, U_2 are auxiliary random variables with cardinality $|\mathcal{U}_1| = |\mathcal{U}_2| \le 7$.

The proof of Theorem 6 contains similar ideas to the proof for the two user case. In the converse proof, instead of three inequalities, there are seven inequalities. For example, we can bound the rate R_1 as

$$R_1 \le \frac{1}{n}\sum_{i=1}^{n} I(Y_i; X_{1,i}|X_{2,i}, X_{3,i}, S_i, S_{i-d_1}, S_{i-d_2}, S_{i-d_3}, S^{i-d_1-1}, S_{i-d_1+1}^{i-d_2-1}) + \varepsilon_n.$$

We can rewrite this equation with the new variable Q, where $Q = i \in \{1, 2, \ldots, n\}$ with probability $\frac{1}{n}$. Hence

$$R_1 \le I(Y_Q; X_{1,Q}|X_{2,Q}, X_{3,Q}, S_Q, S_{Q-d_1}, S_{Q-d_2}, S_{Q-d_3}, S^{Q-d_1-1}, S_{Q-d_1+1}^{Q-d_2-1}) + \varepsilon_n.$$

Now let us denote $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, $X_3 \triangleq X_{3,Q}$, $Y \triangleq Y_Q$, $S \triangleq S_Q$, $\tilde{S}_1 \triangleq S_{Q-d_1}$, $\tilde{S}_2 \triangleq S_{Q-d_2}$, $\tilde{S}_3 \triangleq S_{Q-d_3}$, $U_1 \triangleq (S^{Q-d_1-1}, Q)$, and $U_2 \triangleq (S_{Q-d_1+1}^{Q-d_2-1}, Q)$. Therefore

$$R_1 \le I(X_1; Y|X_2, X_3, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2).$$

The other inequalities are derived in a similar way.

In the achievability proof, we use multiplexing coding and successive decoding. The multiplexers of encoder 1, 2, and 3 are controlled by the delayed CSI $(\tilde{S}_1)$, $(\tilde{S}_1, \tilde{S}_2)$, and $(\tilde{S}_1, \tilde{S}_2, \tilde{S}_3)$, respectively. In successive decoding scheme, instead of decoding the three messages simultaneously, the decoder first decodes one of the messages by itself, where the other users messages are considered as noise. After decoding the first user's message, the decoder turns to decode the second message. When decoding the second message, the decoder uses the information about the first message as side information, etc. This decoding rule aims to achieve the six corner points of the rate region, for example, one of the corner points is

$$\begin{aligned} R_1 &< I(X_1; Y|S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_2 &< I(X_2; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_3 &< I(X_3; Y|X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2) \end{aligned} \quad (100)$$

The analysis of the probability of error, is done in similar way to the two-user case. It is worth noting that from the three users case, it is quite simple to see how to extend the result to m-users FSM-MAC.

Appendix D
Determination of the Two-State MAC Capacity Region

For simplicity we give here the solution to the constrained optimization only for the symmetrical case, i.e., both CSI delays are the same $(d_1 = d_2)$, the solution of the other cases are obtained in a similar way. The optimization problem is

$$R_1 + R_2 = \max_{\mathcal{P}_1(\tilde{s}), \mathcal{P}_2(\tilde{s})} \frac{1}{2}\sum_{\tilde{s}} \pi(\tilde{s}) \sum_{s} K^d(s, \tilde{s}) \times \log\left(1 + \frac{\mathcal{P}_1(\tilde{s}) + \mathcal{P}_2(\tilde{s})}{\sigma_s^2}\right) \quad (101)$$

subject to the power constraints:

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_1(\tilde{s}) \le \mathcal{P}_1 \quad (102)$$

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_2(\tilde{s}) \le \mathcal{P}_2 \quad (103)$$

$$\mathcal{P}_1(\tilde{s}) \ge 0 \ \forall \tilde{s} \quad (104)$$

$$\mathcal{P}_2(\tilde{s}) \ge 0 \ \forall \tilde{s}. \quad (105)$$

$$\mathcal{R} = \bigcup_{W} \begin{pmatrix} R_1 < I(X_1; Y|X_2, X_3, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_2 < I(X_2; Y|X_1, X_3, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_3 < I(X_3; Y|X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_1 + R_2 < I(X_1, X_2; Y|X_3, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_1 + R_3 < I(X_1, X_3; Y|X_2, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_2 + R_3 < I(X_2, X_3; Y|X_1, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \\ R_1 + R_2 + R_3 < I(X_1, X_2; Y|X_3, S, \tilde{S}_1, \tilde{S}_2, \tilde{S}_3, U_1, U_2), \end{pmatrix} \quad (98)$$

The solution can be obtained by the Lagrange multiplier method. Since the objective function is monotonically increasing with respect to $\mathcal{P}_1$, and $\mathcal{P}_2$, it follows that the maximum is achieved when

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_1(\tilde{s}) = \mathcal{P}_1 \tag{106}$$

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_2(\tilde{s}) = \mathcal{P}_2. \tag{107}$$

Since $\log$ is a concave function, and $\pi(\tilde{s}), K^d(s,\tilde{s}) \geq 0$. We get that objective function is concave in both variables $\mathcal{P}_1(\tilde{s})$, and $\mathcal{P}_2(\tilde{s})$. Also the constraints functions (106), and (107) are affine. So we can use the Kuhn–Tucker conditions [30, Ch. 5.3.3] as sufficient conditions to solve the optimization problem. Application of the Kuhn–Tucker conditions gives the following conditions of optimality:

$$\frac{1}{2}\sum_{s}\frac{K^d(s,\tilde{s}_i)}{\sigma_s^2+\mathcal{P}_1^*(\tilde{s}_i)+\mathcal{P}_2^*(\tilde{s}_i)} \leq \nu_1, \forall \tilde{s}_i \in \{s_1,..,s_k\} \tag{108}$$

$$\frac{1}{2}\sum_{s}\frac{K^d(s,\tilde{s}_i)}{\sigma_s^2+\mathcal{P}_1^*(\tilde{s}_i)+\mathcal{P}_2^*(\tilde{s}_i)} \leq \nu_2, \forall \tilde{s}_i \in \{s_1,..,s_k\} \tag{109}$$

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_1^*(\tilde{s}) = \mathcal{P}_1 \tag{110}$$

$$\sum_{\tilde{s}} \pi(\tilde{s})\mathcal{P}_2^*(\tilde{s}) = \mathcal{P}_2 \tag{111}$$

with equality in (108) whenever $\mathcal{P}_1^*(\tilde{s}_i) \geq 0$, and equality in (109) whenever $\mathcal{P}_2^*(\tilde{s}_i) \geq 0$. For the two state Gaussian MAC example in Section VII-A we have

$$K^d = \begin{bmatrix} 1-\frac{g}{g+b}\left(1-(1-g-b)^d\right) & \frac{g}{g+b}(1-(1-g-b)^d) \\ \frac{b}{b+g}(1-(1-g-b)^d) & 1-\frac{b}{b+g}(1-(1-g-b)^d) \end{bmatrix}.$$

Now the solution to the constrained optimization problem is obtained by finding $\mathcal{P}_1^*(\tilde{s}_i)$, and $\mathcal{P}_2^*(\tilde{s}_i)$ that satisfy the Kuhn-Tucker conditions. For simplicity, in order to solve the optimization problem we used CVX, a package for specifying and solving convex optimization problems [28].

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and anonymous referees for their comments, which significantly helped to improve the content and the organization of the paper. The authors gratefully acknowledge B. Zaidel for very helpful discussions.

REFERENCES

[1] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, pp. 868–886, 1996.

[2] H. S. Wang and N. Moayeri, "Finite-state markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol*, vol. 44, pp. 163–171, 1995.

[3] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop*, vol. 2, no. 4, pp. 289–293, 1958.

[4] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Theory*, vol. 9, pp. 19–31, 1980.

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

[6] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, 1999.

[7] 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.

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

[9] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, pp. 445–586, Jun. 2008.

[10] G. Como and S. Yüksel, "On the capacity of finite state multiple access channels with asymmetric partial state feedback," in *WiOPT'09: Proc. 7th Int. Conf. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, IEEE Press*, Piscataway, NJ, 2009, pp. 589–594.

[11] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, 2002.

[12] H. H. Permuter and T. Weissman, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, 2009.

[13] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *ISIT 2010*, Austin, TX, 2010.

[14] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," in *2010 48th Annual Allerton Conf. Communication, Control, and Computing (Allerton)*, 1, 2010, pp. 1188–1195.

[15] M. A. Maddah-Ali and D. Tse, On the Degrees of Freedom of MISO Broadcast Channels With Delayed Feedback EECS Dept., Univ. California, Berkeley, 2010, Technical Report UCB/EECS-2010-122.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley-Interscience, 1991.

[17] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Applications (ISITA-90)*, Nov. 1990, pp. 303–305.

[18] Y. H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory.*, vol. 25, pp. 1488–1499, Apr. 2008.

[19] S. C. Tatikonda, "Control Under Communication Constraints," Ph. D. dissertation, Elect. Eng. Comput. Sci., Mass. Inst. Technol., Cambridge, MA, 2000.

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

[21] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, pp. 644–662, Feb. 2009.

[22] H. H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, 2009.

[23] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3629–3644, Aug. 2009.

[24] G. Kramer, "Directed Information for Channels With Feedback," Ph. D. thesis, Elect. Eng., Swiss Federal Inst. Technol. (ETH) Zurich, , 1998.

[25] S. Pradhan and R. Venkataramanan, "Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2154–2179, 2007.

[26] R. Zamir, Y. Kochman, and U. Erez, "Achieving the gaussian rate-distortion function by prediction," *IEEE Trans. Inf. Theory*, vol. 54, pp. 3354–3364, 2008.

[27] R. Dabora and A. J. Goldsmith, "Capacity theorems for discrete, finite-state broadcast channels with feedback and unidirectional receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5958–5983, 2010.

[28] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, Version 1.21 2010.

[29] H. G. Eggleston, *Convexity*. New York: Cambridge Univ. Press, 1958.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. New-York: Cambridge Univ. Press, 2004.

Uria Basher received his B.Sc. (*summa cum laude*) degree in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2010. He is studying for the M.Sc. degree on the fast track program for honor students in Electrical and Computer Engineering from the Ben-Gurion University, Israel.

Between 2003 and 2006, he was an instructor of advanced electronics systems at the Israeli Defense Forces.

Mr. Basher is a recipient of the Dean's List Award, and excellence scholarships from the Electrical and Computer Engineering Department, Ben-Gurion University, Israel.

Avihay Shirazi received his B.Sc. (*summa cum laude*) degree in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2010. He is studying for the M.Sc. degree on the fast track program for honor students in Electrical and Computer Engineering from the Ben-Gurion University, Israel.

Between 2001 and 2005, he was an officer at the Israeli Defense Forces.

Avihay is a recipient of the Dean's List Award, and excellence scholarships from the Electrical and Computer Engineering Department, Ben-Gurion University, Israel.

Haim H. Permuter (M'08) received his B.Sc. (*summa cum laude*) and M.Sc. (summa cum laude) degree in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and Ph.D. degrees in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. He is currently a senior lecturer at Ben-Gurion University.

Dr. Permuter is a recipient of the Fulbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, Marie Curie Reintegration Fellowship, and the 2009 U.S.-Israel Binational Science Foundation Bergmann Memorial Award.