POLITECNICO DI TORINO
Repository ISTITUZIONALE

Memoryless multiple access channel with asymmetric noisy state information at the encoders

Original
Memoryless multiple access channel with asymmetric noisy state information at the encoders / en, Nevroz; Alajaji, Fady; Yüksel, Serdar; Como, Giacomo. - In: IEEE TRANSACTIONS ON INFORMATION THEORY. - ISSN 0018-9448. - 59:11(2013), pp. 7052-7070. [10.1109/TIT.2013.2276612]

Availability:
This version is available at: 11583/2624526 since: 2015-11-30T18:36:52Z

Publisher:
IEEE

Published
DOI:10.1109/TIT.2013.2276612

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

Memoryless Multiple Access Channel with Asymmetric Noisy State Information at the Encoders

Nevroz Şen, *Student Member, IEEE,* Fady Alajaji, *Senior Member, IEEE,* Serdar Yüksel, *Member, IEEE* and Giacomo Como, *Member, IEEE*

***Abstract*—The problem of reliable communication over the memoryless state-dependent multiple-access channel (MAC) is considered where the encoders and the decoder are provided with various degrees of asymmetric noisy channel state information (CSI). For the case where the encoders observe causal, asymmetric noisy CSI and the decoder observes complete CSI, inner and outer bounds to the capacity region, which are tight for the sum-rate capacity, are provided. Next, single-letter characterizations for the channel capacity regions under each of the following system settings are established: (a) the CSI at the encoders are asymmetric deterministic functions of the CSI at the decoder and the encoders have non-causal noisy CSI; (b) the encoders observe asymmetric noisy CSI with asymmetric delays and the decoder observes complete CSI; (c) a degraded message set scenario with asymmetric noisy CSI at the encoders and complete and/or noisy CSI at the decoder. The main component in these results is a generalization of a recently introduced converse coding approach for the MAC with asymmetric quantized CSI at the encoders and herein considerably extended and adapted for the noisy CSI setup.**

***Index Terms*—Asymmetric channel state information, capacity region, converse coding theorem, Shannon strategies, State-dependent multiple-access channel**

I. Introduction

A. Literature Review

Modeling communication channels with a state process, which governs the channel behavior, fits well for many physical scenarios. For single-user channels, the characterization of the capacity with various degrees of channel state information at the transmitter (CSIT) and at the receiver (CSIR) is well understood. Among them, Shannon [1] provides the capacity formula for a discrete memoryless channel with causal noiseless CSIT, where the state process is independent and identically distributed (i.i.d.), in terms of Shannon strategies (random functions from the state space to the channel input space). In [2] Gel'fand and Pinsker consider the same problem with non-causal side information and establish a single-letter capacity formula. In [3], noisy state observation available at both the transmitter and the receiver is considered and the capacity under such a setting is derived. Later, in [4] this result is shown to be a special case of Shannon's model and the authors also prove that when CSIT is a deterministic function of CSIR, optimal codes can be constructed directly on the input alphabet. In [5], the authors examine the discrete modulo-additive noise channel with causal CSIT which governs the noise distribution, and they determine the optimal strategies that achieve channel capacity. In [6], fading channels with perfect channel state information at the transmitter is considered and it is shown that with instantaneous and perfect CSI, the transmitter can adjust the data rates for each channel state to maximize the average transmission rate. In [7], a single letter characterization of the capacity region for single-user finite-state Markovian channels with quantized state information available at the transmitter and full state information at the decoder is provided. In a closely related direction, finite-state channels (with memory) with output feedback is investigated in [8]. In particular, [8] shows that it is possible to formulate the computation of feedback capacity as a stochastic control problem. In [9], finite-state channels with feedback, where feedback is a time-invariant deterministic function of the output samples, is considered.

The literature on state dependent multiple access channels with different assumptions of CSIR and CSIT (such as causal vs non-causal, perfect vs imperfect) is extensive and the main contributions of the current paper have several interactions with the available results in the literature, which we present in Subsection I-B. Hence, we believe that in order to suitably highlight the contributions of this paper, it is worth to discuss the relevant literature for the multi-user setting in more detail. To start, [10] provides a multi-letter characterization of the capacity region of time-varying MACs with general channel statistics (with/without memory) under a general state process (not necessarily stationary or ergodic) and with various degrees of CSIT and CSIR. In [10], it is also shown that when the channel is memoryless, if the encoders use only the past k asymmetric partial (but not noisy) CSI and the decoder has complete CSI, then it is possible to simplify the multi-letter characterization to a single letter one [10, Theorem 4]. In [11], a general framework for the capacity region of MACs with causal and non-causal CSI is presented. In particular, an achievable rate region is presented for the memoryless state-dependent MAC with correlated CSI and the sum-rate capacity is established under the condition that the state information available to each encoder are independent. In [12], MACs with complete CSIR and noncausal, partial, rate limited CSITs are considered. In particular, for the degraded case, i.e., the case where the CSI available at one of the encoders is a subset of the CSI available at the other encoder, a single

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

The material in this paper was presented in part at the Forty-Ninth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2011 and at the IEEE International Symposium on Information Theory, Cambridge, MA, July 2012..

N. Şen, F. Alajaji and S. Yüksel are with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (email:{nsen,fady,yuksel}@mast.queensu.ca).

G. Como is with the Dept. of Automatic Control, Lund University, SE-221 00 Lund, Sweden (email:giacomo@control.lth.se).

letter formula for the capacity region is provided and when the CSITs are not degraded, inner and outer bounds are derived, see [12, Theorems 1, 2]. In [13] state-dependent MAC in which transmitters observe asymmetric partial quantized CSI causally, and the receiver has full CSI is considered and a single letter characterization of the capacity region is obtained. In [14], memoryless state-dependent MACs with two independent states (see also [15] for the common state), each known causally and strictly causally to one encoder, is considered and an achievable rate region, which is shown to contain an achievable region where each user applies Shannon strategies, is proposed. In [16], another achievable rate region for the same problem is proposed and in [17] it is shown that this region can be strictly larger than the one proposed in [14]. In [14], it is also shown that strictly causal CSI does not increase the sum-rate capacity. In [18], the finite-state Markovian MAC with asymmetric delayed CSITs is studied and its capacity region is determined. In [19], the capacity region of some multiple-user channels with causal CSI is established and inner and outer capacity bounds are provided for the MAC. Another active research direction on the state-dependent MAC regards the so-called cooperative state-dependent MAC where there exists a degraded condition on the message sets. In particular, [20] and [21] characterize the capacity region of the cooperative state-dependent MAC with states non-causally and causally available at the transmitters. For more recent results on the cooperative state-dependent MAC problem see references [22], [23], [24] and [25]. Finally, for a comprehensive survey on channel coding with side information see [26] and for other recent results on the multi-user channels with side information see [27], [28], [29], [30] and [31].

B. Main Contributions and Connections with the Literature

We consider several scenarios where the encoders and the decoder observe various degrees of noisy CSI. The essential requirement we impose is that the noisy CSI available to the decision makers is realized via the corruption of CSI by different noise processes, which give a realistic physical structure of the communication setup. We herein note that the asymmetric noisy CSI assumption is acceptable as typically the feedback links are imperfect and sufficiently far from each other so that the information carried through them is corrupted by different (independent) noise processes. It should also be noted that asymmetric side information has many applications in different multi-user models. Finally, what makes (asymmetric) noisy setups particularly interesting are the facts that

(i) No transmitter CSI contains the CSI available to the other one;

(ii) CSI available to the decoder does not contain any of the CSI available to the two encoders.

When existing results, which provide a single letter capacity formulation, are examined, it can be observed that most of them do not satisfy (i) or (ii) or both (e.g., [13], [10], [11], [12], [18]). Nonetheless, among these, [10] discusses the situation with noisy CSI and makes the observation that the situation where the CSITs and CSIR are noisy versions of the state S_t can be accommodated by their models. However, they also note that if the noises corrupting transmitters and receiver CSI are different, then the encoder CSI will, in general, not be contained in the decoder CSI. Hence, motivated by similar observations in the literature (e.g., [11]), we partially treat the scenarios below and provide inner and outer bounds, which are tight for the sum-rate capacity, for scenario (1) below and provide a single-letter characterization for the capacity region of the latter scenarios:

(1) The state-dependent MAC in which each of the transmitters has an asymmetric causal noisy CSI and the receiver has complete CSI (Theorems 2.1, 2.2 and Corollary 2.1).

(2) The state-dependent MAC in which each of the transmitters has an asymmetric non-causal noisy CSIT which is a deterministic function of the CSIR at the receiver (Theorem 2.3).

(3) The state-dependent MAC in which each of the transmitters has an asymmetrically delayed and asymmetric noisy CSI and the receiver has complete CSI (Theorem 2.4).

(4) The state-dependent MAC with degraded message set where both transmitters transmit a common message and one transmitter (informed transmitter) transmits a private message. The informed transmitter has causal noisy CSI, the other encoder has a delayed noisy CSI and the receiver has various degrees of CSI (Theorems 2.5 and 2.6).

Let us now briefly position these contributions with respect to the available results in the literature. The sum-rate capacity determined in (1) can be thought as an extension of [11, Theorem 4] to the case where the encoders have correlated CSI. The causal setup of (2) is solved in [13]. The solution that we provide to the non-causal case partially solves [12] and extends [11, Theorem 5] to the case where the encoders have correlated CSI. Furthermore, since the causal and non-causal capacities are identical for scenario (2), the causal solution can be considered as an extension of [4, Proposition 1] to a noisy multi-user case. Finally, (4) is an extension of [20, Theorem 4] to a noisy setup.

C. The Converse Coding Approach

The most relevant paper to this work is [13] which provides a converse coding approach for the state-dependent MAC where asymmetric partial state information available at the encoders. In this work, we adopt and expand on the converse technique of this paper and use it in a noisy setup. The converse coding approach of [13] is based on team decision theoretic methods [32] (see also [33], [34] and [35] for recent team decision and control theoretic approaches) where the authors use *memoryless stationary team policies* which play a key role in showing that the past information is irrelevant. As the authors mention in [13, Remark 2], for the validity of their arguments, it would suffice that the state information available at the decoder contains the one available at the two transmitters. In this way, the decoder does not need to estimate the coding policies used in decentralized time-sharing.

For the noisy setup, we need to modify this approach to account for the fact that the decoder does not have access

to the state information at the encoders, and that the past state information does not lead to a tractable recursion. This difficulty is overcome by showing that a product form on the team policies exists in the noisy setup as well.

The rest of the paper is organized as follows. In Section II, we formally state scenarios (1)-(4), and present the main results and several observations. In Section III, we provide two examples in one of which we apply the result of [5] and get the full capacity region by only considering the tightness of the sum-rate capacity. Finally, in Section IV, we present concluding remarks.

Throughout the paper, we will use the following notations. A random variable will be denoted by an upper case letter X and its particular realization by a lower case letter x. For a vector v, and a positive integer i, v_i will denote the i-th entry of v, while $v_{[i]} = (v_1, \cdots, v_i)$ will denote the vector of the first i entries and $v_{[i,j]} = (v_i, \cdots, v_j)$, $i \leq j$ will denote the vector of entries between i, j of v. For a finite set $\mathcal{A}$, $\mathcal{P}(\mathcal{A})$ will denote the simplex of probability distributions over $\mathcal{A}$. Probability distributions are denoted by $P(\cdot)$ and subscripted by the name of the random variables and conditioning, e.g., $P_{U,T|V,S}(u,t|v,s)$ is the conditional probability of $(U = u, T = t)$ given $(V = v, S = s)$. Finally, for a positive integer n, we shall denote by $\mathcal{A}^{(n)} := \bigcup_{0<s<n} \mathcal{A}^s$ the set of $\mathcal{A}$-strings of length smaller than n. We denote the indicator function of an event E by $1_{\{E\}}$. All sets considered hereafter are finite.

II. Main Results

Consider a two-user memoryless state-dependent MAC, with two encoders, a, b, and two independent message sources W_a and W_b which are uniformly distributed in the finite sets $\mathcal{W}_a$ and $\mathcal{W}_b$, respectively. The channel inputs from the encoders are $X^a \in \mathcal{X}_a$ and $X^b \in \mathcal{X}_b$, respectively, and the channel output is $Y \in \mathcal{Y}$. The channel state process is modeled as a sequence $\{S_t\}_{t=1}^{\infty}$ of i.i.d. random variables in some finite space $\mathcal{S}$. Let (S_t^a, S_t^b) denote a pair of random variables available at two encoders, a, b, respectively, at time t. Throughout the paper, by symmetric side information we will refer to the case where $S_t^a = S_t^b$, $\forall t$ and by asymmetric side information to when this does not occur. Furthermore, by noisy side information will refer to the case where (S_t^a, S_t^b, S_t) are correlated according to a given joint distribution $P_{S^a,S^b,S}(s^a, s^b, s)$.

A. Asymmetric Causal Noisy CSIT

Let the two encoders have access to a causal noisy version of the state information S_t at each time $t \geq 1$, modeled by $S_t^a \in \mathcal{S}_a$, $S_t^b \in \mathcal{S}_b$, respectively, where the joint distribution of (S_t, S_t^a, S_t^b) factorizes as

$$P_{S_t^a,S_t^b,S_t}(s_t^a, s_t^b, s_t) = P_{S_t^a|S_t}(s_t^a|s_t) P_{S_t^b|S_t}(s_t^b|s_t) P_{S_t}(s_t). \quad (1)$$

The system is depicted in Fig. 1. Let S_t be available at the receiver and let $\{(S_t, S_t^a, S_t^b)\}_{t=1}^{\infty}$ be a sequence of i.i.d. triples, independent from (W_a, W_b). Hence, for any $n \geq 1$,

$$P_{S_{[n]},S_{[n]}^a,S_{[n]}^b,W_a,W_b}(s_{[n]}, s_{[n]}^a, s_{[n]}^b, w_a, w_b)$$

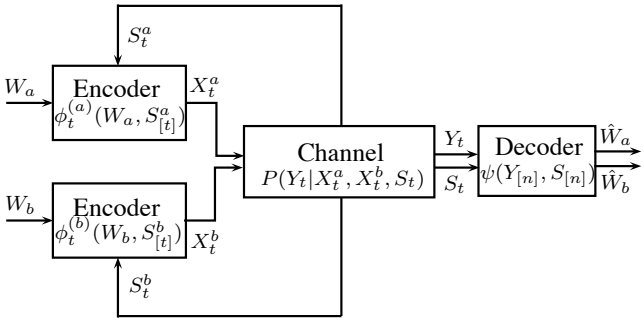

Fig. 1. The multiple-access channel with asymmetric causal noisy CSI.

$$= \prod_{t=1}^{n} \frac{1}{|\mathcal{W}_a|} \frac{1}{|\mathcal{W}_b|} P_{S_t^a|S_t}(s_t^a|s_t) P_{S_t^b|S_t}(s_t^b|s_t) P_{S_t}(s_t). \quad (2)$$

The channel inputs at time t, i.e., X_t^a and X_t^b, are functions of the locally available information $(W_a, S_{[t]}^a)$ and $(W_b, S_{[t]}^b)$, respectively. Let $\mathbf{W} := (W_a, W_b)$ and $\mathbf{X_t} := (X_t^a, X_t^b)$, respectively. Then, the laws governing n-sequences of state, input and output letters are given by

$$P_{Y_{[n]}|\mathbf{W},\mathbf{X}_{[n]},S_{[n]},S_{[n]}^a,S_{[n]}^b}(y_{[n]}|\mathbf{w}, \mathbf{x}_{[n]}, s_{[n]}, s_{[n]}^a, s_{[n]}^b) = \prod_{t=1}^{n} P_{Y_t|X_t^a,X_t^b,S_t}(y_t|x_t^a, x_t^b, s_t), \quad (3)$$

where $P_{Y_t|X_t^a,X_t^b,S_t}(y_t|x_t^a, x_t^b, s_t)$, the channel's transition probability distribution, is given a priori.

Definition 2.1: An $(n, 2^{nR_a}, 2^{nR_b})$ code with block length n and rate pair (R_a, R_b) for a state-dependent MAC with causal noisy state information consists of

(1) A sequence of mappings for each encoder

$$\phi_t^{(a)} : \mathcal{S}_a^t \times \mathcal{W}_a \to \mathcal{X}_a, \ \ t = 1, 2, ...n;$$
$$\phi_t^{(b)} : \mathcal{S}_b^t \times \mathcal{W}_b \to \mathcal{X}_b, \ \ t = 1, 2, ...n.$$

2) An associated decoding function

$$\psi : \mathcal{S}^n \times \mathcal{Y}^n \to \mathcal{W}_a \times \mathcal{W}_b.$$

Let $P_{e,f} := P\left(\psi(Y_{[n]}, S_{[n]}) \neq (w_a, w_b)|\mathbf{W} = \mathbf{w}\right)$.The system's probability of error, $P_e^{(n)}$, is given by

$$P_e^{(n)} = \frac{1}{2^{n(R_a+R_b)}} \sum_{w_a=1}^{2^{nR_a}} \sum_{w_b=1}^{2^{nR_b}} P_{e,f}.$$

A rate pair (R_a, R_b) is achievable if for any $\epsilon > 0$, there exists, for all n sufficiently large an $(n, 2^{nR_a}, 2^{nR_b})$ code such that $\frac{1}{n}\log|\mathcal{W}_a| \geq R_a > 0$, $\frac{1}{n}\log|\mathcal{W}_b| \geq R_b > 0$ and $P_e^{(n)} \leq \epsilon$. The capacity region of the state-dependent MAC, $\mathcal{C}_{FS}$, is the closure of the set of all achievable rate pairs (R_a, R_b) and the sum-rate capacity is defined as $\mathcal{C}_{FS}^{\sum} := \max_{(R_a,R_b)\in\mathcal{C}_{FS}}(R_a + R_b)$.

Before proceeding with the main result, we introduce *memoryless stationary team policies* [13] and their associated rate regions. Let the set of all possible functions from $\mathcal{S}_a$ to $\mathcal{X}_a$ and $\mathcal{S}_b$ to $\mathcal{X}_b$ be denoted by $\mathcal{T}_a := {\mathcal{X}_a}^{|\mathcal{S}_a|}$ and $\mathcal{T}_b := {\mathcal{X}_b}^{|\mathcal{S}_b|}$, respectively. We shall refer to $\mathcal{T}_a$-valued and $\mathcal{T}_b$-valued random vectors as Shannon strategies.

Definition 2.2: [13] A memoryless stationary (in time) team policy is a family

$$\Pi = \{\pi = (\pi_{T^a}(\cdot), \pi_{T^b}(\cdot)) \in \mathcal{P}(\mathcal{T}_a) \times \mathcal{P}(\mathcal{T}_b)\} \quad (4)$$

of probability distribution pairs on $(\mathcal{T}_a, \mathcal{T}_b)$.

For every memoryless stationary team policy π, let $\mathcal{R}_{FS}(\pi)$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$\begin{aligned} R_a &< I(T^a; Y|T^b, S) && (5)\\ R_b &< I(T^b; Y|T^a, S) && (6)\\ R_a + R_b &< I(T^a, T^b; Y|S) && (7) \end{aligned}$$

where S, T^a, T^b and Y are random variables taking values in $\mathcal{S}$, $\mathcal{T}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively, and whose joint probability distribution factorizes as

$$\begin{aligned} &P_{S,T^a,T^b,Y}(s, t^a, t^b, y)\\ &\quad = P_S(s) P_{Y|T^a,T^b,S}(y|t^a, t^b, s)\pi_{T^a}(t^a)\pi_{T^b}(t^b). \quad (8) \end{aligned}$$

Let $\mathcal{C}_{IN} := \overline{co}\left(\bigcup_\pi \mathcal{R}_{FS}(\pi)\right)$ denote the closure of the convex hull of the rate regions $\mathcal{R}_{FS}(\pi)$ given by (5)-(7) associated to all possible memoryless stationary team polices as defined in (4).

Theorem 2.1 (Inner Bound to $\mathcal{C}_{FS}$): $\mathcal{C}_{IN} \subseteq \mathcal{C}_{FS}$.
The achievability proof (which we omit) is based on a random code construction with Shannon strategies and follows the standard arguments involving joint ϵ-typical sequences (e.g., cf. [36, Section 15.2]). Let

$$\mathcal{C}_{OUT} := \Big\{(R_a, R_b) \in \mathbb{R}^+ \times \mathbb{R}^+ : R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S)\Big\},$$

where $\mathbb{R}^+$ is the set of positive reals.

Theorem 2.2 (Outer Bound to $\mathcal{C}_{FS}$): $\mathcal{C}_{FS} \subseteq \mathcal{C}_{OUT}$.

Proof of Theorem 2.2: We need to show that all achievable rates satisfy

$$R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S),$$

i.e., a converse for the sum-rate capacity. Following [13], for $1 \leq t \leq n$, let

$$\alpha_\mu := \frac{1}{n} P_{S_{[t-1]}}(\mu) \text{ and } \eta(\epsilon) := \frac{\epsilon}{1-\epsilon}\log|\mathcal{Y}| + \frac{H(\epsilon)}{1-\epsilon}. \quad (9)$$

Observe that $\lim_{\epsilon\to 0}\eta(\epsilon) = 0$ and

$$\sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu = \frac{1}{n}\sum_{1\leq t\leq n}\sum_{\mu\in\mathcal{S}^{t-1}} P_{S_{[t-1]}}(\mu) = 1,$$

where $\mathcal{S}^{(n)}$ is the set of all $\mathcal{S}$-strings of length less than n.

Recall that $X_t^a = \phi_t^{(a)}\left(W_a, S_{[t]}^a\right) = \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, S_t^a\right)$ and $X_t^b = \phi_t^{(b)}\left(W_b, S_{[t]}^b\right) = \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, S_t^b\right)$, for all $t \geq 1$. Then, we can define the Shannon strategies $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ by putting, for every $s_a \in \mathcal{S}_a$ and $s_b \in \mathcal{S}_b$,

$$\begin{aligned} T_t^a(s_a) &:= \phi_t^{(a)}\left(W_a, S_{[t-1]}^a, s_a\right)\\ T_t^b(s_b) &:= \phi_t^{(b)}\left(W_b, S_{[t-1]}^b, s_b\right). \quad (10) \end{aligned}$$

We now show that the sum of any achievable rate pair can be written as the convex combinations of mutual information terms which are indexed by the realization of past complete CSI.

Lemma 2.1: Let $T_t^a \in \mathcal{T}_a$ and $T_t^b \in \mathcal{T}_b$ be the Shannon strategies induced by $\phi_t^{(a)}$ and $\phi_t^{(b)}$, respectively, as shown in (10). Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,

$$R_a + R_b \leq \sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu) + \eta(\epsilon). \quad (11)$$

Proof: Let $\mathbf{T}_t := (T_t^a, T_t^b)$. By Fano's inequality, we get

$$H(\mathbf{W}|Y_{[n]}, S_{[n]}) \leq H(\epsilon) + \epsilon\log(|\mathcal{W}_a||\mathcal{W}_b|). \quad (12)$$

Observing that

$$\begin{aligned} I(\mathbf{W}; Y_{[n]}, S_{[n]}) &= H(\mathbf{W}) - H(\mathbf{W}|Y_{[n]}, S_{[n]})\\ &= \log(|\mathcal{W}_a||\mathcal{W}_b|) - H(\mathbf{W}|Y_{[n]}, S_{[n]}). \quad (13) \end{aligned}$$

Combining (12) and (13) gives

$$(1-\epsilon)\log(|\mathcal{W}_a||\mathcal{W}_b|) \leq I(\mathbf{W}; Y_{[n]}, S_{[n]}) + H(\epsilon)$$

and

$$\begin{aligned} R_a + R_b &\leq \frac{1}{n}\log(|\mathcal{W}_a||\mathcal{W}_b|)\\ &\leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(\mathbf{W}; Y_{[n]}, S_{[n]}) + H(\epsilon)\right). \quad (14) \end{aligned}$$

Furthermore,

$$\begin{aligned} &I(\mathbf{W}; Y_{[n]}, S_{[n]})\\ &= \sum_{t=1}^n \big[H(Y_t, S_t|S_{[t-1]}, Y_{[t-1]})\\ &\qquad - H(Y_t, S_t|\mathbf{W}, S_{[t-1]}, Y_{[t-1]})\big]\\ &\overset{(i)}{=} \sum_{t=1}^n \big[H(Y_t|S_{[t]}, Y_{[t-1]}) - H(Y_t|\mathbf{W}, S_{[t]}, Y_{[t-1]})\big]\\ &\overset{(ii)}{\leq} \sum_{t=1}^n \big[H(Y_t|S_{[t]}) - H(Y_t|\mathbf{W}, S_{[t]}, Y_{[t-1]}, \mathbf{T}_t)\big]\\ &\overset{(iii)}{=} \sum_{t=1}^n \big[H(Y_t|S_{[t]}) - H(Y_t|S_{[t]}, \mathbf{T}_t)\big]\\ &= \sum_{t=1}^n I(\mathbf{T}_t; Y_t|S_{[t]}) \quad (15) \end{aligned}$$

where (i) is implied by (2), in (ii) $\mathbf{T}_t := (T_t^a, T_t^b)$ are Shannon strategies whose realizations are mappings $t_t^i : S_t^i \to X_t^i$ for $i = \{a, b\}$ and thus (ii) holds since conditioning does not increase entropy. Finally, (iii) follows since

$$\begin{aligned} &P_{Y_t|\mathbf{W},S_t,S_{[t-1]},Y_{[t-1]},T_t^a,T_t^b}(y_t|\mathbf{w}, s_t, s_{[t-1]}, y_{[t-1]}, t_t^a, t_t^b)\\ &\quad = \sum_{s_t^a,s_t^b} P_{Y_t|S_t,S_t^a,S_t^b,T_t^a,T_t^b}(y_t|s_t, s_t^a, s_t^b, t_t^a, t_t^b)\\ &\qquad \times P_{S_t^a,S_t^b|S_t}(s_t^a, s_t^b|s_t)\\ &\quad = P_{Y_t|S_t,T_t^a,T_t^b}(y_t|s_t, t_t^a, t_t^b) \quad (16) \end{aligned}$$

where the first equality is verified by (3) and (2), where $x_t^i = t_t^i(s_t^i)$ for $i = \{a, b\}$. At this point, it is worth to note that by (16), one can remove $S_{[t-1]}$ from (15) in the conditioning. However, we will soon observe why it is crucial to keep it when we prove the product form. Now, let $\chi(\epsilon) := \frac{H(\epsilon)}{n(1-\epsilon)}$ and combining (14)-(15) gives

$$\begin{aligned}
&R_a + R_b \\
&\leq \frac{1}{n}\log(|\mathcal{W}_a||\mathcal{W}_b|) \\
&\leq \left(\frac{1}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]})\right) + \chi(\epsilon) + (n-1)\chi(\epsilon) \\
&\overset{(a)}{\leq} \frac{1}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) + \eta(\epsilon) \\
&\qquad\qquad - \frac{\epsilon}{1-\epsilon}\frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) \\
&= \frac{1}{n}\sum_{t=1}^{n} I(T_t^a, T_t^b; Y_t|S_{[t]}) + \eta(\epsilon) \qquad (17)
\end{aligned}$$

where (a) is valid since $I(T_t^a, T_t^b; Y_t|S_{[t]}) \leq \log|\mathcal{Y}|$. Furthermore,

$$\begin{aligned}
&I(T_t^a, T_t^b; Y_t|S_{[t]}) \\
&\qquad = n\sum_{\mu\in\mathcal{S}^{t-1}} \alpha_\mu I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu), \qquad (18)
\end{aligned}$$

and substituting the above into (17) yields (11). ■

Note that, for any $t \geq 1$, $I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu)$ is a function of the joint conditional distribution of channel state S_t, inputs T_t^a, T_t^b and output Y_t given the past realization $(S_{[t-1]} = \mu)$. Hence, to complete the proof of the outer bound, we need to show that $P_{T_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(t^a, t^b, y, s|\mu)$ factorizes as in (8). This is done in the lemma below. In particular, it is crucial to observe that the knowledge of the past state at the decoder, $S_{[t-1]}$, is enough to provide a product form on T^a and T^b. Let

$$\begin{aligned}
\Upsilon_{\mu_\mathbf{a}}^a(t^a) &:= \{w_a : \phi_t^{(a)}(w_a, s_{[t-1]}^a = \mu_\mathbf{a}) = t^a\}, \\
\Upsilon_{\mu_\mathbf{b}}^b(t^b) &:= \{w_b : \phi_t^{(b)}(w_b, s_{[t-1]}^b = \mu_\mathbf{b}) = t^b\} \qquad (19)
\end{aligned}$$

and

$$\begin{aligned}
\pi_{T^a}^{\mu_\mathbf{a}}(t^a) &:= \sum_{w_a\in\Upsilon_{\mu_\mathbf{a}}^a(t^a)} \frac{1}{|\mathcal{W}_a|}, \\
\pi_{T^b}^{\mu_\mathbf{b}}(t^b) &:= \sum_{w_b\in\Upsilon_{\mu_\mathbf{b}}^b(t^b)} \frac{1}{|\mathcal{W}_b|}, \\
\pi_{T^a}^{\mu}(t^a) &:= \sum_{\mu_\mathbf{a}} \pi_{T^a}^{\mu_\mathbf{a}}(t^a) P_{S_{[t-1]}^a|S_{[t-1]}}(\mu_\mathbf{a}|\mu), \\
\pi_{T^b}^{\mu}(t^b) &:= \sum_{\mu_\mathbf{b}} \pi_{T^b}^{\mu_\mathbf{b}}(t^b) P_{S_{[t-1]}^b|S_{[t-1]}}(\mu_\mathbf{b}|\mu), \qquad (20)
\end{aligned}$$

where $\mu_\mathbf{a}$ and $\mu_\mathbf{b}$ denote particular realizations of $S_{[t-1]}^a$ and $S_{[t-1]}^b$, respectively.

Lemma 2.2: For every $1 \leq t \leq n$ and $\mu \in \mathcal{S}^{t-1}$, the following holds

$$\begin{aligned}
&P_{T_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(t^a, t^b, y, s|\mu) \\
&\qquad = P_S(s) P_{Y|S,T^a,T^b}(y|s, t^a, t^b)\pi_{T^a}^{\mu}(t^a)\pi_{T^b}^{\mu}(t^b). \qquad (21)
\end{aligned}$$

Proof: Let $\mathbf{S} := (S_t, S_t^a, S_t^b)$ and $\mathbf{s} := (s, s_t^a, s_t^b)$. Observe that

$$\begin{aligned}
&P_{T_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(t^a, t^b, y, s|\mu) \\
&:= \sum_{s_t^a,s_t^b} P_{\mathbf{S},T_t^a,T_t^b,Y_t|S_{[t-1]}}(\mathbf{s}, t^a, t^b, y|\mu) \\
&:= \sum_{s_t^a,s_t^b} P_{Y|\mathbf{S},T_t^a,T_t^b}(y|\mathbf{s}, t^a, t^b) P_{\mathbf{S},T_t^a,T_t^b|S_{[t-1]}}(\mathbf{s}, t^a, t^b|\mu) \qquad (22)
\end{aligned}$$

where the second equality is shown in (16). Let us now consider the term $P_{\mathbf{S},T_t^a,T_t^b|S_{[t-1]}}(\mathbf{s}, t^a, t^b|\mu)$ above. We have the following

$$\begin{aligned}
&P_{\mathbf{S},T_t^a,T_t^b|S_{[t-1]}}(\mathbf{s}, t^a, t^b|\mu) \\
&= \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_\mathbf{a}}\sum_{\mu_\mathbf{b}} \\
&\qquad P_{\mathbf{W},S_{[t-1]}^a,S_{[t-1]}^b,\mathbf{S},T_t^a,T_t^b|S_{[t-1]}}(\mathbf{w}, \mu_\mathbf{a}, \mu_\mathbf{b}, \mathbf{s}, t^a, t^b|\mu) \\
&\overset{(i)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_\mathbf{a}}\sum_{\mu_\mathbf{b}} \\
&\qquad P_{\mathbf{W},S_{[t-1]}^a,S_{[t-1]}^b,T_t^a,T_t^b|S_{[t-1]}}(\mathbf{w}, \mu_\mathbf{a}, \mu_\mathbf{b}, t^a, t^b|\mu) \\
&\overset{(ii)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_\mathbf{a}}\sum_{\mu_\mathbf{b}} 1_{\{t^l=\phi_t^{(l)}(w_l,\mu_\mathbf{l}),\ \ l=a,b\}} \\
&\qquad P_{\mathbf{W},S_{[t-1]}^a,S_{[t-1]}^b|S_{[t-1]}}(\mathbf{w}, \mu_\mathbf{a}, \mu_\mathbf{b}|\mu) \\
&\overset{(iii)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}\sum_{\mu_\mathbf{a}}\sum_{\mu_\mathbf{b}} 1_{\{t^l=\phi_t^{(l)}(w_l,\mu_\mathbf{l}),\ \ l=a,b\}} \\
&\qquad \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|} P_{S_{[t-1]}^a,S_{[t-1]}^b|S_{[t-1]}}(\mu_\mathbf{a}, \mu_\mathbf{b}|\mu) \\
&\overset{(iv)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{\mu_\mathbf{a}} P_{S_{[t-1]}^a|S_{[t-1]}}(\mu_\mathbf{a}|\mu) \sum_{\mu_\mathbf{b}} P_{S_{[t-1]}^b|S_{[t-1]}}(\mu_\mathbf{b}|\mu) \\
&\qquad \sum_{w_a\in\mathcal{W}_a} \frac{1}{|\mathcal{W}_a|} 1_{\{t^a=\phi_t^{(a)}(w_a,\mu_\mathbf{a})\}} \\
&\qquad\qquad \sum_{w_b\in\mathcal{W}_b} \frac{1}{|\mathcal{W}_b|} 1_{\{t^b=\phi_t^{(b)}(w_b,\mu_\mathbf{b})\}} \\
&\overset{(v)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{\mu_\mathbf{a}} P_{S_{[t-1]}^a|S_{[t-1]}}(\mu_\mathbf{a}|\mu) \sum_{w_a\in\Upsilon_{\mu_\mathbf{a}}^a(t^a)} \frac{1}{|\mathcal{W}_a|} \\
&\qquad \sum_{\mu_\mathbf{b}} P_{S_{[t-1]}^b|S_{[t-1]}}(\mu_\mathbf{b}|\mu) \sum_{w_b\in\Upsilon_{\mu_\mathbf{b}}^b(t^b)} \frac{1}{|\mathcal{W}_b|} \\
&\overset{(vi)}{=} P_\mathbf{S}(\mathbf{s}) \sum_{\mu_\mathbf{a}} P_{S_{[t-1]}^a|S_{[t-1]}}(\mu_\mathbf{a}|\mu)\pi_{T^a}^{\mu_\mathbf{a}}(t^a) \\
&\qquad \sum_{\mu_\mathbf{b}} P_{S_{[t-1]}^b|S_{[t-1]}}(\mu_\mathbf{b}|\mu)\pi_{T^b}^{\mu_\mathbf{b}}(t^b) \\
&\overset{(vii)}{=} P_\mathbf{S}(\mathbf{s})\pi_{T^a}^{\mu}(t^a)\pi_{T^b}^{\mu}(t^b) \qquad (23)
\end{aligned}$$

where (i) is due to (2) and (10), (ii) is valid by (10), (iii) is due to (2), (iv) is valid by (1) and (10), (v) is valid due to (19) and $(vi) - (vii)$ is valid due to (20). Substituting (23) into (22) proves the lemma. ■

We can now complete the proof of Theorem 2.2. We have

$$\begin{aligned} & R_a + R_b \\ & \leq \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t | S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \\ & = \sum_{\mu \in \mathcal{S}^{(n)}} \alpha_\mu I(T_t^a, T_t^b; Y_t | S_t)_{\pi^\mu_{T^a}(t^a)\pi^\mu_{T^b}(t^b)} + \eta(\epsilon) \\ & \leq \sup_{\left(\pi_{T^a}(t^a)\pi_{T^b}(t^b) \in \Pi\right)} I(T_t^a, T_t^b; Y_t | S_t) + \eta(\epsilon), \end{aligned}$$

where $I(T_t^a, T_t^b; Y_t|S_t)_{\pi^\mu_{T^a}(t^a)\pi^\mu_{T^b}(t^b)}$ denotes the mutual information induced by the product distribution $\pi^\mu_{T^a}(t^a)\pi^\mu_{T^b}(t^b)$ and the second step is valid since $I(T_t^a, T_t^b; Y_t|S_t, S_{[t-1]} = \mu)$ is a function of the joint conditional distribution of channel state S_t, inputs T_t^a, T_t^b and output Y_t given the past realization $(S_{[t-1]} = \mu)$. Hence, since $\lim_{\epsilon \to 0} \eta(\epsilon) = 0$, any achievable pair satisfies $R_a + R_b \leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S)$. ■

As a consequence of Theorems 2.1 and 2.2, we have the following corollary which can be thought of as an extension of [11, Theorem 4] to the case where the encoders have correlated CSI.

Corollary 2.1:

$$\mathcal{C}_{\sum}^{FS} = \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} I(T^a, T^b; Y|S). \tag{24}$$

Proof of Corollary 2.1: We need to show that $\exists\ (R_a, R_b) \in \mathcal{C}_{IN}$ achieving (24). We follows steps akin to [36, p.535] where discrete memoryless MACs are considered. Let us fix $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$ and consider the rate constraints given in $\mathcal{C}_{IN}$

$$\begin{aligned} I(T^a; Y|T^b, S) &= H(T^a|T^b, S) - H(T^a|T^b, Y, S) \\ &= H(T^a) - H(T^a|T^b, Y, S) \qquad (25) \\ I(T^b; Y|T^a, S) &= H(T^b|T^a, S) - H(T^b|T^a, Y, S) \\ &= H(T^b) - H(T^b|T^a, Y, S) \qquad (26) \end{aligned}$$

and

$$\begin{aligned} & I(T^a, T^b; Y|S) \\ & = H(T^a, T^b) - H(T^a, T^b|Y, S) \\ & = H(T^a) + H(T^b) - H(T^a|T^b, Y, S) - H(T^b|Y, S), \qquad (27) \end{aligned}$$

where (25), (26) and (27) are valid since T^a and T^b are independent of each other and independent of S. Observe now that for any $\pi_{T^a}(t^a)\pi_{T^b}(t^b)$, $I(T^a; Y|T^b, S) + I(T^b; Y|T^a, S) \geq I(T^a, T^b; Y|S)$ since $H(T^b|Y, S) \geq H(T^b|T^a, Y, S)$. Therefore, the sum-rate constraint in $\mathcal{C}_{IN}$ is always active and hence, there exists $(R_a, R_b) \in \mathcal{C}_{IN}$ achieving (24). ■

We conclude this section with a number of remarks.

Remark 2.1: One essential step in the proof of Theorem 2.2 is that, once we have the complete CSI, conditioning on which allows a product form on T^a and T^b, there is no loss of optimality (for the sum-rate capacity) in using associated memoryless team policies instead of using all the past information at the receiver.

Remark 2.2: For the validity of Corollary 2.1, it is crucial to have the product form on the pair (T^a, T^b). If this is not the case, we would get that $I(T^a; Y|T^b, S) + I(T^b; Y|T^a, S) = H(T^a|T^b) + H(T^b|T^a) - H(T^a|T^b, Y, S) - H(T^b|T^a, Y, S)$ and $I(T^a, T^b; Y|S) = H(T^a|T^b) + H(T^b) - H(T^a|T^b, Y, S) - H(T^b|Y, S)$. Therefore, it is possible to get an obsolete sum-rate constraint in $\mathcal{C}_{IN}$ and hence, achievability of $\mathcal{C}_{FS}^{\sum}$ is not guaranteed. Note that the channel inputs are not independent since $X^a = T^a(S^a)$ and $X^b = T^b(S^b)$.

Remark 2.3 (Cases of partial and no CSIR): In the situation where the receiver has partial information about the state at time t in the sense that it is provided with process $\{(V_t)\}_{t=1}^\infty$, $V_t \in \mathcal{V}$, which is independent of (W_a, W_b) and satisfies the following

$$\begin{aligned} & P_{S_{[t]}, S^a_{[t]}, S^b_{[t]}, V_{[t]}}(s_{[t]}, s^a_{[t]}, s^b_{[t]}, v_{[t]}) \\ & \quad = P(s^a_{[t]}|v_{[t]}) P(s^b_{[t]}|v_{[t]}) P(s_{[t]}, v_{[t]}), \ \ 1 \leq t \leq n, \qquad (28) \end{aligned}$$

it can be shown that the sum-rate capacity admits a similar expression as in (24) with S replaced by V, see [37, Theorem 5.2.3]. Furthermore, inspired by the coding schemes of the lossless CEO problem [38] as well as of a recently proposed achievable region [14], an inner bound, which demonstrates the rate required to transmit the above partial information about the state in the case where the receiver has no CSI, is shown in [37, Theorem 5.3.2].

B. CSITs as Deterministic Functions of CSIR: Non-Causal Case

In this section we consider the situation where the transmitters have access to partial state information available at the decoder. In particular, let $S_t^i = f^i(S_t^r)$, where $f^i : \mathcal{S}_r \to \mathcal{S}_i$, $i = \{a, b\}$ and $S^r \in \mathcal{S}_r$ such that

$$\begin{aligned} & P_{S_{[n]}, S^a_{[n]}, S^b_{[n]}, S^r_{[n]}, W_a, W_b}(s_{[n]}, s^a_{[n]}, s^b_{[n]}, s^r_{[n]}, w_a, w_b) \\ & \qquad = \prod_{t=1}^n \frac{1}{|\mathcal{W}_a|} \frac{1}{|\mathcal{W}_b|} P_{S_t, S^a_t, S^b_t, S^r_t}(s_t, s^a_t, s^b_t, s^r_t). \qquad (29) \end{aligned}$$

The channel is driven by the state process $\{S_t\}_{t=1}^\infty$ and hence,

$$\begin{aligned} & P_{Y_{[n]}|\mathbf{W}, \mathbf{X}_{[n]}, S_{[n]}, S^a_{[n]}, S^b_{[n]}, S^r_{[n]}}(y_{[n]}|\mathbf{w}, \mathbf{x}_{[n]}, s_{[n]}, s^a_{[n]}, s^b_{[n]}, s^r_{[n]}) \\ & \qquad = \prod_{t=1}^n P_{Y_t|X^a_t, X^b_t, S_t}(y_t|x^a_t, x^b_t, s_t). \qquad (30) \end{aligned}$$

Note that one can define an equivalent channel with conditional output probability

$$\begin{aligned} & P^{eq}_{Y|X^a, X^b, S^r}(y|x^a, x^b, s^r) \\ & \qquad = \sum_{s \in \mathcal{S}} P_{Y|X^a, X^b, S}(y|x^a, x^b, s) P_{S|S^r}(s|s^r). \qquad (31) \end{aligned}$$

Hence, the causal setup of this problem is no more general than the setup in [13] and the main result of this subsection is to show that the result of [13] also holds for non-causal coding.

We keep the channel codes definition identical for the causal and non-causal cases, except for the non-causal case we have; $\phi_t^{(i)} : \mathcal{S}_i^n \times \mathcal{W}_i \to \mathcal{X}_i^n$, $i = \{a, b\}$, $t = 1, \cdots, n$. Let $\mathcal{C}_{NS}^Q$ denote the capacity region. We need to modify Definition 2.2 in order to take the current CSI into account.

Definition 2.3: A memoryless stationary (in time) team policy is a family

$$\bar{\Pi} = \big\{\bar{\pi} = \big(\pi_{X^a|S^a}(\cdot|f^a(s^r)), \pi_{X^b|S^b}(\cdot|f^b(s^r))\big) \in \mathcal{P}(\mathcal{X}_a) \times \mathcal{P}(\mathcal{X}_b)\big\} \quad (32)$$

For every $\bar{\pi}$ defined in (32), $\mathcal{R}^Q_{NS}(\bar{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_a < I(X^a; Y|X^b, S^r) \quad (33)$$
$$R_b < I(X^b; Y|X^a, S^r) \quad (34)$$
$$R_a + R_b < I(X^a, X^b; Y|S^r) \quad (35)$$

where S^r, X^a, X^b and Y are random variables taking values in $\mathcal{S}_r$, $\mathcal{X}_a$, $\mathcal{X}_b$ and $\mathcal{Y}$, respectively, and whose joint probability distribution factorizes as

$$P_{S^r,X^a,X^b,Y}(s^r, x^a, x^b, y) = P_{S^r}(s^r)P_{Y|X^a,X^b,S^r}(y|x^a, x^b, s^r) \times \pi_{X^a|S^a}(x^a|f^a(s^r))\pi_{X^b|S^b}(x^b|f^b(s^r)). \quad (36)$$

Let $\overline{co}\Big(\bigcup_{\bar{\pi}} \mathcal{R}^Q_{NS}(\bar{\pi})\Big)$ denote the closure of the convex hull of the rate regions $\mathcal{R}^Q_{NS}(\bar{\pi})$ given by (33)-(35) associated to all possible memoryless stationary team polices as defined in (32).

Theorem 2.3: $\mathcal{C}^Q_{NS} = \overline{co}\Big(\bigcup_{\bar{\pi}} \mathcal{R}^Q_{NS}(\bar{\pi})\Big)$.

For the achievability proof, see [13, Section III] and observe that any rate which is achievable with causal CSI is also achievable with non-causal CSI. For the converse proof of the non-causal case see Appendix A. The proof for the non-causal case is realized by observing that there is no loss of optimality if not only the past, as shown in [13], but also the future CSI is ignored given that the receiver is provided with complete CSI. A similar observation for independent CSIT is also made see [11, Theorem 5].

Consider now the setup in Section II in order to observe that for the non-causal case the optimality of Shannon strategies is not guaranteed. Recall that, we have

$$I(\mathbf{W}; Y_{[n]}, S_{[n]}) \leq \sum_{t=1}^{n} \big[H(Y_t|S_{[n]}, Y_{[t-1]}) - H(Y_t|\mathbf{W}, S_{[n]}, Y_{[t-1]}, \mathbf{T}_t)\big] \quad (37)$$

where $\mathbf{T}_t := (T^a_t, T^b_t)$. Consider now the right hand side of (37) and observe that

$$P_{Y_t|\mathbf{W},S_{[n]},Y_{[t-1]},T^a_t,T^b_t}(y_t|\mathbf{w}, s_{[n]}, y_{[t-1]}, t^a_t, t^b_t) = \sum_{s^a_t,s^b_t} P_{Y_t|S_t,S^a_t,S^b_t,T^a_t,T^b_t}(y_t|s_t, s^a_t, s^b_t, t^a_t, t^b_t) \times P_{S^a_t,S^b_t|Y_{[t-1]},S_t}(s^a_t, s^b_t|y_{[t-1]}, s_t),$$

and therefore, the past channel outputs cannot be eliminated.

C. Asymmetric Noisy CSIT with Delays

Consider the problem defined in Section II-A where the two encoders have accesses to asymmetrically delayed, where delays are $d_a \geq 1$ and $d_b \geq 1$, respectively, and noisy versions of the state information S_t at each time $t \geq 1$, modeled by $S^a_{t-d_a} \in \mathcal{S}_a$, $S^b_{t-d_b} \in \mathcal{S}_b$, respectively. The rest of the channel model is identical and hence, (1), (2) and (3) are valid throughout this section. We also assume that S_t is fully available at the receiver. A code can be defined as in Definition 2.1, except now

$$\phi^{(a)}_t : \mathcal{S}^{t-d_a}_a \times \mathcal{W}_a \to \mathcal{X}_a, \ \ t = 1, 2, ...n;$$
$$\phi^{(b)}_t : \mathcal{S}^{t-d_b}_b \times \mathcal{W}_b \to \mathcal{X}_b, \ \ t = 1, 2, ...n.^1$$

Let $\mathcal{C}_{DN}$ denotes the capacity region of the delayed setup.

In the main result of this section the team policies are composed of probability distributions on the channel inputs rather than Shannon strategies.

Definition 2.4: A memoryless stationary (in time) team policy is a family

$$\tilde{\Pi} = \big\{\tilde{\pi} = (\pi_{X^a}(\cdot), \pi_{X^b}(\cdot)) \in \mathcal{P}(\mathcal{X}^a) \times \mathcal{P}(\mathcal{X}^b)\big\}. \quad (38)$$

For every memoryless stationary team policy $\tilde{\pi}$, $\mathcal{R}_{DN}(\tilde{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_a < I(X^a; Y|X^b, S) \quad (39)$$
$$R_b < I(X^b; Y|X^a, S) \quad (40)$$
$$R_a + R_b < I(X^a, X^b; Y|S) \quad (41)$$

where S, X^a, X^b and Y are random variables taking values in $\mathcal{S}$, $\mathcal{X}^a$, $\mathcal{X}^b$ and $\mathcal{Y}$, respectively and whose joint probability distribution factorizes as

$$P_{S,X^a,X^b,Y}(s, x^a, x^b, y) = P_S(s)P_{Y|X^a,X^b,S}(y|x^a, x^b, s)\pi_{X^a}(x^a)\pi_{X^b}(x^b). \quad (42)$$

Let $\overline{co}\Big(\bigcup_{\tilde{\pi}} \mathcal{R}_{DN}(\tilde{\pi})\Big)$ denotes the closure of the convex hull of the rate regions $\mathcal{R}_{DN}(\tilde{\pi})$ given by (39)-(41) associated to all possible memoryless stationary team polices as defined in (38).

Theorem 2.4: $\mathcal{C}_{DN} = \overline{co}\Big(\bigcup_{\tilde{\pi}} \mathcal{R}_{DN}(\tilde{\pi})\Big)$.

Achievability can be shown via random coding arguments. For the converse, see Appendix B.

Remark 2.4 (Strictly Causal CSIT): When $d_a = d_b = 1$, Theorem 2.4 is the capacity region of the setup with strictly causal CSITs. This case was considered in the literature, e.g., see [14], [16], [15] and [22], where it is shown that strictly causal side information is helpful. Theorem 2.4 verifies that since the full CSI is available at the receiver and since the decoder does not need to access the current CSI at the encoders, there exists no loss of optimality if the past information at the encoders are ignored.

D. Degraded Message Set with Noisy CSIT

Assume a common message is provided to both encoders and one of the encoders has its own private message. Assume further that the encoder with the private message has causal noisy CSI, whereas the encoder with the common

[1] Obviously, when $d_l \geq t, \ l = a, b$ then $X^a_t = \phi^{(a)}_t(W_a)$ and $X^b_t = \phi^{(b)}_t(W_b)$.

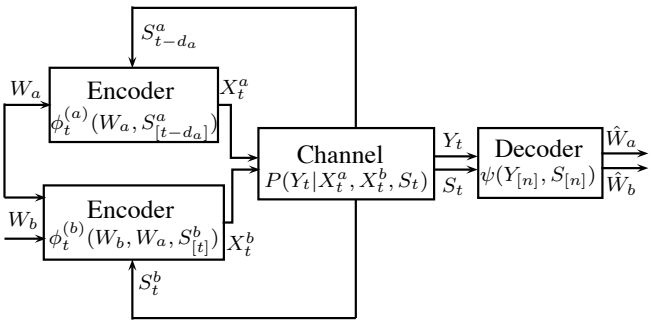

Fig. 2. MAC with degraded message set and with noisy CSI.

message only observes noisy state information with delay $d_a \geq 1$. Let the common and the private messages be W_a and W_b, respectively, and $S^a_{[t-d_a]}$, $d_a \geq 1$, and $S^b_{[t]}$ denote the CSI at encoder a, b, respectively, where (S_t, S^a_t, S^b_t) satisfies (1) and (2). Hence, $X^a_t = \phi^{(a)}_t(W_a, S^a_{[t-d_a]})$ and $X^b_t = \phi^{(b)}_t(W_a, W_b, S^b_{[t]})$; see Fig. 2. Let $\mathcal{C}_C$ denote the capacity region for this channel. Recall that $\mathcal{T}_b = \mathcal{X}_b^{|\mathcal{S}_b|}$.

Definition 2.5: A memoryless stationary (in time) team policy is a family

$$\hat{\Pi} = \left\{\hat{\pi} = \left(\pi_{X^a,T^b}(\cdot,\cdot)\right) \in \mathcal{P}(\mathcal{X}^a \times \mathcal{T}^b)\right\} \tag{43}$$

of probability distributions on $(\mathcal{X}_a, \mathcal{T}_b)$.

Let for every $\hat{\pi}$, $\mathcal{R}_C(\hat{\pi})$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$R_b < I(T^b; Y|X^a, S) \tag{44}$$

$$R_a + R_b < I(X^a, T^b; Y|S) \tag{45}$$

where S, X^a, T^b and Y are random variables taking values in $\mathcal{S}$, $\mathcal{X}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively and whose joint probability distribution factorizes as

$$P_{S,X^a,T^b,Y}(s, x^a, t^b, y) = P_S(s) P_{Y|X^a,T^b,S}(y|x^a, t^b, s)\pi_{X^a,T^b}(x^a, t^b). \tag{46}$$

Let $\overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}_C(\hat{\pi})\right)$ denotes the closure of the convex hull of the rate regions $\mathcal{R}_C(\hat{\pi})$ given by (44) and (45) associated to all possible memoryless stationary team polices as defined in (43).

Theorem 2.5: $\mathcal{C}_C = \overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}_C(\hat{\pi})\right)$.

See Appendix C for the proof.

Remark 2.5: Theorem 2.5 shows that when the common message encoder does not have access to the current noisy CSI (since the delay $d_a \geq 1$), by enlarging the optimization space of the other encoder, via Shannon strategies, the past CSI can be ignored without loss of optimality if the decoder is provided with complete CSI.

One important observation to be made in the degraded message set scenario is that we do not require a product form on the pair (X^a, T^b) (see (46)). In connection with this observation, let us consider the following noisy CSIR setup.

Let the encoder with the private message causally observe the noisy state information, whereas let the encoder with the common message have no CSI, i.e., $X^a_t = \phi^{(a)}_t(W_a)$ and $X^b_t = \phi^{(b)}_t(W_a, W_b, S^b_{[t]})$, and let the decoder also have access to noisy CSI at time t, $S^r_t \in \mathcal{S}_r$; see Fig. 3, where,

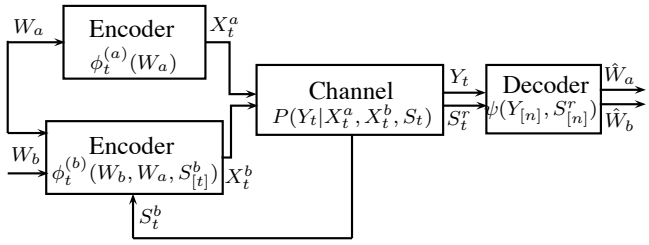

Fig. 3. MAC with degraded message set and with noisy CSIT and CSIR.

$$P_{S_{[n]},S^r_{[n]},S^b_{[n]},W_a,W_b}(s_{[n]}, s^r_{[n]}, s^b_{[n]}, w_a, w_b) = \prod_{t=1}^{n} \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|} P_{S_t,S^r_t,S^b_t}(s_t, s^r_t, s^b_t) \tag{47}$$

and let $\mathcal{C}^G_C$ denote the capacity region for this setup.

Let for every memoryless stationary team policy $\hat{\pi}$ defined in (43), $\mathcal{R}^G_C(\hat{\pi})$ denote the region of all rate pairs $R = (R_a, R_b)$ satisfying,

$$R_b < I(T^b; Y|X^a, S^r) \tag{48}$$

$$R_a + R_b < I(X^a, T^b; Y|S^r) \tag{49}$$

where S^r, X^a, T^b and Y are random variables taking values in $\mathcal{S}_r$, $\mathcal{X}_a$, $\mathcal{T}_b$ and $\mathcal{Y}$, respectively and whose joint probability distribution factorizes as

$$P_{S^r,X^a,T^b,Y}(s^r, x^a, t^b, y) = P_{S^r}(s^r) P_{Y|X^a,T^b,S^r}(y|x^a, t^b, s^r)\pi_{X^a,T^b}(x^a, t^b). \tag{50}$$

Let $\overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}^G_C(\hat{\pi})\right)$ denote the closure of the convex hull of the rate regions $\mathcal{R}^G_C(\hat{\pi})$ given by (48) and (49) associated to all possible $\hat{\pi}$ as defined in (43).

Theorem 2.6: $\mathcal{C}^G_C = \overline{co}\left(\bigcup_{\hat{\pi}} \mathcal{R}^G_C(\hat{\pi})\right)$.

Proof: The achievability proof is identical to that of Theorem 2.5. The converse proof is also similar and therefore, we only provide a sketch. In particular, observe the following lines of equations for the converse proof of the condition on R_b:

$$\begin{aligned}
I(W_b; Y_{[n]}, S^r_{[n]}) &\leq I(W_b; Y_{[n]}, S^r_{[n]}|W_a) \\
&= \sum_{t=1}^{n} \Big[H(Y_t, S^r_t|S^r_{[t-1]}, Y_{[t-1]}, W_a) \\
&\qquad - H(Y_t, S^r_t|S^r_{[t-1]}, Y_{[t-1]}, W_a, W_b)\Big] \\
&\overset{(i)}{=} \sum_{t=1}^{n} \Big[H(Y_t|S^r_{[t]}, Y_{[t-1]}, W_a) \\
&\qquad - H(Y_t|S^r_{[t]}, Y_{[t-1]}, W_a, W_b)\Big] \\
&= \sum_{t=1}^{n} \Big[H(Y_t|S^r_{[t]}, Y_{[t-1]}, W_a, X^a_t)
\end{aligned}$$

$$
\begin{aligned}
& -H(Y_t|S^r_{[t]}, Y_{[t-1]}, W_a, W_b, X^a_t)\Big] \\
\overset{(ii)}{\leq} & \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[t]}, X^a_t) \\
& -H(Y_t|S^r_{[t]}, Y_{[t-1]}, W_a, W_b, X^a_t, T^b_t)\Big] \\
\overset{(iii)}{=} & \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[t]}, X^a_t) - H(Y_t|S^r_{[t]}, X^a_t, T^b_t)\Big] \\
= & \sum_{t=1}^{n} I(T^b_t; Y_t|X^a_t, S^r_{[t]}) \qquad (51)
\end{aligned}
$$

where (i) follows since state is i.i.d., where T^b_t is the Shannon strategy induced by encoder b at time t as shown in (105), and (ii) is valid since conditioning does not increase entropy, and (iii) is valid since

$$
\begin{aligned}
& P_{Y_t|S^r_{[t]}, Y_{[t-1]}, \mathbf{W}, X^a_t, T^b_t}(y_t|s^r_{[t]}, y_{[t-1]}, \mathbf{w}, x^a_t, t^b_t) \\
= & \sum_{s_t\in\mathcal{S}, s^b_t\in\mathcal{S}_b} P_{Y_t|S_t, S^b_t, X^a_t, T^b_t}(y_t|s_t, s^b_t, x^a_t, t^b_t) \\
& \times P_{S^b_t, S_t|S^r_{[t]}, Y_{[t-1]}, \mathbf{W}, X^a_t, T^b_t}(s^b_t, s_t|s^r_{[t]}, y_{[t-1]}, \mathbf{w}, x^a_t, t^b_t) \\
= & \sum_{s_t\in\mathcal{S}, s^b_t\in\mathcal{S}_b} P_{Y_t|S_t, S^b_t, X^a_t, T^b_t}(y_t|s_t, s^b_t, x^a_t, t^b_t) \\
& \times P_{S^b_t, S_t}(s^b_t, s_t|s^r_t) \\
= & P_{Y_t|S^r_t, X^a_t, T^b_t}(y_t|s^r_t, x^a_t, t^b_t) \qquad (52)
\end{aligned}
$$

where the first equality is valid due to (3) and the second equality holds due to (47). Hence, one can directly obtain that

$$
R_b \leq \sum_{\mu_{\mathbf{r}}\in\mathcal{S}^{(n)}_r} \alpha_{\mu_{\mathbf{r}}} I(T^b_t; Y_t|X^a_t, S^r_t, S^r_{[t-1]} = \mu_{\mathbf{r}}) + \eta(\epsilon) \qquad (53)
$$

$$
\begin{aligned}
& R_a + R_b \\
& \leq \sum_{\mu_{\mathbf{r}}\in\mathcal{S}^{(n)}_r} \alpha_{\mu_{\mathbf{r}}} I(X^a_t, T^b_t; Y_t|S^r_t, S^r_{[t-1]} = \mu_{\mathbf{r}}) + \eta(\epsilon) \qquad (54)
\end{aligned}
$$

where $\alpha_{\mu_{\mathbf{r}}} := \frac{1}{n} P_{S^r_{[t-1]}}(\mu_{\mathbf{r}})$. We now need to show that the joint distribution $P_{X^a_t, T^b_t, Y_t, S^r_t|S^r_{[t-1]}}(x^a, t^b, y, s^r|\mu_{\mathbf{r}})$ satisfies (50). Let $\pi^{\mu_{\mathbf{r}}}_{X^a, T^b}(x^a, t^b) := P_{X^a_t, T^b_t|S^r_{[t-1]}}(x^a, t^b|\mu_{\mathbf{r}})$ and observe that

$$
\begin{aligned}
& P_{X^a_t, T^b_t, Y_t, S^r_t|S^r_{[t-1]}}(x^a, t^b, y, s^r|\mu_{\mathbf{r}}) \\
= & \sum_{s^b_t\in\mathcal{S}^b}\sum_{s_t\in\mathcal{S}} P_{Y_t|X^a_t, X^b_t, S_t}(y|x^a, t^b(s^b_t), s) \\
& \times P_{S^b_t, S_t, S_r}(s^b_t, s_t, s^r) P_{X^a_t, T^b_t|S^r_{[t-1]}}(x^a, t^b|\mu_{\mathbf{r}}) \\
= & \pi^{\mu}_{X^a, T^b}(x^a, t^b) P_{S^r_t}(s^r) P_{Y_t|X^a_t, T^b_t, S^r_t}(y|x^a, t^b, s^r) \qquad (55)
\end{aligned}
$$

where the first equality is verified by (3) and by the fact that (X^a_t, T^b_t) is independent of (S_t, S^b_t, S^r_t). ■

Remark 2.6: It should be observed that unlike Theorem 2.5 and results in the previous sections, for the validity of Theorem 2.6, it is not required to have a Markov condition on $P_{S_t, S^b_t, S^r_t}(s_t, s^b_t, s^r_t)$. Furthermore, the result also holds with no CSIR, i.e., $\mathcal{S}_r = \emptyset$ is allowed, and in this case Theorem 2.6 is as an extension of [20, Theorem 4] to a noisy setup.

Note that for the setup given in [20, Theorem 4], Theorem 2.6 provides an equivalent characterization. Recall that in [20, Theorem 4] the informed encoder has full CSI, i.e., $X^b_t = \phi^{(b)}_t(W_a, W_b, S_{[t]})$, both the uniformed encoder and the decoder have no CSI and the capacity region, $\mathcal{C}_{AS}$, is given as the closure of all rate pairs (R_a, R_b) satisfying

$$
\begin{aligned}
R_b & < I(U; Y|X^a) \qquad (56) \\
R_b + R_a & < I(U, X^a; Y) \qquad (57)
\end{aligned}
$$

for some joint measure on $\mathcal{S}\times\mathcal{X}_a\times\mathcal{X}_b\times\mathcal{Y}\times\mathcal{U}$ having the form

$$
\begin{aligned}
& P_{Y|X^a, X^b, S}(y|x^a, x^b, s) P_{X^b|U, X^a, S}(x^b|u, x^a, s) \\
& \times P_S(s) P_{X^a, U}(x^a, u), \qquad (58)
\end{aligned}
$$

where $|\mathcal{U}| \leq |\mathcal{S}||\mathcal{X}_a||\mathcal{X}_b| + 1$. On the other hand, for this setup, Theorem 2.6 gives the capacity region, $\mathcal{C}^G_{FS}$, as $\overline{co}\Big(\bigcup_{\hat{\pi}} \mathcal{R}'_C(\hat{\pi})\Big)$ where $\mathcal{R}'_C(\hat{\pi})$ denotes the region of all rate pairs $R = (R_a, R_b)$ satisfying

$$
\begin{aligned}
R_b & < I(T; Y|X^a) \qquad (59) \\
R_a + R_b & < I(T, X^a; Y) \qquad (60)
\end{aligned}
$$

where $P_{Y, T, X^a, X^b, S}(y, t, x^a, x^b, s)$ factorizes as

$$
P_{Y|X^a, X^b, S}(y|x^a, x^b, s) P_{X^b|S, T}(x^b|s, t) P_S(s) \hat{\pi}_{X^a, T}(x^a, t), \qquad (61)
$$

and $T : \mathcal{S} \to \mathcal{X}_b$.

Although the relation between an auxiliary variable and Shannon strategies is well understood for the single-user case (e.g., see [26, Section 3.2]), we believe that it requires more attention in the multi user case; in particular, note the difference between $|\mathcal{U}|$ and $|\mathcal{T}|$. Hence, we provide a proof for $\mathcal{C}^G_{FS} = \mathcal{C}_{AS}$, see Appendix D.

We conclude this section with the following remark.

Remark 2.7: For the validity of converse proof of Theorem 2.6 it is crucial that X^a_t only depends on W_a. To be more explicit, let us assume $\mathcal{S}_r = \emptyset$ and consider the following steps of the converse

$$
\begin{aligned}
& I(W_b; Y_{[n]}) \\
& \leq \sum_{t=1}^{n} H(Y_t|Y_{[t-1]}, X^a_{[n]}) - H(Y_t|Y_{[t-1]}, \mathbf{W}, X^a_{[n]}, T^b_t) \\
& = \sum_{t=1}^{n} H(Y_t|Y_{[t-1]}, X^a_{[n]}) - H(Y_t|Y_{[t-1]}, X^a_t, T^b_t). \qquad (62)
\end{aligned}
$$

Since S_t is not available to the decoder, the above equality is valid if $X^a_{[n]}$ does not provide any information about S_t. Hence, in other words, whether CSITs are noisy or not, if there is no CSI or noisy CSI at the decoder, the arguments above would fail if the uninformed encoder observes some degree of CSI, i.e., $d_a < \infty$ so that $X^a_{[n]}$ carries information about (S_t, S^b_t, S^r_t).

III. Examples

We present two examples. In the first example we discuss the state dependent modulo-additive MAC with noisy CSIT and complete CSIR (as in Section II-A) and show that the

proposed inner and outer bounds are tight and yield the capacity region. In the second example we consider the problem defined in Section II-B where the channel is a binary multiplier MAC whose state is an interference sequence.

A. Modulo-additive State-dependent MAC with Noisy CSIT

Recall that the results of Section II-A are given in terms of Shannon-strategies. Hence, their computation requires an optimization over an extended space of the input alphabet to a space of strategies and is often hard; in fact, very few explicit solutions exist even in the single-user case. In [5] modulo-additive single-user channel with complete CSIT is considered and a closed-form solution for the capacity is derived. Based on this result, we now consider the modulo-additive state-dependent MAC with asymmetric noisy CSIT and show that for the sum-rate capacity, the optimal set of strategies has uniform distribution. This enables us to determine the entire capacity region by observing that under the uniform distribution both inner and outer bounds are tight.

To be more explicit, we consider a two-user state-dependent MAC in which the channel noise, defined by a process $\{Z_t\}_{t=1}^{\infty}$, is correlated with the state process. The channel is given by $Y = X^a \oplus X^b \oplus Z$ where $\mathcal{X}_a = \mathcal{X}_b = \mathcal{Y} = \mathcal{Z} = \{0, \cdots, q-1\}$ and Z, is conditionally independent of (X^a, X^b) given the state S and in the sequel addition (and subtraction) is understood to be performed mod-q. Assume further that we have the setup of Section II-A. The following theorem is the main result of this example and can be though as an extension of [5, Theorem 1] to a noisy multi-user setting.

Theorem 3.1: The capacity region of the modulo-additive state-dependent MAC defined above is given by the closure of the rate pairs (R_a, R_b) satisfying

$$\begin{aligned} R_a &< \log q - H_{\min} \\ R_b &< \log q - H_{\min} \\ R_a + R_b &< \log q - H_{\min} \end{aligned} \tag{63}$$

where $H_{\min} := \min_{t^a, t^b} H(Z + t^a(S^a) + t^b(S^b)|S)$.

Proof: First, recall the rate condition given in Theorem 2.2;

$$R_a + R_b \quad \leq \quad H(Y|S) - H(Y|T^a, T^b, S). \tag{64}$$

The proof composed of two steps; we first determine the optimal distributions of t^a, t^b, the distributions achieving the sum-rate capacity, and then we show these distributions yield the same inner bound. Let us first consider $H(Y|T^a, T^b, S)$. Clearly, $P_{Y|X^a,X^b,S}(y|x^a, x^b, s) = P_{Z|S}(y - x^a - x^b|s)$ and $H(Y|T^a, T^b, S) \geq \min_{t^a,t^b} H(Y|T^a = t^a, T^b = t^b, S)$. Observe that

$$\begin{aligned} &P_{Y|T^a,T^b,S}(y|t^a, t^b, s) \\ &= \sum_{s^a,s^b} P_{Y|T^a,T^b,S^a,S^b,S}(y|t^a, t^b, s^a, s^b, s) P_{S^a,S^b|S}(s^a, s^b|s) \\ &= \sum_{s^a,s^b} P_{Z|S}(Z = y - t^a(s^a) - t^b(s^b)|s) P_{S^a,S^b|S}(s^a, s^b|s) \\ &= P_{Z+t^a(S^a)+t^b(S^b)|S}(y|s). \end{aligned} \tag{65}$$

where the second step is valid since Z is conditionally independent of (S^a, S^b) given S. Therefore, $H(Y|T^a = t^a, T^b = t^b, S) = H(Z + t^a(S^a) + t^b(S^b)|S)$. Let (t^{a*}, t^{b*}) be two mappings from $\mathcal{S}_a$ to $\mathcal{X}_a$ and $\mathcal{S}_b$ to $\mathcal{X}_b$, respectively, for which $H(Y|T^a = t^{a*}, T^b = t^{b*}, S) = H_{\min}$. Now recall that, by Corollary 2.1, we have

$$\begin{aligned} \mathcal{C}_{\Sigma}^{FS} &= \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} \left[H(Y|S) - H(Y|T^a, T^b, S)\right] \\ &\leq \sup_{\pi_{T^a}(t^a)\pi_{T^b}(t^b)} H(Y|S) - H_{\min}, \end{aligned} \tag{66}$$

and we now determine the policies $\{\pi_{T^a}(t^a),\ t^a \in \mathcal{T}_a\}$ and $\{\pi_{T^b}(t^b),\ t^b \in \mathcal{T}_b\}$ achieving the supremum above. Let us first define the following class of strategies

$$\mathcal{T}_a^* \quad := \quad \{t_\tau^a\};\ t_\tau^a(s^a) = t^{a*}(s^a) + \tau,\ \ \tau = 1, \cdots, q \tag{67}$$

$$\mathcal{T}_b^* \quad := \quad \{t_\tau^b\};\ t_\tau^b(s^b) = t^{b*}(s^b) - \tau,\ \ \tau = 1, \cdots, q. \tag{68}$$

Note that $H(Y|T^a = t^{a*}, T^b = t^{b*}, S) = H(Y|T^a = t_\tau^a, T^b = t_\tau^b, S)$ since $H(Y|T^a = t^a, T^b = t^b, S) = H(Z + t^a(S^a) + t^b(S^b)|S)$. Note that $H(Y|S) \leq \log|\mathcal{Y}| = \log q$, but if we choose T^a and T^b uniformly distributed within $\mathcal{T}_a^*$ and $\mathcal{T}_b^*$, respectively (with zero mass on strategies not in $\mathcal{T}_a^*$ and $\mathcal{T}_b^*$), we would get

$$\begin{aligned} &P_{Y|S}(y|s) \\ &\overset{(i)}{=} \sum_{s^a,s^b} \sum_{t^a \in \mathcal{T}_a^*} \sum_{t^b \in \mathcal{T}_b^*} P_{Y|T^b,T^b,S^a,S^b,S}(y|t^a, t^b, s^a, s^b, s) \\ &\qquad \frac{1}{q^2} P_{S^a,S^b|S}(s^a, s^b|s) \\ &= \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a, s^b|s) \frac{1}{q^2} \\ &\qquad \sum_{t^a \in \mathcal{T}_a^*} \sum_{t^b \in \mathcal{T}_b^*} P_{Z|S}(y - t^a(s^a) - t^b(s^b)|s) \\ &\overset{(ii)}{=} \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a, s^b|s) \frac{1}{q^2} \sum_{t^a \in \mathcal{T}_a^*} 1 \\ &\overset{(iii)}{=} \frac{1}{q} \end{aligned} \tag{69}$$

where (i) valid since T^a and T^b are uniformly distributed, (ii) is due to (68) (i.e., follows from the fact that $t^b \in \mathcal{T}_b^*$ traces all possible values of Z) and finally, (iii) is valid since $|\mathcal{T}_a^*| = q$. Therefore, we get that $\mathcal{C}_{FS}^{\Sigma} = \log q - H_{\min}$ which is achieved by

$$\pi_{T^a}(t^a) = \frac{1}{q},\ \ \forall t^a \in \mathcal{T}_a^*,\ \ \pi_{T^b}(t^b) = \frac{1}{q},\ \ \forall t^b \in \mathcal{T}_b^*. \tag{70}$$

Let us now consider the inner bound. In particular, we need to show that the sets of policies in (70) give $H(Y|T^a, S) = H(Y|T^b, S) = \log q$. Consider $H(Y|T^a, S)$ and observe that

$$\begin{aligned} &P_{Y|T^a,S}(y|t^a, s) \\ &\overset{(iv)}{=} \sum_{s^a,s^b} \sum_{t^b \in \mathcal{T}_b^*} P_{Y|T^b,T^b,S^a,S^b,S}(y|t^a, t^b, s^a, s^b, s) \\ &\qquad \frac{1}{q} P_{S^a,S^b|S}(s^a, s^b|s) \end{aligned}$$

$$= \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q} \sum_{t^b\in\mathcal{T}_b^*} P_{Z|S}(y-t^a(s^a)-t^b(s^b)|s)$$
$$\overset{(v)}{=} \sum_{s^a,s^b} P_{S^a,S^b|S}(s^a,s^b|s)\frac{1}{q}$$
$$= \frac{1}{q} \tag{71}$$

where (iv) is valid since T^b is uniformly distributed and (v) is due to (68) (i.e., follows from the fact that $t^b \in \mathcal{T}_b^*$ traces all possible values of Z). Thus, $H(Y|T^a,S) = \log q$. It can be shown similarly that under (70) $H(Y|T^b,S) = \log q$. ■

Finally, it is easy to see that when there is no side information at the encoders and at the decoder the capacity region of modulo-additive state-dependent MAC is given by the closure of rate pairs (R_a, R_b) where

$$\begin{aligned} R_a &\le \log q - H(Z) \\ R_b &\le \log q - H(Z) \\ R_a + R_b &\le \log q - H(Z). \end{aligned} \tag{72}$$

Observe that we have

$$H(Z+t^a(S^a)+t^b(S^b)|S) \le H(Z|S) + H(t^a(S^a)+t^b(S^b)|S)$$

and

$$\begin{aligned} H_{\min} &= \min_{t^a,t^b} H(Z+t^a(S^a)+t^b(S^b)|S) \\ &\le \min_{t^a,t^b} \left[H(Z|S) + H(t^a(S^a)+t^b(S^b)|S)\right] \\ &\overset{(vi)}{=} H(Z|S) \\ &\overset{(vii)}{<} H(Z) \end{aligned}$$

where (vi) can be achieved with any deterministic mapping and (vii) is valid since Z and S (and hence S) are correlated. Therefore, availability of state information strictly increases, by an amount of at least $I(S;Z)$, the capacity region of the modulo-additive state-dependent MAC.

B. Binary Multiplier State-dependent MAC with Interference

Consider the binary multiplier MAC with state process interfering the output, namely $Y = X^aX^b \oplus S$ where $\mathcal{X}_a = \mathcal{X}_b = \mathcal{Y} = \mathcal{S} = \{0,1\}$. Assume further that the communication setup is given as in Section II-B with $S^r = S \oplus Z^r$ where $Z^r \sim \text{Ber}(p_r)$ is Bernoulli with $P(Z^r=1) = p_r$. Clearly, in this setup we have

$$\begin{aligned} &P_{S_t^a,S_t^b,S_t^r,S_t}(s_t^a,s_t^b,s_t^r,s_t) \\ &\quad = P_{S_t^a|S_t^r}(s_t^a|s_t^r)P_{S_t^b|S_t^r}(s_t^b|s_t^r)P_{S_t,S_t^r}(s_t,s_t^r). \end{aligned} \tag{73}$$

We now show that the capacity region, with both causal and non-causal coding, of this channel is given by the closure of (R_a, R_b) where $R_a < 1 - H(S|S^r)$, $R_b < 1 - H(S|S^r)$ and $R_a + R_b < 1 - H(S|S^r)$.

First recall the capacity region given in Theorem 2.3 and observe that $H(Y|S^r,X^a,X^b) = H(X^aX^b \oplus S|S^r,X^a,X^b) = H(S|S^r,X^a,X^b) = H(S|S^r)$, where the last equality follows from (73). Hence, input distributions do not effect $H(Y|S^r,X^a,X^b)$. Clearly, $H(Y|S^r) \le 1$, $H(Y|S^r,X^a) \le 1$ and $H(Y|S^r,X^b) \le 1$ and we now show that equalities can be achieved. More explicitly, we have the following optimizing distributions which can be obtained using basic inequalities

$$\begin{aligned} &\underset{\pi_{X^a|S^a}(x^a|f^a(s^r)),\pi_{X^b|S^b}(x^b|f^b(s^r))}{\text{argmax}} H(Y|S^r) \\ &= \big\{\pi_{X^a|S^a}(0|f^a(0)) = \pi_{X^a|S^a}(0|f^a(1)) = 0.5, \\ &\qquad \pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\}, \end{aligned} \tag{74}$$

$$\begin{aligned} &\underset{\pi_{X^a|S^a}(x^a|f^a(s^r)),\pi_{X^b|S^b}(x^b|f^b(s^r))}{\text{argmax}} H(Y|S^r,X^a) \\ &= \big\{\pi_{X^a|S^a}(0|f^a(0)) = \pi_{X^a|S^a}(0|f^a(1)) = 0, \\ &\qquad \pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\}, \end{aligned} \tag{75}$$

$$\begin{aligned} &\underset{\pi_{X^a|S^a}(x^a|f^a(s^r)),\pi_{X^b|S^b}(x^b|f^b(s^r))}{\text{argmax}} H(Y|S^r,X^b) \\ &= \big\{\pi_{X^b|S^b}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0, \\ &\qquad \pi_{X^a|S^a}(0|f^b(0)) = \pi_{X^b|S^b}(0|f^b(1)) = 0.5\big\} \end{aligned} \tag{76}$$

and in the rest, let us show that these yield the equalities in the conditional entropies. Let us start with R_a, i.e., $H(Y|S^r,X^b)$. Note that

$$\begin{aligned} H(Y|S^r,X^b) = &\sum_{s^r\in\{0,1\}}\sum_{x^b\in\{0,1\}} P_{S^r}(s^r)\pi_{X^b|S^b}(x^b|f^b(s^r)) \\ &\times H(Y|S^r=s^r,X^b=x^b). \end{aligned} \tag{77}$$

Substituting (76) in (77) gives

$$\begin{aligned} H(Y|S^r,X^b) = &P_{S^r}(0)H(X^a\oplus S|X^b=1,S^r=0) \\ &+P_{S^r}(1)H(X^a\oplus S|X^b=1,S^r=1). \end{aligned} \tag{78}$$

We next show that under (76) $H(X^a\oplus S|X^b=1,S^r=0) = 1$, for which it is enough to show that $P_{X^a\oplus S|X^b,S^r}(0|1,0) = 0.5$. We have

$$\begin{aligned} &P_{X^a\oplus S|X^b,S^r}(0|1,0) \\ &\quad = \sum_{s\in\{0,1\}}\sum_{x^a\in\{0,1\}} P_{X^a\oplus S|S,X^a,X^b,S^r}(0|s,x^a,1,0) \\ &\qquad P_{S|S^r}(s|0)\pi_{X^a|S^a}(x^a|f^a(0)) \\ &\quad = P_{S|S^r}(0|1)\left[0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|0,0,1,0)+\right. \\ &\qquad \left.0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|0,1,1,0)\right] \\ &\qquad +P_{S|S^r}(1|1)\left[0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|1,0,1,0)\right. \\ &\qquad \left.+0.5P_{X^a\oplus S|S,X^a,X^b,S^r}(0|1,1,1,0)\right] \\ &\quad = 0.5, \end{aligned} \tag{79}$$

where (79) is due to (73) and (32). We can similarly show that $P_{X^a\oplus S|X^b,S^r}(0|1,1) = 0.5$ and hence, $H(X^a\oplus S|X^b = 1, S^r = 1) = 1$. Therefore, $H(Y|S^r,X^b) = 1$. Since the above derivation is symmetric, under (75) $H(Y|X^a,S^r) = 1$.

It now remains to show that with (74), $H(Y|S^r)$ is equal to one. It should be observed that

$$\begin{aligned} &P_{X^aX^b\oplus S|S^r}(\cdot|s^r) \\ &\quad \overset{(i)}{=} \sum_{x^a,x^b,s\in\{0,1\}} P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s) \end{aligned}$$

$$\pi_{X^a|S^a}(x^a|f^a(s^r))\pi_{X^b|S^b}(x^b|f^b(s^r))P_{S|S^r}(s|s^r)$$
$$\overset{(ii)}{=} 0.25 \sum_{x^a,x^b\{0,1\}} P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s) \sum_{s\in\{0,1\}} P_{S|S^r}(s|s^r)$$
$$= 0.5$$

where (i) is due to (73) and (32), (ii) is due to (74) and the last step is valid since for given s, there are only two pairs of (x^a, x^b) for which $P_{X^aX^b\oplus S|X^a,X^b,S}(\cdot|x^a,x^b,s) = 1$ (and zero for the other twos). Hence, $H(Y|S^r) = 1$.

Finally, it can be easily shown that the capacity region of $Y = X^aX^b \oplus S$ without CSIT and CSIR is given by the closure of (R_a, R_b) where $R_a < 1 - H(S)$, $R_b < 1 - H(S)$ and $R_a + R_b < 1 - H(S)$. Therefore, availability of noisy CSI at the encoders (both causal and non-causal) and at the decoder increases the capacity region by an amount of $I(S; S^r)$.

IV. Conclusion and Remarks

We have considered several scenarios for the memoryless state-dependent MAC with an i.i.d. state process, asymmetric noisy CSI at the encoders and complete and noisy CSI at the receiver. When the encoders have access to causal noisy CSI, single-letter inner and outer bounds, which are tight for the sum-rate capacity, are obtained. In order to reduce the space of optimization, from Shannon strategies to channel inputs, we consider the case where CSITs are asymmetric deterministic functions of noisy CSIR. The causal setup of this problem is considered in [13] and a single-letter characterization for capacity region is provided. Hence, we also considered the non-causal setup and showed that the causal and non-causal capacity regions are identical.

When the decoder does not have access to access the current CSI at the encoder, which matches with the delayed scenario, we observe that a single-letter characterization of the capacity region can be obtained. We further discuss a degraded message set scenario and show that when the common message encoder does not have access to the current noisy CSI, due to delay, it is possible to obtain a single-letter expression for the capacity region. Since a product form is not required in this case, we observed that as long as the common message encoder does not have access to CSI, then in any noisy setup (the cases of no CSIR or noisy CSIR) it is possible to obtain the capacity region.

Finally, the following problems are worth exploring in the future: the complete characterization of the capacity region for the problem defined in Section II-A and its non-causal extension, the state-dependent MAC with degraded message set where either both encoders observe causal noisy CSI or the informed encoder observes noisy CSI non-causally while the other encoder observes noisy CSI with delay.

Appendix A
Converse Proof of Theorem 2.3: Non-Causal Case

Let

$$\alpha_{\mu_{\mathbf{p},\mathbf{f}}} := \frac{1}{n} P_{S^r_{[1,t-1]},S^r_{[t+1,n]}}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}). \tag{80}$$

Observe that $(\mu_{\mathbf{p}} : \mu_{\mathbf{f}}) \in \mathcal{S}_r^{n-1}$, where $(v : w)$ denotes the concatenation of two vectors v and w, and

$$\sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} := \frac{1}{n} \sum_{1\le t\le n} \sum_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}} P_{S^r_{[1,t-1]},S^r_{[t+1,n]}}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) = 1.$$

Lemma A.1: Assume that a rate pair $R = (R_a, R_b)$, with block length $n \ge 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Let $\Theta_t^a(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) = I(X_t^a; Y_t|X_t^b, S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}})$, $\Theta_t^b(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) = I(X_t^b; Y_t|X_t^a, S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}})$ and $\Theta_t^{a,b}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) = I(X_t^a, X_t^b; Y_t|S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}})$. Then,

$$R_a \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} \Theta_t^a(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) + \eta(\epsilon) \tag{81}$$
$$R_b \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} \Theta_t^b(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) + \eta(\epsilon) \tag{82}$$
$$R_a + R_b \le \sum_{(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} \Theta_t^{a,b}(\mu_{\mathbf{p}}, \mu_{\mathbf{f}}) + \eta(\epsilon) \tag{83}$$

Proof: Let us first consider the sum-rate. With standard steps, we get

$$R_a + R_b \le \frac{1}{1-\epsilon}\frac{1}{n}\left(I(\mathbf{W}; Y_{[n]}, S_{[n]}^r) + H(\epsilon)\right). \tag{84}$$

Note that since $S_{[n]}^r$ is independent of $\mathbf{W}$; $I(\mathbf{W}; Y_{[n]}, S_{[n]}^r) = I(\mathbf{W}; Y_{[n]}|S_{[n]}^r)$ and

$$\begin{aligned}
&I(\mathbf{W}; Y_{[n]}|S_{[n]}^r) \\
&= \sum_{t=1}^n \left[H(Y_t|S_{[n]}^r, Y_{[t-1]}) - H(Y_t|\mathbf{W}, S_{[n]}^r, Y_{[t-1]})\right] \\
&\overset{(i)}{\le} \sum_{t=1}^n \left[H(Y_t|S_{[n]}^r) - H(Y_t|\mathbf{W}, S_{[n]}^r, Y_{[t-1]})\right] \\
&\overset{(ii)}{=} \sum_{t=1}^n \left[H(Y_t|S_{[n]}^r) - H(Y_t|\mathbf{W}, S_{[n]}^r, Y_{[t-1]}, \mathbf{X}_{[n]})\right] \\
&\overset{(iii)}{=} \sum_{t=1}^n \left[H(Y_t|S_{[n]}^r) - H(Y_t|S_{[n]}^r, \mathbf{X}_t)\right] \\
&= \sum_{t=1}^n I(\mathbf{X}_t; Y_t|S_{[n]}^r)
\end{aligned} \tag{85}$$

where (i) holds conditioning does not increase entropy, (ii) holds since $X_t^i = \phi_t^{(i)}(W_i, f^i(S_{[n]}^r))$, $i = \{a, b\}$, and (iii) is due to (3). Combining (84) and (85) similar to (17), gives

$$R_a + R_b \le \frac{1}{n}\sum_{t=1}^n I(X_t^a, X_t^b; Y_t|S_{[n]}^r) + \eta(\epsilon) \tag{86}$$

Furthermore,

$$I(X_t^a, X_t^b; Y_t|S_{[n]}^r) = n \times \sum_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}} \alpha_{\mu_{\mathbf{p},\mathbf{f}}} I(X_t^a, X_t^b; Y_t|S_t^r, S_{[t-1]}^r = \mu_{\mathbf{p}}, S_{[t+1,n]}^r = \mu_{\mathbf{f}}) \tag{87}$$

and substituting the above into (86) yields (83).

Let us now consider encoder a. Using Fano's inequality and standard steps we first get,

$$R_a \le \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_a; Y_{[n]}, S_{[n]}^r) + H(\epsilon)\right). \tag{88}$$

Furthermore,

$$\begin{aligned}
&I(W_a;Y_{[n]},S^r_{[n]})\\
&\overset{(i)}{\leq} I(W_a;Y_{[n]}|S^r_{[n]},W_b)\\
&= \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[n]},Y_{[t-1]},W_b)-H(Y_t|S^r_{[n]},Y_{[t-1]},\mathbf{W})\Big]\\
&\overset{(ii)}{\leq} \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[n]},W_b)-H(Y_t|S^r_{[n]},Y_{[t-1]},\mathbf{W})\Big]\\
&\overset{(iii)}{=} \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[n]},W_b,X^b_{[n]})\\
&\qquad -H(Y_t|S^r_{[n]},Y_{[t-1]},\mathbf{W},\mathbf{X}_{[n]})\Big]\\
&\overset{(iv)}{\leq} \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[n]},X^b_t)-H(Y_t|S^r_{[n]},Y_{[t-1]},\mathbf{W},\mathbf{X}_{[n]})\Big]\\
&\overset{(v)}{=} \sum_{t=1}^{n}\Big[H(Y_t|S^r_{[n]},X^b_t)-H(Y_t|S^r_{[n]},X^b_t,X^a_t)\Big]\\
&= \sum_{t=1}^{n} I(X^a_t;Y_t|X^b_t,S^r_{[n]}) \qquad (89)
\end{aligned}$$

where (i) is due to (2) and conditioning does not increase entropy, (ii) holds since conditioning does not increase entropy, (iii) holds since $X^i_t=\phi^{(i)}(W_i,f^i(S^r_{[n]}))$, $i=\{a,b\}$, (iv) is valid since conditioning does not increase entropy and finally, (v) is valid due to (3) and S^i_t, $i=\{a,b\}$, being a function of S^r_t.

Now combining (88)-(89) and following steps akin to (86) and (87), we can verify (81). To verify (82) for encoder b it is enough to switch the roles of encoder a and (b). ■

Note that for any for any $t\geq 1$, $I(X^a_t;Y_t|X^b_t,S^r_t,S^r_{[t-1]}=\mu_{\mathbf{p}},S^r_{[t+1,n]}=\mu_{\mathbf{f}})$, $I(X^b_t;Y_t|X^a_t,S^r_t,S^r_{[t-1]}=\mu_{\mathbf{p}},S^r_{[t+1,n]}=\mu_{\mathbf{f}})$ and $I(X^a_t,X^b_t;Y_t|S^r_t,S^r_{[t-1]}=\mu_{\mathbf{p}},S^r_{[t+1,n]}=\mu_{\mathbf{f}})$ are functions of $P_{X^a_t,X^b_t,Y_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(x^a_t,x^b_t,y_t,s^r_t|\mu_{\mathbf{p}},\mu_{\mathbf{f}})$ Hence, we need to show that this distribution factorizes as in (36). Let

$$\begin{aligned}
&\Upsilon^a_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^a,f^a(s^r)) :=\\
&\qquad \{w_a:\phi^{(a)}_t\left(w_a,f^a(\mu_{\mathbf{p}},\mu_{\mathbf{f}}),f^a(s^r)\right)=x^a\},\\
&\Upsilon^b_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^b,f^b(s^r)) :=\\
&\qquad \{w_b:\phi^{(b)}_t\left(w_b,f^b(\mu_{\mathbf{p}},\mu_{\mathbf{f}}),f^b(s^r)\right)=x^b\} \qquad (90)
\end{aligned}$$

and

$$\begin{aligned}
\pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^a|S^a}\left(x^a|f^a(s^r)\right) &:= \sum_{w_a\in\Upsilon^a_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^a,f^a(s^r))}\frac{1}{|\mathcal{W}_a|},\\
\pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^b|S^b}\left(x^b|f^b(s^r)\right) &:= \sum_{w_b\in\Upsilon^b_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^b,f^b(s^r))}\frac{1}{|\mathcal{W}_b|}. \qquad (91)
\end{aligned}$$

Lemma A.2: For every $1\leq t\leq n$ and $(\mu_{\mathbf{p}}:\mu_{\mathbf{f}})\in\mathcal{S}^{n-1}_r$, the following holds

$$\begin{aligned}
&P_{X^a_t,X^b_t,Y_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(x^a,x^b,y,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}})\\
&\quad = P_{S^r}(s^r)P_{Y|S^r,X^a,X^b}(y|s^r,x^a,x^b)\\
&\qquad \pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^a|S^a}(x^a|f^a(s^r))\pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^b|S^b}(x^b|f^b(s^r)). \qquad (92)
\end{aligned}$$

Proof: First observe that due to (3) we have

$$\begin{aligned}
&P_{X^a_t,X^b_t,Y_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(x^a,x^b,y,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}})\\
&\quad = P_{Y_t|S^r_t,X^a_t,X^b_t}(y|s^r,x^a,x^b)\\
&\qquad P_{X^a_t,X^b_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(x^a,x^b,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}}). \qquad (93)
\end{aligned}$$

Let us now consider the second term in (93). We have

$$\begin{aligned}
&P_{X^a_t,X^b_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(x^a,x^b,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}})\\
&\quad = \sum_{w_a,w_b}P_{\mathbf{W},X^a_t,X^b_t,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(\mathbf{w},x^a,x^b,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}})\\
&\quad \overset{(i)}{=} \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}1_{\left\{x^l=\phi^{(l)}(w_l,f^l(s^r,\mu_{\mathbf{p}},\mu_{\mathbf{f}})),\ l=a,b\right\}}\\
&\qquad P_{W_a,W_b,S^r_t|S^r_{[t-1]},S^r_{[t+1,n]}}(w_a,w_b,s^r|\mu_{\mathbf{p}},\mu_{\mathbf{f}})\\
&\quad \overset{(ii)}{=} \sum_{w_a\in\mathcal{W}_a}\sum_{w_b\in\mathcal{W}_b}1_{\left\{x^l=\phi^{(l)}(w_l,f^l(s^r,\mu_{\mathbf{p}},\mu_{\mathbf{f}})),\ l=a,b\right\}}\\
&\qquad \frac{1}{|\mathcal{W}_a|}\frac{1}{|\mathcal{W}_b|}P_{S^r_t}(s^r)\\
&\quad = P_{S^r_t}(s^r)\sum_{w_a\in\mathcal{W}_a}\frac{1}{|\mathcal{W}_a|}1_{\left\{x^a=\phi^{(a)}(w_a,f^a(s^r,\mu_{\mathbf{p}},\mu_{\mathbf{f}}))\right\}}\\
&\qquad \sum_{w_b\in\mathcal{W}_b}\frac{1}{|\mathcal{W}_b|}1_{\left\{x^b=\phi^{(b)}(w_b,f^b(s^r,\mu_{\mathbf{p}},\mu_{\mathbf{f}}))\right\}}\\
&\quad \overset{(iii)}{=} \frac{1}{|\mathcal{W}_a|}\sum_{w_a\in\Upsilon^a_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^a,f^a(s^r))}\frac{1}{|\mathcal{W}_b|}\sum_{w_b\in\Upsilon^b_{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}(x^b,f^b(s^r))}\\
&\quad \overset{(iv)}{=} P_{S^r_t}(s^r)\pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^a|S^a}(x^a|f^a(s^r))\pi^{\mu_{\mathbf{p}},\mu_{\mathbf{f}}}_{X^b|S^b}(x^b|f^b(s^r)) \qquad (94)
\end{aligned}$$

where (i) follows since $X^i_t=\phi^{(i)}(W_i,f^i(S^r_{[n]}))$, $i=\{a,b\}$, (ii) is valid since W_a and W_b are independent of $S^r_{[n]}$ and state process being i.i.d. and (iii) follows due to (90) and (iv) follows due to (91). Substituting (94) in (93) completes the proof. ■

We can now complete the proof of Theorem 2.3. With Lemma A.1, it is shown that any achievable rate pair can be approximated by the convex combinations of rate conditions given in (33)-(35) which are indexed by $(\mu_{\mathbf{p}},\mu_{\mathbf{f}})$ and satisfy (36) for joint state-input-output distributions. Hence, since $\lim_{\epsilon\to 0}\eta(\epsilon)=0$, any achievable rate pair belongs to $\overline{co}\Big(\bigcup_{\bar{\pi}}\mathcal{R}^Q_{NS}(\bar{\pi})\Big)$.

Appendix B
Converse Proof of Theorem 2.4

Recall that α_μ is defined in (80).

Lemma B.1: Assume that a rate pair $R=(R_a,R_b)$, with block length $n\geq 1$ and a constant $\epsilon\in(0,1/2)$, is achievable. Then,

$$\begin{aligned}
R_a &\leq \sum_{\mu\in\mathcal{S}^{(n)}}\alpha_\mu I(X^a_t;Y_t|X^b_t,S_t,S_{[t-1]}=\mu)+\eta(\epsilon) \qquad (95)\\
R_b &\leq \sum_{\mu\in\mathcal{S}^{(n)}}\alpha_\mu I(X^b_t;Y_t|X^a_t,S_t,S_{[t-1]}=\mu)+\eta(\epsilon) \qquad (96)\\
R_a+R_b &\leq \sum_{\mu\in\mathcal{S}^{(n)}}\alpha_\mu I(X^a_t,X^b_t;Y_t|S_t,S_{[t-1]}=\mu)+\eta(\epsilon).
\end{aligned}$$

(97)

Proof:

Let us now consider encoder a. We have

$$\begin{aligned} R_a &\leq \frac{1}{n}\log(|\mathcal{W}_a|) \\ &\leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_a;Y_{[n]},S_{[n]})+H(\epsilon)\right). \end{aligned} \tag{98}$$

Furthermore,

$$\begin{aligned} &I(W_a;Y_{[n]},S_{[n]}) \\ &\overset{(i)}{\leq} I(W_a;Y_{[n]},S_{[n]}|W_b,S^b_{[n]}) \\ &= \sum_{t=1}^{n}\Big[H(Y_t,S_t|S_{[t-1]},Y_{[t-1]},W_b,S^b_{[n]}) \\ &\qquad -H(Y_t,S_t|S_{[t-1]},Y_{[t-1]},\mathbf{W},S^b_{[n]})\Big] \\ &\overset{(ii)}{=} \sum_{t=1}^{n}\Big[H(Y_t,S_t|S_{[t-1]},Y_{[t-1]},W_b,S^b_{[n]},X^b_{[n]}) \\ &\qquad -H(Y_t,S_t|S_{[t-1]},Y_{[t-1]},\mathbf{W},S^b_{[n]},X^b_{[n]})\Big] \\ &\overset{(iii)}{\leq} \sum_{t=1}^{n}\big[H(Y_t,S_t|S_{[t-1]},X^b_t) \\ &\qquad -H(Y_t,S_t|S_{[t-1]},Y_{[t-1]},\mathbf{W},S^b_{[n]},X^b_{[n]},X^a_{[n]})\big] \\ &\overset{(iv)}{=} \sum_{t=1}^{n}\big[H(Y_t,S_t|S_{[t-1]},X^b_t) \\ &\qquad -H(Y_t,S_t|S_{[t-1]},X^b_t,X^a_t)\big] \\ &= \sum_{t=1}^{n} I(X^a_t;Y_t|X^b_t,S_{[t]}) \end{aligned} \tag{99}$$

where (i) is due to (2) and conditioning does not increase entropy, (ii) is valid since $X^b_t=\phi^{(b)}_t\left(W_b,S^b_{[t-d_b]}\right)$, (iii) is valid since conditioning does not increase entropy and finally, (iv) is valid by (3). Following similar steps such as (17) and (18) verifies (95). Finally, (96) and (97) can be verified similarly. ■

Lemma B.2: For every $1\leq t\leq n$ and $\mu\in\mathcal{S}^{t-1}$, the following holds

$$\begin{aligned} &P_{X^a_t,X^b_t,Y_t,S_t|S_{[t-1]}}(x^a,x^b,y,s|\mu) \\ &\quad = P_S(s)P_{Y|S,X^a,X^b}(y|s,x^a,x^b)\pi^{\mu}_{X^a}(x^a)\pi^{\mu}_{X^b}(x^b). \end{aligned} \tag{100}$$

Let

$$\Upsilon^i_{\mu_{\mathbf{i}}}(x^i):=\{w_i:\phi^{(i)}_t(w_i,s^i_{[t-d_i]}=\mu_{\mathbf{i}})=x^i\},\ \ i=a,b \tag{101}$$

and

$$\begin{aligned} \pi^{\mu_{\mathbf{i}}}_{X^i}(x^i) &:= \sum_{w_i\in\Upsilon^i_{\mu_{\mathbf{i}}}(x^i)}\frac{1}{|\mathcal{W}_i|}, \\ \pi^{\mu}_{X^i}(x^i) &:= \sum_{\mu_{\mathbf{i}}}\pi^{\mu_{\mathbf{i}}}_{X^i}(x^i)P_{S^i_{[t-d_i]}|S_{[t-1]}}(\mu_{\mathbf{i}}|\mu),\ \ i=a,b. \end{aligned}$$

We can now verify (100) by following the same steps in Lemma 2.2.

Lemmas B.1 and B.2 complete the proof of converse.

Appendix C
Achievability and Converse Proofs of Theorem 2.5

Achievability Proof: Fix $(R_a,R_b)\in\mathcal{R}_C(\hat{\pi})$.

Codebook Generation Fix $\pi_{X^a}(x^a)$ and $\pi_{T^b|X^a}(t^b|x^a)$. For each $w_a\in\{1,\cdots,2^{nR_a}\}$, randomly generate $x^a_{[n],w_a}$, each according to $\prod_{i=1}^n\pi_{X^a_i}(x^a_{i,w_a})$. Reveal this codebook to encoder b and, for each $w_a\in\{1,\cdots,2^{nR_a}\}$ and $w_b\in\{1,\cdots,2^{nR_b}\}$, encoder b randomly generates $t^b_{[n],w_b,w_a}$, each according to $\prod_{i=1}^n\pi_{T^b_i|X^a_i}(t^b_{i,w_b}|x^a_{i,w_a})$. These codeword pairs form the codebook, which is revealed to the decoder.

Encoding The encoding functions are defined as follows: $x^a_i(w_a)=\phi^a_i(w_a,s^a_{[i-d_a]})$ and $x^b_i(w_b)=\phi^b_i(w_b,w_a,s^b_{[i]})=t^b_{i,w_b,w_a}(s^b_i)$ where x^a_{i,w_a} and t^b_{i,w_b,w_a} denote the ith component of $x^a_{[n],w_a}$ and $t^b_{[n],w_b,w_a}$, respectively. Therefore, to send the messages w_a and w_b, transmit the corresponding $x^a_{[n],w_a}$ and $t^b_{[n],w_b,w_a}$, respectively.

Decoding After receiving $(y_{[n]},s_{[n]})$, the decoder looks for the only (w_a,w_b) pair such that $(x^a_{[n],w_a},t^b_{[n],w_b},\ y_{[n]},s_{[n]})$ are jointly $\epsilon-$typical and declares this pair as its estimate $(\hat{w}_a,\hat{w}_b)$.

Error Analysis Let $E_{\alpha,\beta}\overset{\triangle}{=}\{(X^a_{[n],\alpha},T^b_{[n],\beta,\alpha},Y_{[n]},S_{[n]})\in A^n_\epsilon\}$, $\alpha\in\{1,\cdots,2^{nR_a}\}$ and $\beta\in\{1,\cdots,2^{nR_b}\}$ and assume that $(w_a,w_b)=(1,1)$ was sent. Then

$$\begin{aligned} P^n_e &= P\big(E^c_{1,1}\bigcup_{(\alpha,\beta)\neq(1,1)}E_{\alpha,\beta}\big) \\ &\leq P(E^c_{1,1})+\sum_{\alpha=1,\beta\neq1}P(E_{\alpha,\beta})+\sum_{\alpha\neq1,\beta=1}P(E_{\alpha,\beta}) \\ &\qquad +\sum_{\alpha\neq1,\beta\neq1}P(E_{\alpha,\beta}). \end{aligned} \tag{102}$$

Since $\{Y_i,S_i,X^a_i,T^b_i\}_{i=1}^\infty$ is i.i.d., $P(E^c_{1,1})\to0$ for $n\to\infty$. Next, let us consider the second term

$$\begin{aligned} &\sum_{\alpha=1,\beta\neq1}P(E_{\alpha=1,\beta\neq1}) \\ &\quad = \sum_{\alpha=1,\beta\neq1}P((X^a_{[n],1},T^b_{[n],\beta},Y_{[n]},S_{[n]})\in A^n_\epsilon) \\ &\quad \overset{(i)}{=} \sum_{\alpha=1,\beta\neq1}\ \sum_{(x^a_{[n]},t^b_{[n]},y_{[n]},s_{[n]})\in A^n_\epsilon}P_{T^b_{[n]}|X^a_{[n]}}(t^b_{[n]}|x^a_{[n]}) \\ &\qquad P_{X^a_{[n]},Y_{[n]},S_{[n]}}(x^a_{[n]},y_{[n]},s_{[n]}) \\ &\quad \leq \sum_{\alpha=1,\beta\neq1}|A^n_\epsilon|2^{-n[H(T^b|X^a)-\epsilon]}2^{-n[H(X^a,Y,S)-\epsilon]} \\ &\quad \leq 2^{nR_b}2^{-n[H(T^b|X^a)+H(X^a,Y,S)-H(X^a,T^b,Y,S)-3\epsilon]} \\ &\quad \overset{(ii)}{=} 2^{n[R_b-I(T^b;Y|S,X^a)-3\epsilon]} \end{aligned} \tag{103}$$

where (i) holds since $T^b_{[n],\beta}$ is independent of $(Y_{[n]},S_{[n]})$ given $X^a_{[n],1}$ and (ii) follows since

$$\begin{aligned} &H(T^b|X^a)+H(X^a,Y,S)-H(X^a,T^b,Y,S) \\ &\quad = H(T^b|X^a)+H(X^a,Y,S) \\ &\qquad -H(Y|X^a,T^b,S)-H(X^a,T^b,S) \\ &\quad = H(X^a,Y,S)-H(Y|X^a,T^b,S)-H(X^a,S) \\ &\quad = I(T^b;Y|S,X^a) \end{aligned}$$

where the second equality follows since T^b and S are independent given X^a. Next,

$$\begin{aligned}
&\sum_{\alpha\neq 1,\beta\neq 1} P(E_{\alpha\neq 1,\beta\neq 1}) \\
&= \sum_{\alpha\neq 1,\beta\neq 1} P((X^a_{[n],\alpha}, T^b_{[n],\beta}, Y_{[n]}, S_{[n]}) \in A^n_\epsilon) \\
&\overset{(iii)}{=} \sum_{\alpha\neq 1,\beta\neq 1} \sum_{(x^a_{[n]},t^b_{[n]},y_{[n]},s_{[n]})\in A^n_\epsilon} P_{T^b_{[n]},X^a_{[n]}}(t^b_{[n]}, x^a_{[n]}) \\
&\qquad P_{Y_{[n]},S_{[n]}}(y_{[n]}, s_{[n]}) \\
&\leq \sum_{\alpha\neq 1,\beta\neq 1} |A^n_\epsilon| 2^{-n[H(T^b,X^a)-\epsilon]} 2^{-n[H(Y,S)-\epsilon]} \\
&\leq 2^{n(R_a+R_b)} 2^{-n[H(T^b,X^a)+H(Y,S)-H(X^a,T^b,Y,S)-3\epsilon]} \\
&\overset{(iv)}{=} 2^{n[R_a+R_b-I(X^a,T^b;Y|S)-3\epsilon]}
\end{aligned} \tag{104}$$

where (iii) holds since for $\alpha,\beta \neq 1$, $(T^b_{[n],\beta}, X^a_{[n],\alpha})$ is independent of $(Y_{[n]}, S_{[n]})$ and (iv) follows since

$$\begin{aligned}
&H(T^b, X^a) + H(Y,S) - H(X^a, T^b, Y, S) \\
&= H(T^b, X^a) + H(Y,S) \\
&\qquad -H(Y|X^a, S, T^b) - H(X^a, S, T^b) \\
&= H(T^b, X^a) + H(Y,S) - H(Y|X^a, S, T^b) \\
&\qquad -H(X^a, T^b) - H(S) \\
&= I(X^a, T^b; Y|S),
\end{aligned}$$

and the rate conditions of the $\mathcal{R}_C(\hat{\pi})$ imply that each term tends in (102) tends to zero as $n \to \infty$. Finally, observe that the analysis for the error event $\sum_{\alpha\neq 1,\beta=1} P(E_{\alpha,\beta})$ is identical to the case of $\sum_{\alpha\neq 1,\beta\neq 1} P(E_{\alpha,\beta})$ which induces the same sum-rate constraint. ■

Note that the main motivation in indexing mutual information terms by the past CSI, is to get a product form on the team policies. In the degraded message set setup, we do not require a product form and therefore, the convex combination argument is not essential. However, we herein keep this indexing (see (46)) to avoid the use of a time sharing auxiliary random variable.

Converse Proof: Since $X^b_t = \phi^{(b)}_t\left(W_a, W_b, S^b_{[t-1]}, S^b_t\right)$, we have

$$T^b_t = \phi^{(b)}_t\left(W_a, W_b, S^b_{[t-1]}\right) \in \mathcal{X}_b^{|\mathcal{S}_b|}. \tag{105}$$

Lemma C.1: Let $T^b_t \in \mathcal{T}_b$ be the Shannon strategy induced by $\phi^{(b)}_t$ as shown in (105). Assume that a rate pair $R = (R_a, R_b)$, with block length $n \geq 1$ and a constant $\epsilon \in (0, 1/2)$, is achievable. Then,

$$R_b \leq \sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu I(T^b_t; Y_t|X^a_t, S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \tag{106}$$

$$R_a + R_b \leq \sum_{\mu\in\mathcal{S}^{(n)}} \alpha_\mu I(X^a_t, T^b_t; Y_t|S_t, S_{[t-1]} = \mu) + \eta(\epsilon) \tag{107}$$

where α_μ and $\eta(\epsilon)$ are defined in (9).

Proof: Let us first consider the sum-rate condition. Since,

$$\begin{aligned}
&I(\mathbf{W}; Y_{[n]}, S_{[n]}) \\
&\leq \sum_{t=1}^n \left[H(Y_t|S_{[t]}) - H(Y_t|\mathbf{W}, S_{[t]}, Y_{[t-1]}, X^a_t, T^b_t)\right] \\
&\overset{(i)}{=} \sum_{t=1}^n \left[H(Y_t|S_{[t]}) - H(Y_t|S_{[t]}, X^a_t, T^b_t)\right] \\
&= \sum_{t=1}^n I(X^a_t, T^b_t; Y_t|S_{[t]}),
\end{aligned} \tag{108}$$

where (i) can be shown in a similar way as (16), we have,

$$R_a + R_b \leq \frac{1}{n}\sum_{t=1}^n I(X^a_t, T^b_t; Y_t|S_{[t]}) + \eta(\epsilon) \tag{109}$$

and

$$I(X^a_t, T^b_t; Y_t|S_{[t]}) = n\sum_{\mu\in\mathcal{S}^{t-1}} \alpha_\mu I(X^a_t, T^b_t; Y_t|S_t, S_{[t-1]} = \mu). \tag{110}$$

Substituting the above into (109) yields (107).

Let us now consider encoder b. With Fano's inequality and standard steps, we get

$$R_b \leq \frac{1}{1-\epsilon}\frac{1}{n}\left(I(W_b; Y_{[n]}, S_{[n]}) + H(\epsilon)\right). \tag{111}$$

Following similar reasonings as in (99) we get,

$$\begin{aligned}
&I(W_b; Y_{[n]}, S_{[n]}) \\
&\leq I(W_b; Y_{[n]}, S_{[n]}|W_a, S^a_{[n]}) \\
&= \sum_{t=1}^n \Big[H(Y_t|S_{[t]}, Y_{[t-1]}, W_a, S^a_{[n]}) \\
&\qquad -H(Y_t|S_{[t]}, Y_{[t-1]}, W_a, W_b, S^a_{[n]})\Big] \\
&= \sum_{t=1}^n \Big[H(Y_t|S_{[t]}, Y_{[t-1]}, W_a, S^a_{[n]}, X^a_{[n]}) \\
&\qquad -H(Y_t|S_{[t]}, Y_{[t-1]}, W_a, W_b, S^a_{[n]}, X^a_{[n]})\Big] \\
&\leq \sum_{t=1}^n \Big[H(Y_t|S_{[t]}, X^a_t) \\
&\qquad -H(Y_t|S_{[t]}, Y_{[t-1]}, W_a, W_b, S^a_{[n]}, X^a_{[n]}, T^b_t)\Big] \\
&\overset{(i)}{=} \sum_{t=1}^n \left[H(Y_t|S_{[t]}, X^a_t) - H(Y_t|S_{[t]}, X^a_t, T^b_t)\right] \\
&= \sum_{t=1}^n I(T^b_t; Y_t|X^a_t, S_{[t]})
\end{aligned} \tag{112}$$

where (i) is valid since

$$\begin{aligned}
&P_{Y_t|S_{[t]},Y_{[t-1]},\mathbf{W},S^a_{[n]},X^a_{[n]},T^b_t}(y_t|s_{[t]}, y_{[t-1]}, \mathbf{w}, s^a_{[n]}, x^a_{[n]}, t^b_t) \\
&= \sum_{s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t, s^b_t, x^a_t, t^b_t) \\
&P_{S^b_t|S_{[t]},Y_{[t-1]},\mathbf{W},S^a_{[n]},X^a_{[n]},T^b_t}(s^b_t|s_{[t]}, y_{[t-1]}, \mathbf{w}, s^a_{[n]}, x^a_{[n]}, t^b_t) \\
&= \sum_{s^b_t\in\mathcal{S}_b} P_{Y_t|S_t,S^b_t,X^a_t,T^b_t}(y_t|s_t, s^b_t, x^a_t, t^b_t) P_{S^b_t|S_t}(s^b_t|s_t) \\
&= P_{Y_t|S_t,X^a_t,T^b_t}(y_t|s_t, x^a_t, t^b_t).
\end{aligned} \tag{113}$$

where the first equality is due to (3) and the second equality is due to (1) and (2). Following (16), we can directly verify (106). ■

We now need to show that the joint conditional distribution $P_{X_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(x^a,t^b,y,s|\mu)$ factorizes as in (46). Let first $\pi^{\mu}_{X^a,T^b}(x^a,t^b) := P_{X_t^a,T_t^b|S_{[t-1]}}(x^a,t^b|\mu)$ and observe that

$$\begin{aligned} &P_{X_t^a,T_t^b,Y_t,S_t|S_{[t-1]}}(x^a,t^b,y,s|\mu) \\ &\quad= \sum_{s_t^b\in\mathcal{S}^b} P_{Y_t|X_t^a,X_t^b,S_t}(y|x^a,t^b(s_t^b),s) \\ &\qquad P_{S_t^b|S_t}(s_t^b|s_t)P_{S_t}(s)P_{X_t^a,T_t^b|S_{[t-1]}}(x^a,t^b|\mu) \\ &\quad= \pi^{\mu}_{X^a,T^b}(x^a,t^b)P_{S_t}(s)P_{Y_t|X_t^a,T_t^b,S_t}(y|x^a,t^b,s) \quad (114) \end{aligned}$$

where the equalities are verified by (3), by (1) and by the fact that (X_t^a,T_t^b) is independent of S_t. ■

We can now complete the converse proof of Theorem 2.5. With Lemma C.1 it is shown that any achievable rate pair can be approximated by the convex combinations of rate conditions which are indexed by $\mu\in\mathcal{S}^{(n)}$ and satisfy (46) for joint state-input-output distributions. Hence, any achievable pair $(R_a,R_b)\in\overline{co}\big(\bigcup_{\hat{\pi}}\mathcal{R}_C(\hat{\pi})\big)$.

Appendix D
Proof of $\mathcal{C}_{FS}^G=\mathcal{C}_{AS}$

Let us first show that $\mathcal{C}_{FS}^G\subseteq\mathcal{C}_{AS}$. Recall that $T\in|\mathcal{T}|=|\mathcal{X}_b|^{|\mathcal{S}|}$ and $|\mathcal{U}|\leq|\mathcal{X}_a||\mathcal{X}_b||\mathcal{S}|+1$. Hence, we have either $|\mathcal{U}|>|\mathcal{T}|$ or else. In the case where $|\mathcal{U}|<|\mathcal{T}|$, we note that $|\mathcal{U}|$ is limited to a finite set without loss of generality. Hence, we can always take $|\mathcal{U}|$ at least $|\mathcal{T}|$ such that it satisfies (56), (57) and (58). Then we can directly conclude that $\mathcal{C}_{FS}^G\subseteq\mathcal{C}_{AS}$ since $P_{X^b|S,T}(x^b|s,t)=P_{X^b|S,T}(x^b|s,t,x^a)=1_{\{x^b=t(s)\}}$ and this is a special case of $P_{X^b|U,X^a,S}(x^b|u,x^a,s)$.

In order to prove the other direction, i.e., $\mathcal{C}_{AS}\subseteq\mathcal{C}_{FS}^G$, let $\mathcal{C}_{AS}^E$ be the closure of all rate pairs (R_a,R_b) satisfying

$$\begin{aligned} R_b &< I(U;Y|X^a) \quad (115)\\ R_b+R_a &< I(U,X^a;Y) \quad (116) \end{aligned}$$

for some joint measure on $\mathcal{S}\times\mathcal{X}_a\times\mathcal{X}_b\times\mathcal{Y}\times\mathcal{U}$ having the form

$$P_{Y|X^a,X^b,S}(y|x^a,x^b,s)1_{\{x^b=\mathbf{m}(s,x^a,u)\}}P_S(s)P_{X^a,U}(x^a,u), \quad (117)$$

for some $\mathbf{m}:\mathcal{U}\times\mathcal{X}_a\times\mathcal{S}\to\mathcal{X}_b$, where $|\mathcal{U}|\leq|\mathcal{S}||\mathcal{X}_a||\mathcal{X}_b|+1$, and we first show that $\mathcal{C}_{AS}=\mathcal{C}_{AS}^E$, and following this, we show that $\mathcal{C}_{AS}^E\subseteq\mathcal{C}_{FS}^G$.

Lemma D.1: $\mathcal{C}_{AS}=\mathcal{C}_{AS}^E$.

Proof: It is obvious that $\mathcal{C}_{AS}^E\subseteq\mathcal{C}_{AS}$ and hence, we need to show that $\mathcal{C}_{AS}\subseteq\mathcal{C}_{AS}^E$. Let $\bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s)$ be a joint distribution in the form of (58), i.e.,

$$\begin{aligned} \bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s) &= \bar{P}_{X^b|X^a,U,S}(x^b|x^a,u,s) \\ &\quad\times P_S(s)\bar{P}_{X^a,U}(x^a,u). \quad (118) \end{aligned}$$

Let $\bar{\mathbf{\Lambda}}$ denote a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ matrix where $\bar{\mathbf{\Lambda}}_{i,jkl}=\bar{P}_{X^b|X^a,U,S}(i|j,k,l)$, $1\leq i\leq|\mathcal{X}_b|$, $1\leq j\leq|\mathcal{X}_a|$, $1\leq k\leq|\mathcal{U}|$ and $1\leq l\leq|\mathcal{S}|$. Hence, $\bar{\mathbf{\Lambda}}$ is a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ row stochastic matrix, i.e., $\bar{\mathbf{\Lambda}}_{i,jkl}\geq 0$, $\forall i,j,k,l$ and $\sum_{i=1}^{|\mathcal{X}_b|}\bar{\mathbf{\Lambda}}_{i,jkl}=1$, $\forall j,k,l$. Let $\mathbf{\Lambda}$ denote a $|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|$-by-$|\mathcal{X}_b|$ binary stochastic matrix, that is a matrix with each row has exactly one non-zero element, which is 1. Observe now that any row stochastic matrix can be written as a convex combination of binary stochastic matrices (e.g., see [39, Lemma 5] and [40, Proposition IV.1]). Therefore, we have

$$\bar{\mathbf{\Lambda}}=\sum_{i=1}^{k}\lambda_i\mathbf{\Lambda}^{(i)},\ \ \sum_{i=1}^{k}\lambda_i=1, \quad (119)$$

where $\mathbf{\Lambda}^{(i)}$ is a binary stochastic matrix and by [39, Lemma 5], $k\leq(|\mathcal{X}_a||\mathcal{U}||\mathcal{S}|)^2$.

Let, for the joint distribution $\bar{P}_{X^b,X^a,U,S}(x^b,x^a,u,s)$,

$$\begin{aligned} \bar{R}_b &< I(U;Y|X^a)_{\bar{\mathbf{\Lambda}}}, \quad (120)\\ \bar{R}_a+\bar{R}_b &< I(U,X^a;Y)_{\bar{\mathbf{\Lambda}}}. \quad (121) \end{aligned}$$

Hence, $(\bar{R}_a,\bar{R}_b)\in\mathcal{C}_{AS}$. Now, observe that for a fixed distribution $P_{X^a,U}(x^a,u)$, both $I(U,X^a;Y)$ and $I(U;Y|X^a)$ are convex in $P_{Y|X^a,U}(y|x^a,u)$ and hence, convex in $P_{X^b|X^a,U,S}(\cdot|x^a,u,s)$. This and (119) imply that

$$\begin{aligned} I(U;Y|X^a)_{\bar{\mathbf{\Lambda}}} &\leq \sum_{i=1}^{k}\lambda_i I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}}, \quad (122)\\ I(U,X^a;Y)_{\bar{\mathbf{\Lambda}}} &\leq \sum_{i=1}^{k}\lambda_i I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}}, \quad (123) \end{aligned}$$

where $I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}}$ and $I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}}$ denote the mutual information terms induced by $\mathbf{\Lambda}^{(i)}$.

Now, let (R_a^i,R_b^i), $1\leq i\leq k$, be such that

$$\begin{aligned} R_b^i &\leq I(U;Y|X^a)_{\mathbf{\Lambda}^{(i)}},\\ R_b^i+R_a^i &\leq I(U,X^a;Y)_{\mathbf{\Lambda}^{(i)}}, \end{aligned}$$

and hence, $(R_a^i,R_b^i)\in\mathcal{C}_{AS}^E$, $1\leq i\leq k$. Let $(R_a^f,R_b^f)=\sum_{i=1}^{k}\lambda_i(R_a^i,R_b^i)$. Since a convex combination of achievable rates is also achievable, so $(R_a^f,R_b^f)\in\mathcal{C}_{AS}^E$. This observation and inequalities (120)-(123) complete the claim that $(\bar{R}_a,\bar{R}_b)\in\mathcal{C}_{AS}^E$. ■

Up to now, we have shown that $\mathcal{C}_{FS}^G\subseteq\mathcal{C}_{AS}$ and $\mathcal{C}_{AS}^E=\mathcal{C}_{AS}$. In order to prove that $\mathcal{C}_{FS}^G=\mathcal{C}_{AS}$, it remains to show that $\mathcal{C}_{AS}^E\subseteq\mathcal{C}_{FS}^G$. Note that $\mathcal{C}_{AS}^E$ still depends on $P_{X^a,U}(x^a,u)$ in which $|\mathcal{U}|$ can be larger than $|\mathcal{T}|$. Hence, in the next lemma we basically show that for every $P_{X^a,U}(x^a,u)$, there exists a $\hat{\pi}_{T^a,U}(t^a,u)$ which induces the same rate constraints as induced by $P_{X^a,U}(x^a,u)$.

Lemma D.2: $\mathcal{C}_{AS}^E\subseteq\mathcal{C}_{FS}^G$.

Proof: Fix a distribution $P^*_{Y,X^a,X^b,U,S}(y,x^a,x^b,u,s)$ satisfying (117), i.e.,

$$\begin{aligned} P^*_{Y,X^a,X^b,U,S}(y,x^a,x^b,u,s) &= P^*_{Y|X^a,X^b,S}(y|x^a,x^b,s) \\ &\quad 1_{\{x^b=\mathbf{m}(s,x^a,u)\}}P_S(s)P^*_{X^a,U}(x^a,u). \quad (124) \end{aligned}$$

Observe that for every $\mathbf{m}$ satisfying $x^b=\mathbf{m}(u,x^a,s)$, one can define

$$x^b=\mathbf{m}(u,x^a,s)=\bar{\mathbf{m}}(x^a,u)(s),\ \ \bar{\mathbf{m}}(x^a,u)\in\mathcal{T}, \quad (125)$$

where $\mathcal{T}$ is the set of all mappings from $\mathcal{S}$ to $\mathcal{X}_b$. Now, let

$$\Big(I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)},I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)}\Big)$$

be the mutual information pair induced by $P^*_{Y,X^a,U}(y,x^a,u)$. We have

$$\begin{aligned}
&I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)}\\
&\quad= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\\
&\qquad\qquad \log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_Y(y)P^*_{U,X^a}(u,x^a)}\\
&\quad= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_Y(y)P^*_{U,X^a}(u,x^a)}\\
&\quad\overset{(i)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y,U,X^a,T}(y,u,x^a,t)}{P^*_Y(y)P^*_{U,X^a,T}(u,x^a,t)}\\
&\quad\overset{(ii)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y|X^a,T}(y|x^a,t)P^*_{U,T,X^a}(u,t,x^a)}{P^*_Y(y)P^*_{U,T,X^a}(u,t,x^a)}\\
&\quad= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y,X^a,T}(y,x^a,t)}{P^*_Y(y)P^*_{X^a,T}(x^a,t)}\\
&\quad= \sum_{t\in\mathcal{T}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,T}(y,x^a,t)\\
&\qquad\qquad \log\frac{P^*_{Y,X^a,T}(y,x^a,t)}{P^*_Y(y)P^*_{X^a,t}(x^a,t)}\\
&\quad= I(T,X^a;Y)_{P^*_{Y,X^a,T}(y,x^a,t)}, \qquad (126)
\end{aligned}$$

where (i) is valid since $\bar{\mathbf{m}}(x^a,u)\in\mathcal{T}$, i.e., for each (x^a,u) there exists only one $t\in\mathcal{T}$ such that $P_{T|X^a,U}(t|x^a,u)=1$, (ii) is valid since

$$\begin{aligned}
&P^*_{Y|X^a,T,U}(y|x^a,t,u)\\
&\quad\overset{(iii)}{=} \sum_{s\in\mathcal{S}} P^*_{Y|X^a,T,U,S}(y|x^a,t,u,s)P_S(s)\\
&\quad\overset{(iv)}{=} \sum_{s\in\mathcal{S}} P_{Y|X^a,T,S}(y|x^a,t,s)P_S(s)\\
&\quad= \sum_{s\in\mathcal{S}} P^*_{Y,S|X^a,T}(y,s|x^a,t) = P^*_{Y|X^a,T}(y|x^a,t) \qquad (127)
\end{aligned}$$

where (iii) is valid since S and (X^a,T,U) are independent and (iv) is valid due to (3). Similarly, we have

$$\begin{aligned}
&I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)}\\
&\quad= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\\
&\qquad\qquad \log\frac{P^*_{Y,U|X^a}(y,u|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U|X^a}(u|x^a)}\\
&\quad= \sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U}(y,x^a,u)\\
&\qquad\qquad \log\frac{P^*_{Y,U,X^a}(y,u,x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U,X^a}(u,x^a)}\\
&\quad\overset{(v)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y,U,X^a,T}(y,u,x^a,t)}{P^*_{Y|X^a}(y|x^a)P^*_{U,X^a,T}(u,x^a,t)}\\
&\quad\overset{(vi)}{=} \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y|T,X^a}(y|t,x^a)P^*_{U,T,X^a}(u,t,x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{U,T,X^a}(u,t,x^a)}\\
&\quad= \sum_{t\in\mathcal{T}}\sum_{u\in\mathcal{U}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,U,T}(y,x^a,u,t)\\
&\qquad\qquad \log\frac{P^*_{Y,T|X^a}(y,t|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{T|X^a}(t|x^a)}\\
&\quad= \sum_{t\in\mathcal{T}}\sum_{y\in\mathcal{Y}}\sum_{x^a\in\mathcal{X}_a} P^*_{Y,X^a,T}(y,x^a,t)\\
&\qquad\qquad \log\frac{P^*_{Y,T|X^a}(y,t|x^a)}{P^*_{Y|X^a}(y|x^a)P^*_{T|X^a}(t|x^a)}\\
&\quad= I(T;Y|X^a)_{P^*_{Y,X^a,T}(y,x^a,t)}, \qquad (128)
\end{aligned}$$

where (v) and (vi) follow from the same reasonings of (i) and (ii), respectively. Let $R_b^{'} < I(U;Y|X^a)_{P^*_{Y,X^a,U}(y,x^a,u)}$ and $R_b^{'}+R_a^{'} < I(U,X^a;Y)_{P^*_{Y,X^a,U}(y,x^a,u)}$ which imply $(R_a^{'},R_b^{'})\in\mathcal{C}^E_{AS}$. Observe now that for a distribution in the form of $P^*_{Y,X^a,T}(y,x^a,t)$, one can define $\hat{\pi}_{X^a,T}(x^a,t)=P^*_{X^a,T}(x^a,t)$. Therefore, since $\mathcal{C}^G_{FS}=\overline{co}\Big(\bigcup_{\hat{\pi}}\mathcal{R}_C^{'}(\hat{\pi})\Big)$, and due to (126) and (128), $(R_a^{'},R_b^{'})\in\mathcal{C}^G_{FS}$, which completes the claim.

■

References

[1] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop,* vol. 2, pp. 289-293, 1958.

[2] S. I. Gelfand and M. S. Pinsker, "Coding for channels with random parameters," *Prob. of Cont. and Inform.,* vol. 9, pp. 19-31, 1980.

[3] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *IEE-Proceedings-I* vol. 139, pp. 113-117, Apr. 1992.

[4] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory,* vol. 45, no. 6, pp. 2007-2019, Sep. 1999.

[5] U. Erez and R. Zamir, "Noise prediction for channels with side information at the transmitter," *IEEE Trans. Inform. Theory,* vol. 46, no. 4, pp. 1610-1617, July 2000.

[6] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory,* vol. 43, no. 6, pp. 1986-1992, Nov. 1997.

[7] S. Yüksel and S. Tatikonda, "Capacity of Markov channels with partial state feedback," *Proc. IEEE Int. Symp. Information Theory* Nice, France, pp. 1861-1865, June 2007.

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

[9] H. Permuter, T. Weissman and A.J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inform. Theory,* vol. 55, no.2, pp. 644-662 , Feb. 2009.

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

[11] S. A. Jafar, "Capacity with causal and non-causal side information-a unified view," *IEEE Trans. Inform. Theory,* vol. 52 no. 12, pp. 5468-5474, Dec. 2006.

[12] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inform. Theory,* vol. 51, no. 11, pp. 3992-4003, Nov. 2005.

[13] G. Como and S. Yüksel, "On the capacity of memoryless finite state multiple access channels with asymmetric state information at the encoders," *IEEE Trans. Inform. Theory,* vol. 57, no. 3, pp. 1267-1273, March 2011.

[14] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: double state," *IEEE Trans. Inform. Theory,* vol. 59, no. 3, pp. 1379-1393, Mar. 2013.

[15] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal state information: common state," *IEEE Trans. Inform. Theory,* vol. 59, no. 1, pp. 32-50, Jan. 2013.

[16] M. Li, O. Simeone and A. Yener, "Multiple access channels with states causally known at transmitters," available at [arxiv:1011.6639v2].

[17] A. Lapidoth and Y. Steinberg, "A note on multiple-access channels with strictly-causal state information," available at [arXiv:1106.0380v1].

[18] U. Basher, A. Shirazi and H. Permuter, "Capacity region of finite state multiple-access channel with delayed state information at the transmitters," *IEEE Trans. Inform. Theory,* vol. 58, no. 6, pp. 3430-3452, June 2012.

[19] S. Sigurjónsson and Y. H. Kim, "On multiple user channels with causal state information at the iransmitters," *Proc. IEEE Int. Symp. Information Theory* Adelaide, Australia, July 2005.

[20] A. Somekh-Baruch, S. Shamai and S. Verdú; "Cooperative multiple-access encoding with states available at one transmitter," *IEEE Trans. Inform. Theory,* vol. 54, no. 10, pp. 4448-4469, Oct. 2008.

[21] S. Kotagiri and J. Laneman, "Multiaccess channels with state known to one encoder: A case of degraded message sets," *Proc. IEEE Int. Symp. Information Theory* Nice, France, pp. 1566-1570, June 2007.

[22] A. Zaidi, P. Piantanida and S. Shamai, "Multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder," *Proc. IEEE Int. Symp. Information Theory* St. Petersburg, Russia, July 2011.

[23] A. Zaidi, P. Piantanida and S. Shamai (Shitz), "Capacity region of multiple access channel with states known noncausally at one encoder and only strictly causally at the other encoder," available at [arXiv:1201.3278].

[24] H. Permuter, S. Shamai and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inform. Theory,* vol. 57, no. 10, pp. 6379-6396, Oct. 2011.

[25] M. Zamanighomi, M. J. Emadi, F. S. Chaharsooghi, and M. R. Aref, "Achievable rate region for multiple access channel with correlated channel states and cooperating encoders," available at [arXiv:1102.1165].

[26] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information: Subject review," *Found. Trends Commun. Inf. Theory,* vol. 4, no. 6, pp. 445-586 , 2007.

[27] A. Zaidi, S. Kotagari, J. N. Laneman and L. Vandendorpe, "Cooperative relaying with state available non-causally at the relay," *IEEE Trans. Inform. Theory,* vol. 56, no. 5, pp. 2272-2298, May 2010.

[28] A. Somekh-Baruch, S. Shamai and S. Verdú; "Cognitive interference channels with state information," *Proc. IEEE Int. Symp. Information Theory* Toronto, Canada, pp. 1353-1357, July 2008.

[29] A. Zaidi, S. Kotagari, J. N. Laneman and L. Vandendorpe, "Multiaccess channels with state known to one encoder: Another case of degraded message sets," *Proc. IEEE Int. Symp. Information Theory* Seoul, Korea, pp. 2376-2380, July 2009.

[30] M. Li, O. Simeone, and A. Yener, "Message and state cooperation in a relay channel when only the relay knows the state," available at [arXiv:1102.0768].

[31] A. Zaidi, S. Shamai (Shitz), P. Piantanida and L. Vandendorpe, "Bounds on the capacity of the relay channel with noncausal state at source," available at [arXiv:1104.1057].

[32] H. S. Witsenhausen, "Equivalent stochastic control problems," *Mathematics of Control, Signal and Systems,* vol. 1, pp. 3-11, Springer-Verlag, 1988.

[33] S. Yüksel, "On optimal causal coding of partially observed Markov sources in single and multi-terminal settings," *IEEE Trans. Inform. Theory,* vol. 59, no. 1, pp. 424-437, Jan. 2013.

[34] A. Mahajan and D. Teneketzis, "Optimal performance of networked control systems with non-classical information structures," *SIAM Jour. Cont. and Opt.* vol. 48, no. 3, pp. 1377-1404, May 2009.

[35] A. Nayyar and D. Teneketzis, "On the structure of real-time encoders and decoders in a multi-terminal communication system," *IEEE Trans. Inform. Theory,* vol. 57, no. 9, pp. 6196-6214, Sep. 2011.

[36] T. M. Cover and J. Thomas, *Elements of Information Theory,* New York: Wiley, 2nd edition, 2006.

[37] N. Şen, *Channel Capacity in the Presence of Feedback and Side Information,* Ph.D. Dissertation, Dept. of Mathematics and Statistics, Queen's University, July 2013.

[38] S. I. Gelfand and M. S. Pinsker, "Coding of sources on the basis of observations with incomplete observations," *Prob. Pered. and Inform.,* vol. 15, no. 2, pp. 45-57, July 1979.

[39] G. Högnäs, "Random semigroup acts on a finite set," *J. Austral. Math. Soc.,* vol. 23 (Series A), pp. 481-498, 1977.

[40] U. Niesen, C. Fragouli, D. Tuninetti, "On capacity of line networks," *IEEE Trans. Inform. Theory,* vol. 53, no. 11, pp. 4039-4058, Nov. 2007.

Nevroz Şen received the B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Turkey, 2003 and the M.Eng. and Ph.D. degrees in mathematics and engineering from Queen's University, Canada, in 2009 and 2013, respectively. He is currently a postdoctoral researcher in the Department of Electrical and Computer Engineering at the McGill University.

In 2004-2008, he was with the research and development department of Nortel NETAŞ, Turkey, where he worked as a software design engineer. His research interests include information theory and stochastic control.

PLACE PHOTO HERE

Fady Alajaji (S'90–M'94–SM'00) received the B.E. degree with distinction from the American University of Beirut, Lebanon, and the M.Sc. and Ph.D. degrees from the University of Maryland, College Park, all in electrical engineering, in 1988, 1990 and 1994, respectively. He held a postdoctoral appointment in 1994 at the Institute for Systems Research, University of Maryland.

In 1995, he joined the Department of Mathematics and Statistics at Queen's University, Kingston, Ontario, where he is currently a Professor of Mathematics and Engineering. Since 1997, he has also been cross-appointed in the Department of Electrical and Computer Engineering at the same university. From 2003 to 2008, he served as chair of the Queen's Mathematics and Engineering program. His research interests include information theory, digital communications, error control coding, joint source-channel coding and data compression.

Dr. Alajaji currently serves as Area Editor and Editor for Source and Source-Channel Coding for the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as organizer and Technical Program Committee member of several international conferences and workshops. He received the Premier's Research Excellence Award from the Province of Ontario.

Serdar Yüksel received his BSc degree in Electrical and Electronics Engineering from Bilkent University in 2001; MS and PhD degrees in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2003 and 2006, respectively. He was a post-doctoral researcher at Yale University for a year before joining Queen's University as an Assistant Professor of Mathematics and Engineering in the Department of Mathematics and Statistics, where he is now an Associate Professor. He has been awarded the 2013 CAIMS/PIMS Early Career Award in Applied Mathematics. His research interests are on stochastic and decentralized control, information theory and applied probability.