

The Finite State MAC With Cooperative Encoders and Delayed CSI

Ziv Goldfeld, *Student Member, IEEE*, Haim H. Permuter, *Senior Member, IEEE*,
and Benjamin M. Zaidel, *Member, IEEE*

Abstract—In this paper, we consider the finite-state multiple access channel (MAC) with partially cooperative encoders and delayed channel state information (CSI). Here, partial cooperation refers to the communication between the encoders via finite-capacity links. The channel states are assumed to be governed by a Markov process. Full CSI is assumed at the receiver, while at the transmitters, only delayed CSI is available. The capacity region of this channel model is derived by first solving the case of the finite-state MAC with a common message. Achievability for the latter case is established using the notion of strategies, however, we show that optimal codes can be constructed directly over the input alphabet. This results in a single codebook construction that is then leveraged to apply simultaneous joint decoding. Simultaneous decoding is crucial here because it circumvents the need to rely on the capacity region's corner points, a task that becomes increasingly cumbersome with the growth in the number of messages to be sent. The common message result is then used to derive the capacity region for the case with partially cooperating encoders. Next, we apply this general result to the special case of the Gaussian vector MAC with diagonal channel transfer matrices, which is suitable for modeling, e.g., orthogonal frequency division multiplexing-based communication systems. The capacity region of the Gaussian channel is presented in terms of a convex optimization problem that can be solved efficiently using numerical tools. The region is derived by first presenting an outer bound on the general capacity region and then suggesting a specific input distribution that achieves this bound. Finally, numerical results are provided that give valuable insight into the practical implications of optimally using conferencing to maximize the transmission rates.

Index Terms—Capacity region, common message, convex optimization, cooperative encoders, delayed CSI, diagonal vector gaussian multiple-access channel, finite-state channel, multiple-access channel, simultaneous decoding, strategy letters.

I. INTRODUCTION

TEMPORAL variations, a characteristic typical of wireless channels, may occur due to atmospheric changes, changes in the environment, the mobility of transmitters and/or

receivers or time-varying intentional or unintentional interference. Since accurate channel state information (CSI) at both the transmitting and the receiving ends is crucial for efficient communications, measures are commonly incorporated in the communication protocol to enable channel state estimation. For example, the long term evolution (LTE) cellular communication standard relies on pilot signals transmitted at pre-scheduled time intervals and frequency slots to estimate the channel's state [1]. Performed at the receiver, these estimations are then typically fed back to the transmitter, but obtaining perfect CSI at both ends of the channel in practical systems is a formidable challenge. More often than not, CSI is subject to channel estimation errors and feedback is not instantaneous due to some inevitable processing delay, and as a result, receivers and transmitters typically have access to only partial CSI. The impact of such partial CSI on the achievable performance, therefore, has attracted much attention in recent years. In the case of multiuser communication, performance is affected not only by channel characteristics, but also by interactions between the users. In particular, different forms of cooperation between the transmitting and receiving ends, a subject of growing interest in recent years (see [2], [3]), may significantly enhance performance. This paper aims to investigate the combined impact of both partial CSI and cooperation. More specifically, we focus on a two-user finite state Markov multiple access channel (FSM-MAC), with *partially* cooperative encoders and *delayed* CSI, as illustrated in Fig. 1 and explained in the following text.

In the communication scenario under discussion, each of the two encoders wishes to send an independent private message through a time-varying MAC to the decoder. Delayed CSI is assumed to be available at the encoders, while full delayless CSI is assumed at the decoder. Different users may be subject to different CSI delays. It is further assumed that prior to each transmission block, the two encoders are allowed to hold a conference. More specifically, it is assumed that the encoders can communicate with each other over noise-free communication links of given capacities. We restrict the discussion to the case in which the conference held between the encoders is independent of the CSI.

The non-state-dependent MAC with partially cooperative encoders was first introduced by Willems [4], who also derived the capacity region for the discrete memoryless setting. Special cases of this channel model include that in which the encoders are ignorant of each other's messages (i.e., the capacities of the communication links between them are both zero) and that in which the encoders fully cooperate (i.e., the capacities of

Manuscript received March 27, 2013; revised December 15, 2013; accepted July 18, 2014. Date of publication August 8, 2014; date of current version September 11, 2014. This work was supported in part by the European Research Council, Israel Science Foundation, under Grant 684/11, and in part by the Israel Ministry of Defence. This paper was presented at the 2012 IEEE International Symposium on Information Theory and IEEE 27th Convention of Electrical and Electronics Engineers in 2012.

Z. Goldfeld and H. H. Permuter are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84 110, Israel (e-mail: zgzg1984@gmail.com; haimp@bgu.ac.il).

B. M. Zaidel is with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel (e-mail: benjamin.zaidel@gmail.com).

Communicated by Y. Liang, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2014.2346494

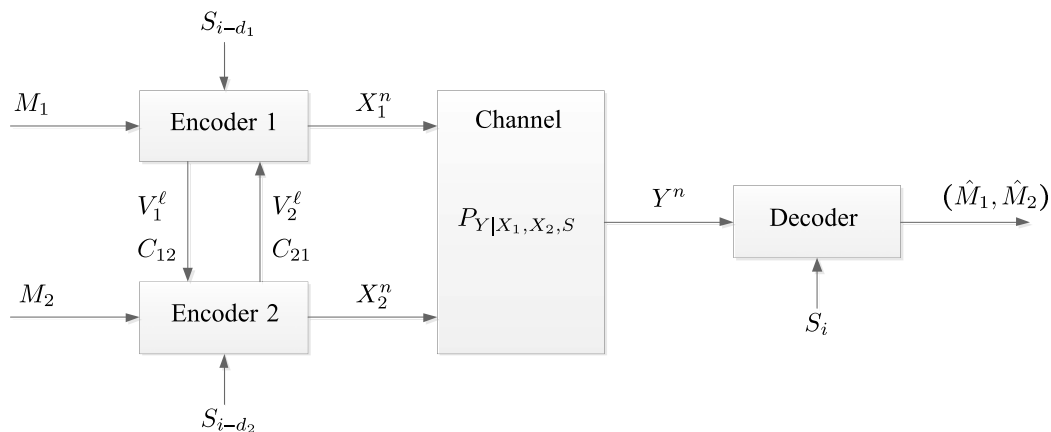


Fig. 1. FSM-MAC with partially cooperative encoders, CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2 .

the communication links are infinite). The first setting, where no conference is held, corresponds to the classical MAC, for which the capacity region was determined by Ahlswede [5] and Liao [6]. In contrast, in the second setting, where total cooperation is available, the encoders can act as one by fully sharing their private messages via the conference. The capacity region for this case is the part of the first quadrant below the so-called total cooperation line. This triangle-shaped region always contains the capacity region for the classical MAC.

In his proof of achievability for the conferencing MAC, Willems [4] introduced a coding scheme based on the capacity region for the MAC with a common message, derived by Slepian and Wolf in [7]. Willems showed that in order to achieve the capacity region, the encoders should use the cooperation link to share parts of their private messages and then use a coding scheme for the ordinary MAC with a common message. Although Willems's model allows interactive communication between the encoders, it was shown both in [4] and later in [8] that a single round of communication between the encoders (referred to as a "pair of simultaneous monologues" in [4]) suffices to achieve optimality.

Additional multiuser settings that involve cooperation between users through communication links of finite capacities have been extensively treated in the literature. See [9] and [10] for studies of the MAC, [2] and [11]–[16] for studies of the interference channel with cooperating nodes, [17] for the broadcast channel, [18] and [19] for cooperative relaying and [3] and [20] and references therein for cooperation in cellular architectures. A comprehensive survey of cooperation and its role in communication can be found in [21]. It is important to note, however, that in all of the above settings the channel was not assumed to be time-varying.

Multiuser settings that combine both time-varying channels and user cooperation are obviously of major interest as well. A Gaussian fading MAC with cooperating encoders that have access to delayless CSI was considered in [22] and [23]. As in our case, these works assume that cooperation is allowed only before the CSI becomes available at the encoders. The case in which the CSI becomes available to the encoders prior to transmission is treated in [24], where a MAC with perfect

noncausal CSI is considered. The coding scheme introduced in [24] uses conferencing to share parts of the messages as well as CSI.

The notion of modeling time-varying channels as state-dependent channels dates back to Shannon [25], who characterized the capacity of the state-dependent, memoryless point-to-point channel with independent and identically distributed (i.i.d.) states available causally at the encoder. To establish achievability, Shannon presented a code construction that relied on "strategies" (or "strategy letters") [26], a notion we also exploit in this paper. Gelfand and Pinsker [27], and later Heegard and El Gamal [28], studied the case in which the encoder observes the channel states noncausally. In both [27] and [28] a single letter expression for the capacity is derived using random binning. In [29], Goldsmith and Varaiya considered a fading channel with perfect CSI at both the transmitter and the receiver. It was shown that in such a case, the optimal strategy is to employ waterfilling over time.

As was already stated, because perfect CSI is difficult to obtain in practical systems, models that involve partial or imperfect CSI have attracted a lot of attention in recent years. At first, different settings involving an i.i.d. state sequence with imperfect CSI were treated. Initially, various point-to-point channel scenarios with partial CSI were studied. Among others, the causal, noncausal, rate-limited and noisy cases were addressed [30]–[32]. Extension of the result to the MAC with rate-limited CSI can be found in [33]. In [34], the authors derive the capacity region for the MAC with asymmetric quantized CSI at the encoders, where the quantization models the imperfection in the channel state estimation (full CSI at the decoder is assumed). Later, in [35] Lapidoth and Steinberg provided an inner bound for the capacity region of the MAC with strictly causal CSI at the encoders. In contrast to the point-to-point setting, where strictly causal CSI regarding an i.i.d. state sequence does not increase capacity, the capacity region of the MAC with causal CSI is strictly larger than the corresponding region without CSI. Li *et al.* presented an improved inner bound for the same setting in [36]. A comprehensive monograph on channel coding in the presence of side information can be found in [37], where an i.i.d.

state sequence is assumed. An information theoretic model for a single user channel involving delayed CSI and a state process that is no longer restricted to be memoryless and i.i.d. was first introduced by Viaswanathan [38], who derived the capacity while assuming a FSM channel. This result was later generalized by Caire and Shamai in [26], where they addressed a point-to-point channel in which the CSIs at both encoder and decoder admit some general joint probability law. A general capacity formula, which relies on the notion of *inf-information rate* [39], is then provided for the case of state processes with memory. The result is then shown to boil down to a single-letter characterization in the case in which perfect CSI is available to the receiver, the CSI at the transmitter is given by a deterministic function of the channel state, and the two processes are jointly stationary and ergodic. By an appropriate choice of the above deterministic function, the result for Viswanathan's delayed CSI model [38] is obtained as a special case of the result in [26]. A generalization of the point-to-point results of [26] to the MAC was presented by Das and Narayan in [40]. The generality of the channel model therein leads to multiletter characterization of the capacity region in various settings, which unfortunately provides limited insight into practical encoding schemes for channel models in this framework.

Taking a practically oriented approach, we focus in this paper on a specific channel definition that leads to single-letter results. Following [38], we model temporal variations by means of a FSM channel [41], [42]. The channel state is determined on a per symbol basis and governed by the underlying FSM process. An important extension of this idea to the multiuser case was introduced by Basher *et al.* in [43], presenting the FSM-MAC with delayed CSI and non-cooperating encoders, i.e., where no conferencing is held (see also [44] for a related source coding analysis). In the proof of the capacity region for this model, achievability was established by employing a coding scheme based on rate-splitting and multiplexing-coding combined with successive decoding at the receiver. Successive decoding was used in [43] to demonstrate that the two corner points of the capacity region are achievable. The whole capacity region is then achievable via time-sharing. Although the setting in [43] constitutes a special case of the general model in [40], the main contribution of [43] is the single-letter characterization of the capacity region and the detailed construction of the coding scheme.

In the current paper, accounting for the availability of a conferencing link between the encoders, we take a different approach than that taken in [43]. We base the proof of achievability on the coding scheme for the MAC with a common message as presented in [4], and therefore, we start by deriving the capacity region for the FSM-MAC with a common message and the same CSI properties as in [43]. We thus provide a solution to what has been, until now, an unsolved problem. Next, using the achievable scheme for the common message setting, the achievability of the conferencing region is established. We note that the large number of corner points induced by the presence of an additional transmission rate (namely, the rate of the common message) render the provision of an achievable coding scheme for the common

message setting based on achieving the region's corner points an awkward task. Moreover, the use of rate-splitting and multiplexing-coding when a common message is involved yields a rather complex coding scheme which we sought to avoid.

Therefore, we present an alternative coding scheme that employs strategy letters in the code construction (see [25], [26], [40]) and simultaneous decoding. However, unlike the case of Shannon's classical result for the point-to-point channel with causal encoder CSI, here we show that optimal codes can be constructed directly over the input alphabet (as also shown for certain special cases in [40]). Namely, a single codebook is generated for each of the three messages over a super-alphabet that corresponds to the different realizations of the delayed CSI available at the encoders. At each time instance, a symbol that is correlated with the *current* available delayed CSI is selected by the encoders and transmitted to the channel. Thus, in contrast to previous works involving delayed CSI (see [38], [43]), here rate-splitting is no longer required. The decoder then uses its access to full CSI (which deterministically defines the delayed state sequences as well) to reduce each codeword (originally constructed over a super-alphabet) to a sequence over the input alphabet and executes a simultaneous decoding scheme based on joint typicality. Indeed, one of the most significant contributions of our paper is this coding scheme for the MAC with a common message and delayed CSI. Not only does it successfully avoid the unnecessary complexity of its rate-splitting and multiplexing counterpart and relies on a simpler codebook construction, it also achieves every possible point in the region rather than only the corner points. Furthermore, this two-user coding scheme is easily extendable to the case of multiple users with a *single* common message.

Based on the general results for the FSM-MAC with conferencing, we continue with the derivation of the capacity region for the special case of a vector Gaussian FSM-MAC with diagonal channel transfer matrices. This channel model can be used to represent an orthogonal frequency-division multiplexing (OFDM)-based communication system, employing single receive and transmit antennas, where the diagonal entries of the channel matrices represent the orthogonal sub-channels used by the OFDM scheme.

To derive the capacity region for the latter channel, we use a multivariate extension of a novel tool first derived in [45] (namely, a necessary and sufficient condition for a Gaussian triplet of random variables to satisfy a certain Markov relation), and demonstrate that Gaussian multivariate distributions maximize certain mutual information expressions under a Markovity constraint. The scalar version of this tool was employed by Lapidot *et al.* [46] to provide an outer bound for the capacity region of the scalar Gaussian non-state-dependent MAC with conferencing encoders. Wigger and Kramer also used this tool in their solution for the capacity region of the three-user, non-state-dependent MIMO MAC with conferencing [47]. The need to use the tool from [45] stems from the fact that the input distribution of the conferencing channel must admit a certain Markovity constraint. For cases in which no Markov relation needs to be satisfied, the traditional approach

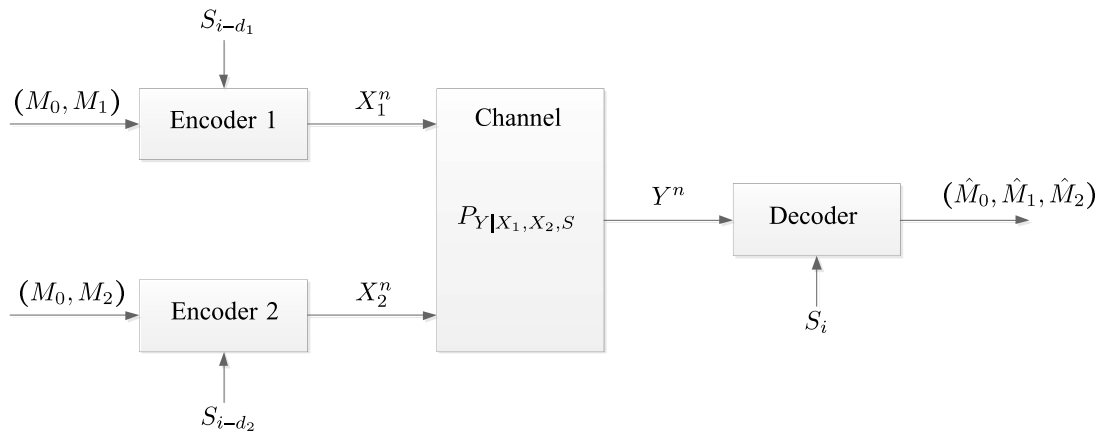


Fig. 2. FSM-MAC with a common message, full CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2 .

to proving the optimality of Gaussian multivariate distributions involves employing either the Vector Max-Entropy Theorem (a direct extension of [48, Th. 12.1.1]) or a conditional version of it. Here, however, this approach fails since replacing a non-Gaussian vector satisfying the Markovity condition by a Gaussian vector of the same covariance matrix may result in a Gaussian vector that violates the Markovity condition. To overcome this issue we use a sufficient and necessary condition on the (auto- and cross-) covariance matrices of the involved Gaussian random vectors for them to admit a Markov relation [49, Section 2, Theorem 1].

We note that although Gaussian input vectors are shown to be optimal in this setting, the original form of the capacity region involves a non-convex optimization problem. To circumvent this difficulty, new variables are introduced to convert the optimization problem into a convex problem that can then be solved using numerical tools such as CVX [50]. The capacity region for the corresponding scalar Gaussian channel can be immediately derived from the result for the vector channel setting and serves as an extension of the result in [46] to the state-dependent case. The capacity region of the vector Gaussian FSM-MAC with a common message and the same CSI properties can also be easily derived from the result for the conferencing channel by exploiting the strong correspondence between the two models and using a simple analogy.

To gain some insight into the practical implications of the results we conclude this paper with a specific example, namely, a scalar AWGN channel with two possible states ('Good' and 'Bad'). Numerical results are included to demonstrate the impact of different channel parameters on the capacity region and the optimal input distribution. Our interpretation of interactions between the different parameters produces valuable insights.

The remainder of the paper is organized as follows. In Section II we describe the two communication models of interest – the FSM-MAC with a common message and delayed CSI and the FSM-MAC with partially cooperative encoders and delayed CSI. In Sections III and IV, we state the capacity results for the common message and conferencing

models, respectively. Each result is followed by its proof. Section V follows with the definition of the vector Gaussian FSM-MAC with diagonal channel transfer matrices and the derivation of the maximization problem defining its capacity region. The regions for the corresponding common message model and the scalar setting are given as special cases. The two-state Gaussian example is discussed in this section as well. Finally, Section VI summarizes the main achievements and insights presented in this paper along with some possible future research directions and extensions.

II. CHANNEL MODELS AND NOTATION

In this paper, we investigate the capacity region of the FSM-MAC with partially cooperative encoders, full CSI at the decoder (receiver) and delayed CSI at the encoders (transmitters), as illustrated in Fig. 1. To this end, we first consider a different setting, which is the FSM-MAC with a common message and the same CSI properties, as depicted in Fig. 2. The derivation of the capacity region for the latter common message setting forms the basis for the achievability proof for the former setting where a conferencing link exists between the encoders. Since most definitions for both channels follow similar lines, we start by defining the common message setting and then extend the description for the setting of partially cooperative encoders.

We use the following notations. Matrices are denoted by nonitalicized capital letters, e.g., \mathbf{X} . Calligraphic letters denote sets, e.g., \mathcal{X} , while the cardinality of a set \mathcal{X} is denoted by $|\mathcal{X}|$. \mathcal{X}^n stands for the n -fold Cartesian product of \mathcal{X} . An element of \mathcal{X}^n is denoted by $x^n = (x_1, x_2, \dots, x_n)$, and its substrings as $x_i^j = (x_i, x_{i+1}, \dots, x_j)$; when $i = 1$, the subscript is omitted. We use the notation $x^{n \setminus i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. Whenever the dimension n is clear from the context, vectors (or sequences) are denoted by boldface letters, e.g., \mathbf{x} . Random variables are denoted by uppercase letter, e.g., X , with similar conventions for random vectors. X_i^j stands for the sequence of random variables $(X_i, X_{i+1}, \dots, X_j)$, while \mathbf{X} stands for X^n . The probability of an event \mathcal{A} is denoted by $\mathbb{P}[\mathcal{A}]$, while $\mathbb{P}[\mathcal{A}|\mathcal{B}]$ denotes conditional probability of \mathcal{A} given \mathcal{B} .

Probability mass functions (PMFs) are denoted by the capital letter P with a subscript that identifies the random variable and its possible conditioning. For example, for two jointly distributed random variables X and Y , let P_X , $P_{X,Y}$, and $P_{X|Y}$ denote, respectively, the PMF of X , the joint PMF of (X, Y) , and the conditional PMF of X given Y . In particular, when X and Y are discrete, $P_{X|Y}$ represents the stochastic matrix whose elements are given by $P_{X|Y}(x|y) = \mathbb{P}[X = x|Y = y]$. We omit the subscripts if the arguments of the distribution are lower case versions of the random variables.

A. FSM-MAC With a Common Message and Delayed CSI

The FSM-MAC with a common message considered in this paper is illustrated in Fig. 2. The MAC setting consists of two senders and one receiver. Each sender $j \in \{1, 2\}$ chooses a pair of indices, (m_0, m_j) , uniformly from the set $\{1, \dots, 2^{nR_0}\} \times \{1, \dots, 2^{nR_j}\}$, where m_0 denotes the common message and m_j , $j \in \{1, 2\}$, denotes the private message of the corresponding sender. The choices of m_0 , m_1 and m_2 are independent. The input to the channel from encoder $j \in \{1, 2\}$ is denoted by $X_j^n = (X_{j,1}, X_{j,2}, \dots, X_{j,n})$, and the output of the channel is denoted by $Y^n = (Y_1, Y_2, \dots, Y_n)$.

At each instance of time, the FSM channel is assumed to be in one of a finite number of states $\mathcal{S} = \{s_1, s_2, \dots, s_k\}$. In each state, the channel is a discrete memoryless channel (DMC), with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and output alphabet \mathcal{Y} . Let the random variable S_i denote the channel state at time i . Similarly, we denote by $X_{1,i}, X_{2,i}$ and Y_i the inputs and the output of the channel at time i . The channel transition probability distribution at time i depends on the state S_i and the inputs $X_{1,i}, X_{2,i}$ at time i , and it is given by $P(y_i|x_{1,i}, x_{2,i}, s_i)$. The channel output at any time i is assumed to depend only on the channel inputs and state at time i . Hence,

$$P(y_i|x_{1,i}, x_{2,i}, s_i) = P(y_i|x_{1,i}, x_{2,i}, s_i). \quad (1)$$

The state process, $\{S_i\}_{i=1}^n$, is assumed to be an irreducible, aperiodic, finite-state, homogeneous and stationary Markov chain and is therefore ergodic. The state process is independent of the channel inputs and output when conditioned on the previous states, i.e.,

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

Furthermore, we assume that the state process is independent of the messages M_0, M_1 and M_2 , i.e.,

$$P(s^n, m_0, m_1, m_2) = \prod_{i=1}^n P(s_i|s_{i-1})P(m_0)P(m_1)P(m_2). \quad (3)$$

We assume that full CSI is available at the decoder (i.e., the decoder knows S_i at each time instance i). However, the encoders are only assumed to have access to delayed CSI, with delays d_1 and d_2 for Encoder 1 and Encoder 2, respectively.

We let S_{i-d_j} , $j \in 1, 2$, denote the channel state at time $i - d_j$, and assume without loss of generality that $d_1 \geq d_2$. Now, let \mathbf{K} be the one-step state-transition probability matrix of the Markov process that governs the channel states, and let π be its steady state probability distribution. The joint distribution of (S_i, S_{i-d}) is stationary and is given by

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

where $\mathbf{K}^d(s_l, s_j)$ is the (l, j) -th element of the d -step transition probability matrix \mathbf{K}^d of the Markov state process. To simplify the notation, we define the joint distribution of the random variables $(S, \tilde{S}_1, \tilde{S}_2)$ as the joint distribution of $(S_i, S_{i-d_1}, S_{i-d_2})$, i.e.,

$$P_{S, \tilde{S}_1, \tilde{S}_2}(s_l, s_j, s_v) = \pi(s_j)\mathbf{K}^{d_1-d_2}(s_v, s_j)\mathbf{K}^{d_2}(s_l, s_v), \quad (5)$$

where $(s_j, s_l, s_v) \in \mathcal{S}^3$.

Definition 1 (Code Description): A $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code for the FSM-MAC with CSI at the decoder and delayed CSI at the encoders with delays d_1 and d_2 consists of:

- 1) Three sets of integers $\mathcal{M}_0 = \{1, 2, \dots, 2^{nR_0}\}$, $\mathcal{M}_1 = \{1, 2, \dots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, \dots, 2^{nR_2}\}$, referred to as the message sets.
- 2) Two encoding functions f_j , $j \in \{1, 2\}$. Each function f_j is defined by means of a sequence of functions $f_{j,i}$, $i \in \{1, 2, \dots, n\}$, that depend only on the pair of messages (M_0, M_j) , and the channel states up to time $i - d_j$. The output of Encoder j at time i , $X_{j,i} \in \mathcal{X}_j$, is given by

$$X_{j,i} = \begin{cases} f_{j,i}(M_0, M_j), & 1 \leq i \leq d_j \\ f_{j,i}(M_0, M_j, S^{i-d_j}), & d_j + 1 \leq i \leq n. \end{cases} \quad (6)$$

- 3) A decoding function:

$$\psi : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2. \quad (7)$$

The average probability of error for the $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code is given in (8) at the bottom of the page. We use standard definitions of achievability and of the capacity region [48]. Namely, a rate triplet (R_0, R_1, R_2) is *achievable* for the FSM-MAC if there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. The *capacity region* is the closure of the set of achievable rates (R_0, R_1, R_2) .

B. FSM-MAC With Partially Cooperative Encoders and Delayed CSI

The FSM-MAC with partially cooperative encoders and delayed CSI is depicted in Fig. 1. The channel definition relies on Subsection II-A, while taking the common message set to be $\mathcal{M}_0 = \emptyset$. Here, however, conferencing between the encoders is introduced under the assumption that conferencing links of fixed and finite capacities C_{12} and C_{21} exist between the encoders. Accordingly, the amount of information exchanged between the encoders during the conference is

$$P_e^{(n)} = \frac{1}{2^{n(R_0+R_1+R_2)}} \sum_{(m_0, m_1, m_2)} \sum_{s^n} P_{S^n}(s^n) \mathbb{P} \left[\psi(Y^n, s^n) \neq (m_0, m_1, m_2) \mid (M_0, M_1, M_2) = (m_0, m_1, m_2), S^n = s^n \right] \quad (8)$$

bounded by C_{12} and C_{21} . The conference is assumed to take place prior to the transmission of a codeword through the channel and consists of ℓ consecutive pairs of communications, simultaneously transmitted by the encoders. Each communication depends on the message to be transmitted by the sending encoder and previously *received* communications from the other encoder. We denote the communications transmitted from encoder $j \in \{1, 2\}$ to the other encoder by V_j^ℓ . Note that here the state process is also assumed to be independent of the conference communications, i.e.,

$$P(s^n, v_1^\ell, v_2^\ell) = P(s^n)P(v_1^\ell, v_2^\ell) = \prod_{i=1}^n P(s_i | s_{i-1})P(v_1^\ell, v_2^\ell). \quad (9)$$

Definition 2 Code Description: A $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code for the FSM-MAC with CSI at the decoder, delayed CSI at the encoders with delays d_1 and d_2 , and conferencing links with capacities C_{12} and C_{21} consists of:

- 1) Two sets of integers $\mathcal{M}_1 = \{1, 2, \dots, 2^{nR_1}\}$ and $\mathcal{M}_2 = \{1, 2, \dots, 2^{nR_2}\}$, referred to as the message sets.
- 2) Two encoders, where each encoder is completely described by an encoding function, f_j , and a set of ℓ ($\ell \geq 1$) communication functions, $\{h_{j,1}, h_{j,2}, \dots, h_{j,\ell}\}$, $j \in \{1, 2\}$ (similar definitions were also used in [4]).
- 3) The encoding function, f_j , maps the message M_j , $j \in \{1, 2\}$, and what was learned from the conference with the other encoder into channel codewords of length n . Each function f_j is defined by means of a sequence of functions $f_{j,i}$ that depend only on the message M_j , the received communications from the other encoder in the conferencing stage, and the channel states up to time $i - d_j$. We emphasize that since encoding occurs only after the conferencing stage has finished, each $f_{j,i}$ depends on all received communications.
- 4) Each of the two communication functions $h_{1,i}$ and $h_{2,i}$, $i \in \{1, 2, \dots, \ell\}$, maps the message M_1 (or M_2 , respectively) and the sequence of previously received communications from the other encoder V_2^{i-1} (or V_1^{i-1} , respectively), onto the i -th communication $V_{1,i}$ (or $V_{2,i}$, respectively). More specifically, the communications are defined as:

$$V_{1,i} = h_{1,i}(M_1, V_2^{i-1}); \quad V_{2,i} = h_{2,i}(M_2, V_1^{i-1}). \quad (10)$$

- 5) The encoding function for Encoder 1 satisfies

$$X_{1,i} = \begin{cases} f_{1,i}(M_1, V_2^\ell), & 1 \leq i \leq d_1 \\ f_{1,i}(M_1, V_2^\ell, S^{i-d_1}), & d_1 + 1 \leq i \leq n, \end{cases} \quad (11)$$

and the encoding function for Encoder 2 is defined analogously (using the private message M_2 , the communications V_1^ℓ and the delay d_2).

- 6) The random variable $V_{j,i}$, for $j \in \{1, 2\}$ and $i \in \{1, 2, \dots, \ell\}$ ranges over the finite alphabet $\mathcal{V}_{j,i}$.

A conference is (C_{12}, C_{21}) -permissible if the sets of communication functions are such that [4]:

$$\sum_{i=1}^{\ell} \log |\mathcal{V}_{1,i}| \leq nC_{12}; \quad \sum_{i=1}^{\ell} \log |\mathcal{V}_{2,i}| \leq nC_{21}. \quad (12)$$

- 7) A decoding function:

$$\psi : \mathcal{Y}^n \times \mathcal{S}^n \rightarrow \mathcal{M}_1 \times \mathcal{M}_2. \quad (13)$$

The average probability of error for the $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code is given by (14) at the bottom of the page. The achievable rates and the capacity region for this channel are defined analogously to their definitions in Section II-A.

III. THE CAPACITY REGION OF THE FSM-MAC WITH A COMMON MESSAGE AND DELAYED TRANSMITTER CSI

In this section we state the capacity region of the FSM-MAC with a common message and delayed transmitter CSI, after which we present its proof.

Theorem 1: The capacity region of the FSM-MAC with a common message, CSI at the decoder and asymmetrically delayed CSI at the encoders with delays d_1 and d_2 , such that $d_1 \geq d_2$, is the union of all sets of rate triplets $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ satisfying:

$$R_1 \leq I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) \quad (15a)$$

$$R_2 \leq I(X_2; Y | X_1, U, S, \tilde{S}_1, \tilde{S}_2) \quad (15b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | U, S, \tilde{S}_1, \tilde{S}_2) \quad (15c)$$

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2), \quad (15d)$$

where the union is over all joint distributions $P_{U|\tilde{S}_1} P_{X_1|\tilde{S}_1,U} P_{X_2|\tilde{S}_1,\tilde{S}_2,U}$. The joint distribution of $(S, \tilde{S}_1, \tilde{S}_2)$ is specified in (5) and $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| \cdot |\mathcal{S}|^3 + 2$. Furthermore, the capacity region is convex.

Proof:

A. Converse

We need to show that for every achievable rate triplet (R_0, R_1, R_2) , there exists a joint distribution $P_{S,\tilde{S}_1,\tilde{S}_2} P_{U|\tilde{S}_1} P_{X_1|\tilde{S}_1,U} P_{X_2|\tilde{S}_1,\tilde{S}_2,U} P_{Y|X_1,X_2,S}$ such that the inequalities in (15) are satisfied. Since (R_0, R_1, R_2) is an achievable rate triplet, there exists a $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code with a probability of error $P_e^{(n)}$ that becomes arbitrarily small with the increase of the block length (see (8)). By Fano's inequality,

$$H(M_0, M_1, M_2 | Y^n, S^n) \leq n(R_0 + R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n, \quad (16)$$

where clearly $\epsilon_n \rightarrow 0$ as $P_e^{(n)} \rightarrow 0$. It therefore follows that

$$H(M_1 | Y^n, S^n) \leq H(M_0, M_1, M_2 | Y^n, S^n) \leq n\epsilon_n \quad (17)$$

$$H(M_2 | Y^n, S^n) \leq H(M_0, M_1, M_2 | Y^n, S^n) \leq n\epsilon_n \quad (18)$$

$$H(M_1, M_2 | Y^n, S^n) \leq H(M_0, M_1, M_2 | Y^n, S^n) \leq n\epsilon_n. \quad (19)$$

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{(m_1, m_2)} \sum_{s^n} P_{S^n}(s^n) \mathbb{P}[\psi(Y^n, s^n) \neq (m_1, m_2) \mid (M_1, M_2) = (m_1, m_2), S^n = s^n] \quad (14)$$

For the sake of brevity we focus here on the upper bound on R_1 , while noting that all other upper bounds in (15) can be analogously derived using the same auxiliary random variable definition. It now follows that

$$\begin{aligned}
nR_1 &= H(M_1) \\
&\stackrel{(a)}{\leq} I(M_1; Y^n, S^n) + n\epsilon_n \\
&\stackrel{(b)}{=} I(M_1; Y^n | S^n) + n\epsilon_n \\
&\stackrel{(c)}{\leq} H(M_1 | S^n, M_0, M_2) - H(M_1 | Y^n, S^n, M_0, M_2) + n\epsilon_n \\
&\stackrel{(d)}{=} \sum_{i=1}^n I(M_1; Y_i | S^n, M_0, M_2, Y^{i-1}) + n\epsilon_n \\
&\stackrel{(e)}{=} \sum_{i=1}^n \left[H(Y_i | S^n, X_2^n, M_0, M_2, Y^{i-1}) \right. \\
&\quad \left. - H(Y_i | S^n, X_1^n, X_2^n, M_0, M_1, M_2, Y^{i-1}) \right] + n\epsilon_n \\
&\stackrel{(f)}{\leq} \sum_{i=1}^n \left[H(Y_i | X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, M_0, S^{i-d_1-1}) \right. \\
&\quad \left. - H(Y_i | X_{1,i}, X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, M_0, S^{i-d_1-1}) \right] + n\epsilon_n \\
&\stackrel{(g)}{=} \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + n\epsilon_n \quad (20)
\end{aligned}$$

where:

- (a) follows from (17);
- (b) follows because M_1 and S^n are independent;
- (c) follows because M_1 and (M_0, M_2) are independent given S^n (first term) and since conditioning reduces entropy (second term);
- (d) follows by the mutual information chain rule;
- (e) follows because X_1^n is a deterministic function of (M_0, M_1, S^n) and X_2^n is a deterministic function of (M_0, M_2, S^n) ;
- (f) follows since conditioning reduces entropy (first term), and because when conditioned on $(X_{1,i}, X_{2,i})$ and S_i , the channel output at time i is independent of $(M_0, M_1, M_2, S^{i-1}, S_{i+1}^n, X_{1,i+1}^{i-1}, X_{1,i+1}^n, X_{2,i+1}^{i-1}, X_{2,i+1}^n, Y^{i-1})$ (second term);
- (g) follows by defining $U_i \triangleq (M_0, S^{i-d_1-1})$.

Note that the definition of the auxiliary random variable U_i represents the common message and the common knowledge of the state sequence at time i (except for S_{i-d_1}), which, in fact, encompasses all common information shared by the two encoders at this instant of time. We can therefore conclude that the rate R_1 must satisfy the following upper bound:

$$R_1 \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + \epsilon_n. \quad (21)$$

In a completely analogous manner it can be shown that

$$R_2 \leq \frac{1}{n} \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + \epsilon_n \quad (22)$$

$$\begin{aligned}
R_1 + R_2 &\leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_i, S_{i-d_1}, S_{i-d_2}, U_i) + \epsilon_n \quad (23)
\end{aligned}$$

$$\begin{aligned}
R_0 + R_1 + R_2 &\leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_i, S_{i-d_1}, S_{i-d_2}) + \epsilon_n. \quad (24)
\end{aligned}$$

The upper bounds in (21)–(24) can also be rewritten by introducing a new time sharing random variable Q that is uniformly distributed over the set $\{1, 2, \dots, n\}$. For example, the upper bound in (21) can be rewritten as

$$\begin{aligned}
R_1 &\leq \frac{1}{n} \sum_{i=1}^n I(Y_Q; X_{1,Q} | X_{2,Q}, S_Q, S_{Q-d_1}, S_{Q-d_2}, U_Q, Q=i) \\
&\quad + \epsilon_n \\
&= I(Y_Q; X_{1,Q} | X_{2,Q}, S_Q, S_{Q-d_1}, S_{Q-d_2}, U_Q, Q) + \epsilon_n. \quad (25)
\end{aligned}$$

By rewriting the rate bounds (22)–(24) in the same manner as (21) is rewritten into (25), it is clear that the obtained region is convex. This follows directly by the presence of the time sharing random variable Q in the conditioning of all the mutual information terms.

Next, by denoting $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, $Y \triangleq Y_Q$, $S \triangleq S_Q$, $\tilde{S}_1 \triangleq S_{Q-d_1}$, $\tilde{S}_2 \triangleq S_{Q-d_2}$ and $U \triangleq (U_Q, Q)$, we get:

$$R_1 \leq I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n \quad (26a)$$

$$R_2 \leq I(X_2; Y | X_1, U, S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n \quad (26b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | U, S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n \quad (26c)$$

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n, \quad (26d)$$

where (26d) holds due to

$$\begin{aligned}
R_0 + R_1 + R_2 &\leq I(X_{1,Q}, X_{2,Q}; Y_Q | S_Q, S_{Q-d_1}, S_{Q-d_2}, Q) + \epsilon_n \\
&\stackrel{(a)}{\leq} I(X_{1,Q}, X_{2,Q}, U_Q; Y_Q | S_Q, S_{Q-d_1}, S_{Q-d_2}, Q) + \epsilon_n \\
&\stackrel{(b)}{\leq} I(X_{1,Q}, X_{2,Q}, U_Q, Q; Y_Q | S_Q, S_{Q-d_1}, S_{Q-d_2}) + \epsilon_n \\
&\stackrel{(c)}{\leq} I(X_1, X_2, U; Y | S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n \\
&\stackrel{(d)}{=} I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n. \quad (27)
\end{aligned}$$

Here:

- (a) and (b) follow from the fact that conditioning reduces entropy;
- (c) follows from the definition of $(X_1, X_2, Y, U, S, \tilde{S}_1, \tilde{S}_2)$;
- (d) follows from the Markov relation $Y - (X_1, X_2, S) - (U, \tilde{S}_1, \tilde{S}_2)$, which is induced from the channel model. Taking the limit as $n \rightarrow \infty$, one obtains the bounds as in (15).

To complete the proof of the converse, it is left to show that the following Markov relations hold:

$$U - \tilde{S}_1 - (S, \tilde{S}_2) \quad (28a)$$

$$X_1 - (\tilde{S}_1, U) - (S, \tilde{S}_2) \quad (28b)$$

$$X_2 - (\tilde{S}_1, \tilde{S}_2, U) - (X_1, S). \quad (28c)$$

The proof of (28) is given in Appendix A.

B. Achievability

To establish achievability, we need to show that for a fixed $\epsilon > 0$, a fixed distribution

$$P_{U|\tilde{S}_1} P_{X_1|U, \tilde{S}_1} P_{X_2|U, \tilde{S}_1, \tilde{S}_2}, \quad (29)$$

and rates (R_0, R_1, R_2) that satisfy the inequalities in (15), there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes such that $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

Without loss of generality, we assume that the finite-state space is the set $\mathcal{S} = \{1, 2, \dots, k\}$. By the underlying assumptions of the channel model, we take the delays (d_1, d_2) to be fixed and finite integers. Moreover, throughout this proof we use the following notations. For an arbitrary finite set $\mathcal{A} = \{a_1, a_2, \dots, a_{|\mathcal{A}|}\}$, we denote by $(x_a)_{a \in \mathcal{A}}$ a column vector of size $|\mathcal{A}|$ with elements $\{x_{a_1}, x_{a_2}, \dots, x_{a_{|\mathcal{A}|}}\}$. As stated in Section II, sequences of length n are denoted by bold lowercase letters, while random sequences are denoted by bold uppercase letters. Consider now the following encoding and decoding scheme.

1) *Codebook Generation*: Generate a common message codebook that comprises 2^{nR_0} codewords $\mathbf{t}_0(m_0)$, $m_0 \in \mathcal{M}_0$, assembled from n symbols from the super-alphabet $\mathcal{U}^{|\mathcal{S}|}$, which are drawn in an i.i.d. manner. Each codeword $\mathbf{t}_0(m_0)$ is distributed according to the product distribution

$$\mathbb{P}[\mathbf{T}_0 = \mathbf{t}_0] = \prod_{i=1}^n P_{T_0}(t_{0,i}), \quad (30)$$

where $t_{0,i} \in \mathcal{U}^{|\mathcal{S}|}$, for $i \in \{1, 2, \dots, n\}$ (each $t_{0,i}$ can thus be treated as a column vector of size $|\mathcal{S}|$ with elements in \mathcal{U} ordered by the natural order of the set \mathcal{S}), and

$$P_{T_0}(t_0) = \mathbb{P}[T_0 = t_0] = \prod_{\tilde{s}_1 \in \mathcal{S}} P_{U|\tilde{s}_1=\tilde{s}_1}(u_{\tilde{s}_1}|\tilde{s}_1), \quad (31)$$

where $t_0 = (u_{\tilde{s}_1})_{\tilde{s}_1 \in \mathcal{S}}$. Each codeword can hence be viewed as a matrix of dimension $|\mathcal{S}| \times n$ with elements in \mathcal{U} , where each row is associated with a different (delayed) state $\tilde{s}_1 \in \mathcal{S}$. Accordingly, we denote by $t_{0,i}(m_0, \tilde{s}_{1,i})$ the $\tilde{s}_{1,i}$ th element of the i th symbol of the codeword $\mathbf{t}_0(m_0)$.

Next, generate the codebook associated with the private message set \mathcal{M}_1 that comprises 2^{nR_1} codewords $\mathbf{t}_1(m_1)$, $m_1 \in \mathcal{M}_1$, assembled from n symbols from the super-alphabet $\mathcal{X}_1^{|\mathcal{U}| \cdot |\mathcal{S}|}$ drawn in an i.i.d. manner. Accordingly, the distribution of each codeword $\mathbf{t}_1(m_1)$ is given by

$$\mathbb{P}[\mathbf{T}_1 = \mathbf{t}_1] = \prod_{i=1}^n P_{T_1}(t_{1,i}), \quad (32)$$

where $t_{1,i} \in \mathcal{X}_1^{|\mathcal{U}| \cdot |\mathcal{S}|}$, for $i \in \{1, 2, \dots, n\}$ (here each $t_{1,i}$ can be treated as a column vector of size $|\mathcal{U}| \cdot |\mathcal{S}|$ with elements in \mathcal{X}_1), and

$$P_{T_1}(t_1) = \mathbb{P}[T_1 = t_1] = \prod_{(u, \tilde{s}_1) \in \mathcal{U} \times \mathcal{S}} P_{X_1|U=u, \tilde{s}_1=\tilde{s}_1}(x_{1,(u, \tilde{s}_1)}|u, \tilde{s}_1), \quad (33)$$

where $t_1 = (x_{1,(u, \tilde{s}_1)})_{(u, \tilde{s}_1) \in \mathcal{U} \times \mathcal{S}}$. Each codeword can therefore be viewed as a matrix of dimension $(|\mathcal{U}| \cdot |\mathcal{S}|) \times n$ with elements in \mathcal{X}_1 , where each row is associated with a different pair $(u, \tilde{s}_1) \in \mathcal{U} \times \mathcal{S}$. The element of the i th symbol of the codeword $\mathbf{t}_1(m_1)$ that is associated with the pair $(t_{0,i}(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})$ is denoted by $t_{1,i}(m_1, t_{0,i}(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})$.

Finally, generate the codebook associated with the private message set \mathcal{M}_2 in a manner analogous to codebook generation for \mathcal{M}_1 , but here, the symbols of each codeword

are elements in $\mathcal{X}_2^{|\mathcal{U}| \cdot |\mathcal{S}|^2}$. Namely, the distribution of each codeword $\mathbf{t}_2(m_2)$, $m_2 \in \mathcal{M}_2$, is

$$\mathbb{P}[\mathbf{T}_2 = \mathbf{t}_2] = \prod_{i=1}^n P_{T_2}(t_{2,i}), \quad (34)$$

where $t_{2,i} \in \mathcal{X}_2^{|\mathcal{U}| \cdot |\mathcal{S}|^2}$, for $i \in \{1, 2, \dots, n\}$, and

$$P_{T_2}(t_2) = \mathbb{P}[T_2 = t_2] = \prod_{(u, \tilde{s}_1, \tilde{s}_2) \in \mathcal{U} \times \mathcal{S}^2} P_{X_2|U=u, \tilde{s}_1=\tilde{s}_1, \tilde{s}_2=\tilde{s}_2}(x_{2,(u, \tilde{s}_1, \tilde{s}_2)}|u, \tilde{s}_1, \tilde{s}_2), \quad (35)$$

where $t_2 = (x_{2,(u, \tilde{s}_1, \tilde{s}_2)})_{(u, \tilde{s}_1, \tilde{s}_2) \in \mathcal{U} \times \mathcal{S}^2}$. Again, each codeword can be viewed as a matrix of dimension $(|\mathcal{U}| \cdot |\mathcal{S}|^2) \times n$ with elements in \mathcal{X}_2 , where each row is associated with a different triplet $(u, \tilde{s}_1, \tilde{s}_2) \in \mathcal{U} \times \mathcal{S}^2$. We denote by $t_{2,i}(m_2, t_{0,i}(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i})$ the element of the i th symbol of the codeword $\mathbf{t}_2(m_2)$ that is associated with the triplet $(t_{0,i}(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i})$. All codebooks are revealed to both encoders and to the decoder.

2) *Encoding*: To transmit the message triplet $(m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$, the encoders adhere to the following scheme. First, note that the delayed CSI becomes available at *both* encoders only after the first d_1 channel uses. Therefore, the initial d_1 symbols transmitted by Encoder j , $j \in \{1, 2\}$, are arbitrarily chosen from \mathcal{X}_j (note that the choice of these symbols has no effect on the achievable rates since d_1 is fixed and finite, see the discussion in the sequel). The manner in which the encoders choose the symbols transmitted at times $i \in \{d_1 + 1, \dots, n\}$ is described next.

Encoder 1: At each time $i \in \{d_1 + 1, \dots, n\}$, Encoder 1 has access to $\tilde{s}_{1,i}$. This delayed CSI is first used to choose an element from the codeword associated with the common message m_0 . Namely, Encoder 1 starts by choosing $u_i(m_0, \tilde{s}_{1,i}) \triangleq t_{0,i}(m_0, \tilde{s}_{1,i})$. Having $u_i(m_0, \tilde{s}_{1,i})$, Encoder 1 refers to the codeword $\mathbf{t}_1(m_1)$ and chooses $x_{1,i}(m_1, u_i(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}) \triangleq t_{1,i}(m_1, u_i(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})$. Encoder 1 then sends $x_{1,i}(m_1, u_i(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})$ to the channel.

Encoder 2: Encoder 2 selects its channel input in manner analogous to that used by Encoder 1. First, recall that at each time $i \in \{d_1 + 1, \dots, n\}$, Encoder 2 has access to both $(\tilde{s}_{1,i}, \tilde{s}_{2,i})$. Thus, using $\tilde{s}_{1,i}$, the encoder first chooses $u_i(m_0, \tilde{s}_{1,i}) = t_{0,i}(m_0, \tilde{s}_{1,i})$. The i th channel input from Encoder 2 is then chosen to be $x_{2,i}(m_2, u_i(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i}) \triangleq t_{2,i}(m_2, u_i(m_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i})$.

An illustration of the codewords' structure for some pair $(m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1$ and the corresponding transmitted symbols is shown in Fig. 3. The structure of the codewords associated with the message set \mathcal{M}_2 is analogous, and is therefore omitted from the figure. We also note that the above construction of the codebooks and transmitted sequences can be regarded as a manifestation of the functional representation lemma [51, Appendix B].

3) *Decoding Process*: Upon receiving the whole channel output sequence \mathbf{y} and the sequence of channel states \mathbf{s} (assumed fully known at the receiver), a joint typicality

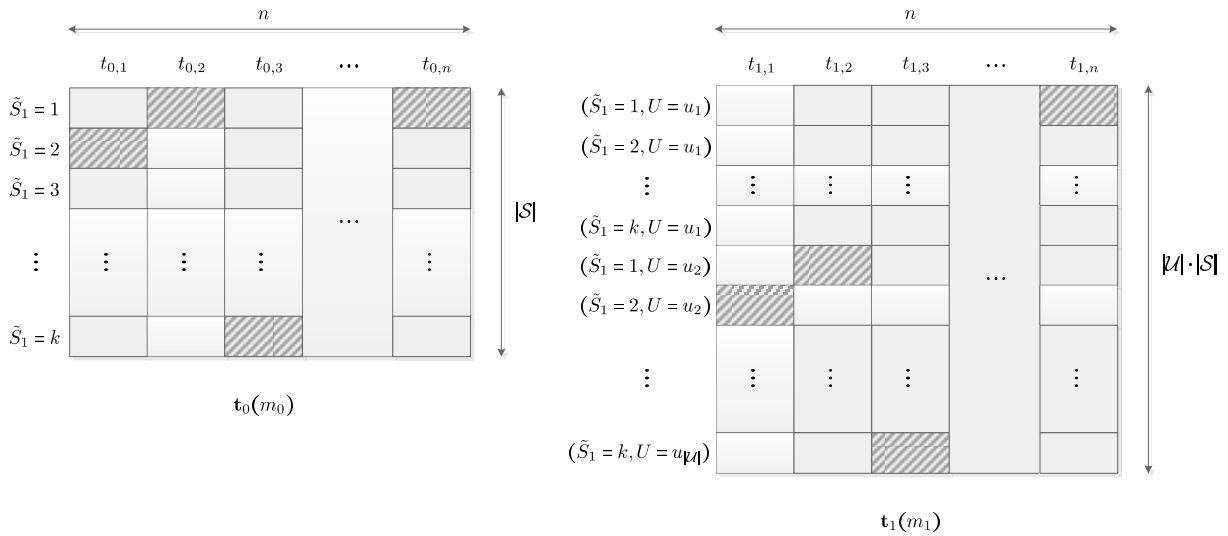


Fig. 3. The codewords $\mathbf{t}_0(m_0)$ and $\mathbf{t}_1(m_1)$ that are associated with some message pair $(m_0, m_1) \in \mathcal{M}_0 \times \mathcal{M}_1$. The shaded regions correspond to the symbols chosen by Encoder 1 when the sequence of delayed CSI realizations is $\tilde{s}_1 = (2, 1, k, \dots, 1)$, and when $t_{0,1}(m_0, 2) = u_2$, $t_{0,2}(m_0, 1) = u_2$, $t_{0,3}(m_0, k) = u_{|A|}$ and $t_{0,n}(m_0, 1) = u_1$.

decision rule is employed to decode the transmitted sequences. Note that since the delayed CSI available to each of the encoders is a deterministic function of the state sequence, the decoder can immediately reconstruct the sequences \tilde{s}_1 and \tilde{s}_2 from the latter.

The decoder searches for a triplet $(\hat{m}_0, \hat{m}_1, \hat{m}_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$ such that (36) at the bottom of the page is satisfied, where $\mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y)$ denotes the jointly ϵ -strongly typical set as defined in [51, Ch. 2], and

$$\begin{aligned} \mathbf{u}(\hat{m}_0, \tilde{s}_1) &\triangleq \{u_i(\hat{m}_0, \tilde{s}_{1,i})\}_{i=1}^n = \{t_{0,i}(\hat{m}_0, \tilde{s}_{1,i})\}_{i=1}^n \\ \mathbf{x}_1(\hat{m}_1, \mathbf{u}(\hat{m}_0, \tilde{s}_1), \tilde{s}_1) &\triangleq \left\{x_{1,i}(\hat{m}_1, u_i(\hat{m}_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})\right\}_{i=1}^n \\ &= \left\{t_{1,i}(\hat{m}_1, u_i(\hat{m}_0, \tilde{s}_{1,i}), \tilde{s}_{1,i})\right\}_{i=1}^n \\ \mathbf{x}_2(\hat{m}_2, \mathbf{u}(\hat{m}_0, \tilde{s}_1), \tilde{s}_1, \tilde{s}_2) &\triangleq \left\{x_{2,i}(\hat{m}_2, u_i(\hat{m}_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i})\right\}_{i=1}^n \\ &= \left\{t_{2,i}(\hat{m}_2, u_i(\hat{m}_0, \tilde{s}_{1,i}), \tilde{s}_{1,i}, \tilde{s}_{2,i})\right\}_{i=1}^n. \end{aligned}$$

If such a unique triplet is found, it is declared as the decoded message triplet; otherwise, the decoder outputs a random message triplet.

We note here that although Encoder 1 (respectively, Encoder 2) arbitrarily chooses the first d_1 symbols of each codeword, these choices have a vanishing effect on the typicality test (36). This is since the underlying assumption of the channel model is that both d_1 and d_2 are fixed, while the total block length n can grow without bound. Therefore, to simplify the analysis that follows, we henceforth ignore the fact that the first symbols of each transmitted sequence do not follow the prescribed input distribution (29).

By error probability analysis (see Appendix B), we get that for the probability of error to vanish as $n \rightarrow \infty$, the rate constraints in (15) must be satisfied. We have thus shown that if (R_0, R_1, R_2) is inside the rate region specified in Theorem 1, then there exists a sequence of $(n, 2^{nR_0}, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes such that $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. This completes the proof of the achievability part. \blacksquare

Remark 1: The cardinality bound on the auxiliary random variable U is straightforwardly established using the convex cover method (see [51, Appendix C] for details) and is therefore omitted.

Remark 2: The capacity region of the corresponding channel in which no CSI is available to Encoder 1 can be obtained from the capacity region in Theorem 1 by omitting \tilde{S}_1 from all mutual information expressions in (15) and from the joint distribution over which the union is taken. Similarity, when considering the case in which the CSI at both encoders is absent, the capacity region can also be obtained from the result in Theorem 1 by omitting \tilde{S}_1 and \tilde{S}_2 from the corresponding expressions. Now, based on the underlying assumptions on the properties of the Markov state process, when d_1 is increased without bound, the delayed channel state S_{i-d_1} becomes independent of the pair (S_i, S_{i-d_2}) , which, in turn, implies that \tilde{S}_1 and (S, \tilde{S}_2) are independent (see (5)). Thus, it can be shown that when d_1 is increased without bound, the capacity region in Theorem 1 approaches the corresponding region of the case in which no CSI is available to Encoder 1. Using similar arguments, one can show that when d_2 is also increased without bound, the capacity region reduces to the corresponding capacity region with no encoder CSI.

$$\left(\mathbf{u}(\hat{m}_0, \tilde{s}_1), \mathbf{x}_1(\hat{m}_1, \mathbf{u}(\hat{m}_0, \tilde{s}_1), \tilde{s}_1), \mathbf{x}_2(\hat{m}_2, \mathbf{u}(\hat{m}_0, \tilde{s}_1), \tilde{s}_1, \tilde{s}_2), \mathbf{s}, \tilde{s}_1, \tilde{s}_2, \mathbf{y}\right) \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y), \quad (36)$$

Remark 3: Based on practical considerations, a rate-splitting and multiplexing coding scheme as used, in [43], can be considered for the current setting. We note, however, that the single codebooks approach employed here exhibits a simpler construction, and also lends itself more easily to error probability analysis.

IV. THE CAPACITY REGION OF THE FSM-MAC WITH PARTIALLY COOPERATIVE ENCODERS AND DELAYED TRANSMITTER CSI

In this section we state the capacity region of the FSM-MAC with partially cooperative encoders and delayed transmitter CSI followed by its proof.

Theorem 2: The capacity region of FSM-MAC with partially cooperative encoders, cooperation link capacities C_{12} and C_{21} , CSI at the decoder and asymmetrically delayed CSI at the encoders with delays d_1 and d_2 , such that $d_1 \geq d_2$, is the union of all sets of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying:

$$R_1 \leq I(X_1; Y|X_2, U, S, \tilde{S}_1, \tilde{S}_2) + C_{12} \quad (37a)$$

$$R_2 \leq I(X_2; Y|X_1, U, S, \tilde{S}_1, \tilde{S}_2) + C_{21} \quad (37b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|U, S, \tilde{S}_1, \tilde{S}_2) + C_{12} + C_{21} \quad (37c)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2), \quad (37d)$$

where the union is over all distribution $P_{U|\tilde{S}_1} P_{X_1|\tilde{S}_1,U} P_{X_2|\tilde{S}_1,\tilde{S}_2,U}$. The joint distribution of $(S, \tilde{S}_1, \tilde{S}_2)$ is specified in (5) and $|\mathcal{U}| \leq |\mathcal{X}_1| \cdot |\mathcal{X}_2| \cdot |S|^3 + 2$. Furthermore, the capacity region is convex.

Proof:

A. Converse

Given an achievable rate (R_1, R_2) , we need to show that there exists a joint distribution of the form $P_{S,\tilde{S}_1,\tilde{S}_2} P_{U|\tilde{S}_1} P_{X_1|\tilde{S}_1,U} P_{X_2|\tilde{S}_1,\tilde{S}_2,U} P_{Y|X_1,X_2,S}$ such that the inequalities in (37) are satisfied. Since (R_1, R_2) is an achievable rate-pair, there exists an $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ code with an arbitrarily small error probability $P_e^{(n)}$. By Fano's inequality (and with some abuse of notation),

$$H(M_1, M_2|Y^n, S^n) \leq n(R_1 + R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n, \quad (38)$$

where $\epsilon_n \rightarrow 0$ as $P_e^{(n)} \rightarrow 0$. It hence follows that

$$H(M_1|Y^n, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\epsilon_n \quad (39)$$

$$H(M_2|Y^n, S^n) \leq H(M_1, M_2|Y^n, S^n) \leq n\epsilon_n. \quad (40)$$

As in the proof of Theorem 1, we focus on the upper bound on R_1 and note that the upper bounds on all other rates can be straightforwardly obtained in an analogous manner. For R_1

we have the following:

$$\begin{aligned} nR_1 &= H(M_1) \\ &\stackrel{(a)}{\leq} I(M_1; Y^n, S^n) + n\epsilon_n \\ &\stackrel{(b)}{=} I(M_1; Y^n|S^n) + n\epsilon_n \\ &\stackrel{(c)}{\leq} H(M_1|S^n, M_2) - H(M_1|V_1^\ell, V_2^\ell, Y^n, S^n, M_2) + n\epsilon_n \\ &\stackrel{(d)}{=} I(M_1; V_1^\ell, V_2^\ell|S^n, M_2) + I(M_1; Y^n|V_1^\ell, V_2^\ell, S^n, M_2) \\ &\quad + n\epsilon_n \\ &\stackrel{(e)}{=} H(V_1^\ell|S^n, M_2) + \sum_{i=1}^n I(M_1; Y_i|V_1^\ell, V_2^\ell, S^n, M_2, Y^{i-1}) \\ &\quad + n\epsilon_n \\ &\stackrel{(f)}{\leq} H(V_1^\ell) + \sum_{i=1}^n \left[H(Y_i|V_1^\ell, V_2^\ell, S^n, X_2^n, M_2, Y^{i-1}) \right. \\ &\quad \left. - H(Y_i|V_1^\ell, V_2^\ell, S^n, X_1^n, X_2^n, M_1, M_2, Y^{i-1}) \right] + n\epsilon_n \\ &\stackrel{(g)}{\leq} \sum_{j=1}^{\ell} H(V_{1,j}) \\ &\quad + \sum_{i=1}^n \left[H(Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1}) \right. \\ &\quad \left. - H(Y_i|X_{1,i}, X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1}) \right] \\ &\quad + n\epsilon_n \\ &\leq \sum_{j=1}^{\ell} \log |V_{1,j}| \\ &\quad + \sum_{i=1}^n I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, V_1^\ell, V_2^\ell, S^{i-d_1-1}) \\ &\quad + n\epsilon_n \\ &\stackrel{(h)}{\leq} nC_{12} + \sum_{i=1}^n I(X_{1,i}; Y_i|X_{2,i}, S_i, S_{i-d_1}, S_{i-d_2}, U_i) + n\epsilon_n \end{aligned} \quad (41)$$

where:

(a) follows from (39);

(b) follows because M_1 and S^n are independent;

(c) follows because M_1 and M_2 are independent given S^n (first term) and since conditioning reduces entropy (second term);

(d) follows by adding and subtracting the term $H(M_1|V_1^\ell, V_2^\ell, S^n, M_2)$;

(e) follows because V_1^ℓ and V_2^ℓ are fully determined by (S^n, M_1, M_2) while V_2^ℓ is a deterministic function of (M_2, V_1^ℓ) (first term), and from the the mutual information chain rule (second term);

(f) follows since conditioning reduces entropy (first term), and because X_1^n is a deterministic function of (M_1, V_1^ℓ, S^n) while X_2^n is a deterministic function of (M_2, V_2^ℓ, S^n) (second and third terms);

(g) follows since conditioning reduces entropy (first and second terms) and because when conditioned on $(X_{1,i}, X_{2,i})$ and S_i , the channel output at time i is independent of $(V_1^\ell, V_2^\ell, M_1, M_2, S^{i-1}, S_{i+1}^n, X_{1,i-1}^n, X_{1,i+1}^n, X_{2,i-1}^n, X_{2,i+1}^n, Y^{i-1})$;

(h) follows from (12) and by defining $U_i = (V_1^\ell, V_2^\ell, S^{i-d_1-1})$.

Note that the auxiliary random variable U at time i was defined as $U_i \triangleq (V_1^\ell, V_2^\ell, S^{i-d_1-1})$. Accordingly, it represents the information shared during the conference (i.e., the parts of the private messages available to both encoders) and the common knowledge of the states. This is completely analogous to the role of U_i in the common message setting (see Theorem 1 and Section III-A).

Applying similar arguments to R_2 and $R_1 + R_2$, one can conclude that any achievable rate-pair (R_1, R_2) must satisfy the following inequalities:

$$R_1 \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, S_i, S_{i-d_2}, S_{i-d_1}, U_i) + C_{12} + \epsilon_n \quad (42)$$

$$R_2 \leq \frac{1}{n} \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, S_i, S_{i-d_2}, S_{i-d_1}, U_i) + C_{21} + \epsilon_n \quad (43)$$

$$R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_i, S_{i-d_2}, S_{i-d_1}, U_i) + C_{12} + C_{21} + \epsilon_n \quad (44)$$

$$R_1 + R_2 \leq \frac{1}{n} \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_i, S_{i-d_2}, S_{i-d_1}) + \epsilon_n. \quad (45)$$

The expressions on the right-hand side of the inequalities in (42)–(45) represent empirical averages of mutual information (taken over the code symbols). These inequalities can be alternatively represented by introducing a new time-sharing random variable Q , uniformly distributed over $\{1, \dots, n\}$, as in Subsection III-A. Starting again with the upper bound on R_1 , this yields

$$\begin{aligned} R_1 &\leq \frac{1}{n} \sum_{i=1}^n I(Y_Q; X_{1,Q} | X_{2,Q}, S_Q, S_{Q-d_2}, S_{Q-d_1}, U_Q, Q = i) \\ &\quad + C_{12} + \epsilon_n \\ &= I(Y_Q; X_{1,Q} | X_{2,Q}, S_Q, S_{Q-d_2}, S_{Q-d_1}, U_Q, Q) \\ &\quad + C_{12} + \epsilon_n \end{aligned} \quad (46)$$

Applying the same procedure to the rest of the upper bounds, while denoting $X_1 \triangleq X_{1,Q}$, $X_2 \triangleq X_{2,Q}$, $Y \triangleq Y_Q$, $S \triangleq S_Q$, $\tilde{S}_1 \triangleq S_{Q-d_1}$, $\tilde{S}_2 \triangleq S_{Q-d_2}$ and $U \triangleq (U_Q, Q)$, we get

$$R_1 \leq I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) + C_{12} + \epsilon_n \quad (47a)$$

$$R_2 \leq I(X_2; Y | X_1, U, S, \tilde{S}_1, \tilde{S}_2) + C_{21} + \epsilon_n \quad (47b)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | U, S, \tilde{S}_1, \tilde{S}_2) + C_{12} + C_{21} + \epsilon_n \quad (47c)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2) + \epsilon_n, \quad (47d)$$

where the justification for (47d) follows similar steps to those presented in (27). Moreover, note that the fact that the obtained region is convex follows from the same arguments given in Section III-A.

Completion of the proof of the converse relies on showing that the following Markov relations hold:

$$U - \tilde{S}_1 - (S, \tilde{S}_2) \quad (48a)$$

$$X_1 - (\tilde{S}_1, U) - (S, \tilde{S}_2) \quad (48b)$$

$$X_2 - (\tilde{S}_1, \tilde{S}_2, U) - (X_1, S), \quad (48c)$$

which is accomplished by applying the same line of arguments employed in Appendix A, while replacing M_0 with (V_1^ℓ, V_2^ℓ) .

B. Achievability

To prove the achievability of the capacity region, we need to show that for a fixed distribution of the form $P_{U|\tilde{S}_1} P_{X_1|\tilde{S}_1, U} P_{X_2|\tilde{S}_1, \tilde{S}_2, U}$ and for (R_1, R_2) that satisfy the inequalities in (37), there exists a sequence of $(n, \ell, 2^{nR_1}, 2^{nR_2}, d_1, d_2)$ codes for which $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. The idea behind this proof is to convert the conferencing problem into a setting that corresponds to the FSM-MAC with a common message considered in Section III, and then rely on Theorem 1 to show that the capacity region with conferencing is, indeed, achievable. This is accomplished by sharing as much of the original private messages (m_1, m_2) as possible through the conferencing links to construct a common message. The parts of the original messages not shared by the encoders constitute the private messages in the new setting. Next, the coding scheme introduced in Section III-B for the FSM-MAC with a common message can be employed.

We start by defining:

$$\tilde{R}_1 = \min\{R_1, C_{12}\} \quad (49a)$$

$$\tilde{R}_2 = \min\{R_2, C_{21}\}, \quad (49b)$$

and rewriting the inequalities in (37) as

$$(R_1 - \tilde{R}_1) \leq I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) \quad (50a)$$

$$(R_2 - \tilde{R}_2) \leq I(X_2; Y | X_1, U, S, \tilde{S}_1, \tilde{S}_2) \quad (50b)$$

$$(R_1 - \tilde{R}_1) + (R_2 - \tilde{R}_2) \leq I(X_1, X_2; Y | U, S, \tilde{S}_1, \tilde{S}_2) \quad (50c)$$

$$\begin{aligned} (\tilde{R}_1 + \tilde{R}_2) + (R_1 - \tilde{R}_1) + (R_2 - \tilde{R}_2) \\ \leq I(X_1, X_2; Y | S, \tilde{S}_1, \tilde{S}_2). \end{aligned} \quad (50d)$$

In view of this representation, we construct a coding scheme by splitting the sets $\mathcal{M}_j = \{1, 2, \dots, 2^{nR_j}\}$, for $j \in \{1, 2\}$, into $2^{n\tilde{R}_j}$ cells, each containing $2^{n(R_j - \tilde{R}_j)}$ messages, and introducing the functions

$$c_1 : \mathcal{M}_1 \rightarrow \{1, 2, \dots, 2^{n\tilde{R}_1}\} \quad (51a)$$

$$c_2 : \mathcal{M}_2 \rightarrow \{1, 2, \dots, 2^{n\tilde{R}_2}\} \quad (51b)$$

$$e_1 : \mathcal{M}_1 \rightarrow \{1, 2, \dots, 2^{n(R_1 - \tilde{R}_1)}\} \quad (51c)$$

$$e_2 : \mathcal{M}_2 \rightarrow \{1, 2, \dots, 2^{n(R_2 - \tilde{R}_2)}\}. \quad (51d)$$

Here, for every message m_j , where $j \in \{1, 2\}$, c_j returns its cell number, $c_j(m_j)$, while e_j returns its index number, $e_j(m_j)$, within the cell $c_j(m_j)$. For the sake of simplicity, we assume here that $2^{n\tilde{R}_1}$, $2^{n\tilde{R}_2}$, $2^{n(R_1 - \tilde{R}_1)}$ and $2^{n(R_2 - \tilde{R}_2)}$ are integers, although the same approach can be formalized for real numbers as well. Also note that the partitioning above is deterministic.

$$R_1 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1) \right) + C_{12}, \quad (55a)$$

$$R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) \right) + C_{21}, \quad (55b)$$

$$R_1 + R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) \right) + C_{12} + C_{21}, \quad (55c)$$

$$R_1 + R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 P_{2,i}(\tilde{s}_2, \tilde{s}_1) + 2g_{1,i}(s)g_{2,i}^*(s) \sqrt{(P_{1,i}(\tilde{s}_1) - \gamma_{1,i}(\tilde{s}_1))(P_{2,i}(\tilde{s}_2, \tilde{s}_1) - \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2))} \right), \quad (55d)$$

Now, for every message pair $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ we define the triplet (m'_0, m'_1, m'_2) where

$$m'_1 \triangleq e_1(m_1) \quad (52a)$$

$$m'_2 \triangleq e_2(m_2) \quad (52b)$$

$$m'_0 \triangleq (c_1(m_1), c_2(m_2)). \quad (52c)$$

Note that the above definitions dictate that $m'_1 \in \{1, 2, \dots, 2^{n(R_1-\tilde{R}_1)}\}$, $m'_2 \in \{1, 2, \dots, 2^{n(R_2-\tilde{R}_2)}\}$ and $m'_0 \in \{1, 2, \dots, 2^{n\tilde{R}_1}\} \times \{1, 2, \dots, 2^{n\tilde{R}_2}\}$. Since by definition $\tilde{R}_1 \leq C_{12}$ and $\tilde{R}_2 \leq C_{21}$, it is possible for Encoder 1 to transmit $c_1(m_1)$ to Encoder 2 and for Encoder 2 to transmit $c_2(m_2)$ to Encoder 1 via the respective conferencing links. Therefore, following the conferencing stage, both encoders know $m'_0 = (c_1(m_1), c_2(m_2))$. m'_1 and m'_2 are viewed as the new private messages.

The above setting can hence be viewed as a FSM-MAC with a common message. The messages to be transmitted are given by the triplet (m'_0, m'_1, m'_2) , where $m'_0 \in \{1, 2, \dots, 2^{n\tilde{R}_1}\} \times \{1, 2, \dots, 2^{n\tilde{R}_2}\}$, $m'_1 \in \{1, 2, \dots, 2^{n(R_1-\tilde{R}_1)}\}$ and $m'_2 \in \{1, 2, \dots, 2^{n(R_2-\tilde{R}_2)}\}$, while (50) holds by assumption. By Theorem 1, it now immediately follows that the new message triplet (m'_0, m'_1, m'_2) can be transmitted to the decoder with an arbitrarily small probability of error. The decoder can, therefore, reliably reconstruct the message pair (m_1, m_2) and the rate-region (37) is therefore achievable. ■

Remark 4: Using arguments similar to those presented in Remark 2, we obtain that when either (or both) of the delays is increased without bound, the capacity region in Theorem 2 approaches the corresponding region where the CSI at the appropriate encoder(s) is absent.

V. THE VECTOR GAUSSIAN FSM-MAC WITH DIAGONAL CHANNEL TRANSFER MATRICES, CONFERENCING AND DELAYED CSI

In this section we consider the vector Gaussian FSM-MAC with diagonal channel transfer matrices, partially cooperative encoders and delayed CSI. For every time instance

$t \in \{1, \dots, n\}$, the channel model under consideration is:

$$\mathbf{Y}_t = \mathbf{G}_1(s_t)\mathbf{X}_{1,t} + \mathbf{G}_2(s_t)\mathbf{X}_{2,t} + \mathbf{Z}_t, \quad (53)$$

where $\{\mathbf{G}_1(s)\}_{s \in \mathcal{S}}$ and $\{\mathbf{G}_2(s)\}_{s \in \mathcal{S}}$ are $N \times N$ diagonal matrices, which are deterministic functions of the channel state $S = s$. We denote the diagonal entries of these matrices by $g_{1,i}(s)$ and $g_{2,i}(s)$, respectively, for $i \in \{1, \dots, N\}$ and $s \in \mathcal{S}$. Moreover, we assume that $\{\mathbf{G}_1(s)\}_{s \in \mathcal{S}}, \{\mathbf{G}_2(s)\}_{s \in \mathcal{S}} \subset \mathbb{C}^{N \times N}$. For every $t \in \{1, 2, \dots, n\}$, $\mathbf{X}_{1,t}, \mathbf{X}_{2,t} \in \mathbb{C}^N$ and $\mathbf{Y}_t \in \mathbb{C}^N$ are the channel input vectors and the channel output vector, respectively. $\{\mathbf{Z}_t\}_{t=1}^n$ is a proper complex zero mean additive white Gaussian noise (AWGN) process, independent of $\mathbf{X}_{1,t}$ and $\mathbf{X}_{2,t}$ for every $t \in \{1, 2, \dots, n\}$. Thus, each noise sample is distributed according to $\mathbf{Z}_t \sim \mathcal{CN}(0, \mathbf{I})$, where \mathbf{I} is the identity matrix of dimensions $N \times N$. The input vector signals are assumed to satisfy the average power constraints

$$\text{tr}(\Sigma_{X_1 X_1}) \leq \bar{P}_1; \quad \text{tr}(\Sigma_{X_2 X_2}) \leq \bar{P}_2, \quad (54)$$

where we use the standard notation $\Sigma_{XY} = \mathbb{E}[\mathbf{X}\mathbf{Y}^\dagger]$, and \mathbf{A}^\dagger denotes the conjugate transpose of the matrix \mathbf{A} .

The motivation for examining the channel model in (53) stems from the fact that it can be used to represent an OFDM-based communication system, employing single receive and transmit antennas. OFDM is an efficient technique used to mitigate frequency selective fading, which is typical in modern wideband communication systems (see [1], [52]). The underlying idea behind OFDM is to split the channel's bandwidth into N separate sub-channels through which orthogonal signals are transmitted. By doing so, not only is the impact of intersymbol interference (ISI) dramatically reduced, but the transfer functions of each of the sub-channels boil down to multiplicative scalar gains. These gains are modeled by the diagonal entries of the channel matrices defined above. In this section we derive the maximization problem that specifies the capacity region for the vector Gaussian channel under consideration and convert it into a convex problem. The solution of this convex maximization problem, which can be easily obtained using a numerical tool such as CVX [50], also yields the optimal power allocation strategy among the sub-channels, which is another essential factor in an OFDM-based transmission.

A. Capacity Region

Theorem 3: The capacity region of the power-constrained vector Gaussian FSM-MAC with diagonal channel transfer matrices, partially cooperative encoders, cooperation link capacities C_{12} and C_{21} , delayed CSI and average power constraints (\bar{P}_1, \bar{P}_2) is the union of all sets of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying (55) at the top of the previous page, where the union is over all $\{\gamma_{1,i}(\tilde{s}_1)\}_{i \in \{1, \dots, N\}, \tilde{s}_1 \in \mathcal{S}}$, $\{\gamma_{2,i}(\tilde{s}_1, \tilde{s}_2)\}_{i \in \{1, \dots, N\}, (\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2}$, $\{P_{1,i}(\tilde{s}_1)\}_{i \in \{1, \dots, N\}, \tilde{s}_1 \in \mathcal{S}}$, $\{P_{2,i}(\tilde{s}_1, \tilde{s}_2)\}_{i \in \{1, \dots, N\}, (\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2} \subset \mathbb{R}$ that satisfy the constraints:

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{i=1}^N P_{1,i}(\tilde{s}_1) \leq \bar{P}_1 \quad (56a)$$

$$\sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_{i=1}^N P_{2,i}(\tilde{s}_1, \tilde{s}_2) \leq \bar{P}_2 \quad (56b)$$

$$0 \leq \gamma_{1,i}(\tilde{s}_1) \leq P_{1,i}(\tilde{s}_1), \quad \forall i \in \{1, \dots, N\}, \quad \tilde{s}_1 \in \mathcal{S} \quad (56c)$$

$$0 \leq \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) \leq P_{2,i}(\tilde{s}_1, \tilde{s}_2), \quad \forall i \in \{1, \dots, N\}, (\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2. \quad (56d)$$

The corresponding capacity region for the analogous setting with a common message can be obtained from Theorem 3 by taking:

$$\tilde{R}_0 = C_{12} + C_{21} \quad (57a)$$

$$\tilde{R}_1 = \max\{0, R_1 - C_{12}\} \quad (57b)$$

$$\tilde{R}_2 = \max\{0, R_2 - C_{21}\}, \quad (57c)$$

where \tilde{R}_0 denotes the common message rate, and \tilde{R}_1 and \tilde{R}_2 denote the rates of the private messages (according to the common message channel definition in Section II-A). The result is summarized in the following Corollary.

Corollary 1: The capacity region of the power-constrained vector Gaussian FSM-MAC with diagonal channel transfer matrices, a common message, delayed CSI and average power constraints (\bar{P}_1, \bar{P}_2) is the union of all sets of rate triplets $(R_0, R_1, R_2) \in \mathbb{R}_+^3$ satisfying (58) at the bottom of the page, where the union is over the domain satisfying the constraints (56).

Note that the capacity regions in Theorem 3 and Corollary 1 are both given in the form of a convex optimization problem, which can be solved efficiently using numerical tools. In the following proof we first derive a slightly different, yet equivalent, region for the Gaussian conferencing model. This equivalent capacity region involves a nonconvex optimization problem that we then convert into a convex problem by an appropriate change of optimization variables.

Proof: A straightforward extension of the result stated in Theorem 2 yields the capacity region of the *general* vector FSM-MAC with partially cooperative encoders, delayed CSI and power constraints as in (54). The region is given by the closure of the set of rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ that satisfy (see (37))

$$R_1 \leq I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{U}, \mathcal{S}, \tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2) + C_{12} \quad (59a)$$

$$R_2 \leq I(\mathbf{X}_2; \mathbf{Y}|\mathbf{X}_1, \mathbf{U}, \mathcal{S}, \tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2) + C_{21} \quad (59b)$$

$$R_1 + R_2 \leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|\mathbf{U}, \mathcal{S}, \tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2) + C_{12} + C_{21} \quad (59c)$$

$$R_1 + R_2 \leq I(\mathbf{X}_1, \mathbf{X}_2; \mathbf{Y}|\mathcal{S}, \tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2), \quad (59d)$$

for some joint distribution of the form

$$P_{\mathcal{S}\tilde{\mathcal{S}}_1\tilde{\mathcal{S}}_2} P_{\mathbf{U}|\tilde{\mathcal{S}}_1} P_{\mathbf{X}_1|\tilde{\mathcal{S}}_1, \mathbf{U}} P_{\mathbf{X}_2|\tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2, \mathbf{U}}, \quad (60)$$

where \mathbf{U} is an auxiliary random vector with bounded cardinality. The convexity of the capacity region in (59) follows from arguments of a nature similar to those presented in Section III-A, namely, by relying on a time-sharing random variable. Note that the structure of the PDF in (60) implies the Markov relations:

$$\mathbf{U} - \tilde{\mathcal{S}}_1 - (\mathcal{S}, \tilde{\mathcal{S}}_2) \quad (61a)$$

$$\mathbf{X}_1 - (\mathbf{U}, \tilde{\mathcal{S}}_1) - (\mathcal{S}, \tilde{\mathcal{S}}_2) \quad (61b)$$

$$\mathbf{X}_2 - (\mathbf{U}, \tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2) - (\mathcal{S}, \mathbf{X}_1). \quad (61c)$$

The proof of Theorem 3 consists of two main parts. First, we provide an outer bound for the general capacity region in (59). Then, by choosing a jointly proper complex Gaussian distribution for $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$, we show that the upper bound is indeed achievable and thus characterizes the actual capacity region.

$$R_1 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1) \right), \quad (58a)$$

$$R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) \right), \quad (58b)$$

$$R_1 + R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 \gamma_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) \right), \quad (58c)$$

$$R_0 + R_1 + R_2 \leq \sum_{\tilde{s}_1} \pi(\tilde{s}_1) \sum_{\tilde{s}_2} K^{d_1-d_2}(\tilde{s}_2, \tilde{s}_1) \sum_s K^{d_2}(s, \tilde{s}_2) \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + |g_{2,i}(s)|^2 P_{2,i}(\tilde{s}_2, \tilde{s}_1) + 2g_{1,i}(s)g_{2,i}^*(s) \sqrt{(P_{1,i}(\tilde{s}_1) - \gamma_{1,i}(\tilde{s}_1))(P_{2,i}(\tilde{s}_2, \tilde{s}_1) - \gamma_{2,i}(\tilde{s}_1, \tilde{s}_2))} \right), \quad (58d)$$

The outer bound for the capacity region is obtained by substituting the random vectors $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$ in (59) with appropriately chosen jointly proper complex Gaussian random vectors $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$, which satisfy a certain Markovian relation. We conclude that the chosen random vectors $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$ indeed admit the desired Markov relation using the following lemma [49, Sec. 2, Th. 1].

Lemma 5: Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ be jointly proper complex Gaussian random vectors. Then $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ form a Markov chain $\mathbf{A} - \mathbf{B} - \mathbf{C}$ if and only if their covariance matrices satisfy:

$$\Sigma_{AC} = \Sigma_{AB} \Sigma_{BB}^{-1} \Sigma_{BC}. \quad (62)$$

As before, we restrict the detailed derivation to the upper bound on R_1 , while noting that all other bounds in (59) can be straightforwardly treated in an analogous manner. To this end, we rewrite the bound on R_1 as (see (5))

$$R_1 \leq \sum_{\tilde{s}_1 \in \mathcal{S}} \pi(\tilde{s}_1) \sum_{\tilde{s}_2 \in \mathcal{S}} K^{d_1 - d_2}(\tilde{s}_2, \tilde{s}_1) \times \sum_{s \in \mathcal{S}} K^{d_2}(s, \tilde{s}_2) I(\mathbf{X}_1; \mathbf{Y} | \mathbf{X}_2, \mathbf{U}, s, \tilde{s}_1, \tilde{s}_2) + C_{12} \quad (63)$$

and proceed with upper bounding each of the mutual information terms in the sum. Consider:

$$\begin{aligned} & I(\mathbf{X}_1; \mathbf{Y} | \mathbf{X}_2, \mathbf{U}, s, \tilde{s}_1, \tilde{s}_2) \\ & \stackrel{(a)}{=} h(\mathbf{G}_1(s)\mathbf{X}_1 + \mathbf{Z} | \mathbf{U}, s, \tilde{s}_1) - h(\mathbf{Z}) \\ & \stackrel{(b)}{\leq} h(\mathbf{G}_1(s)\mathbf{X}_1 + \mathbf{Z} | \mathbf{V}, s, \tilde{s}_1) - h(\mathbf{Z}) \\ & \stackrel{(c)}{\leq} h(\mathbf{G}_1(s)\mathbf{X}_1^G + \mathbf{Z} | \mathbf{V}^G, s, \tilde{s}_1) - h(\mathbf{Z}) \\ & \leq \sum_{i=1}^N \left\{ h(g_{1,i}(s)X_{1,i}^G + Z_i, V_i^G | s, \tilde{s}_1) - h(V_i^G | \tilde{s}_1) \right\} - h(\mathbf{Z}) \\ & = \sum_{i=1}^N \left\{ \log \left((\pi e)^2 \left[1 + |g_{1,i}(s)|^2 \left(\mathbb{E}[|X_{1,i}^G|^2 | \tilde{s}_1] \right. \right. \right. \right. \\ & \quad \left. \left. \left. - \frac{|\mathbb{E}[X_{1,i}^G (V_i^G)^* | \tilde{s}_1]|^2}{\mathbb{E}[|V_i^G|^2 | \tilde{s}_1]} \right) \right] \cdot \mathbb{E}[|V_i^G|^2 | \tilde{s}_1] \right) \\ & \quad \left. - \log \left((\pi e) \mathbb{E}[|V_i^G|^2 | \tilde{s}_1] \right) - \log(\pi e) \right\} \\ & \stackrel{(d)}{=} \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) \right. \\ & \quad \left. \times \left[1 - \frac{|\mathbb{E}[X_{1,i}^G (V_i^G)^* | \tilde{s}_1]|^2}{P_{1,i}(\tilde{s}_1) \mathbb{E}[|V_i^G|^2 | \tilde{s}_1]} \right] \right) \\ & \stackrel{(e)}{=} \sum_{i=1}^N \log \left(1 + |g_{1,i}(s)|^2 \beta_{1,i}(\tilde{s}_1) P_{1,i}(\tilde{s}_1) \right) \end{aligned} \quad (64)$$

where:

(a) follows from (53) and the Markov relations (61);
 (b) follows by substituting the random vector \mathbf{U} , for any given $\tilde{S}_1 = \tilde{s}_1$, with a new random vector: $\mathbf{V}(\tilde{s}_1) \triangleq \mathbb{E}[\mathbf{X}_1 | \mathbf{U}, \tilde{s}_1]$. Note that this is the optimal estimator in the minimum mean square error (MMSE) sense of \mathbf{X}_1 given \mathbf{U} , for each specified

delayed CSI $\tilde{S}_1 = \tilde{s}_1$. By substituting \mathbf{U} (for some $\tilde{S}_1 = \tilde{s}_1$) with $\mathbf{V}(\tilde{s}_1)$ we increase the first entropy term in view of the fact that $\mathbf{V}(\tilde{s}_1)$ is a deterministic function of the pair $(\mathbf{U}, \tilde{s}_1)$, while $h(\mathbf{Z})$ is not affected by the substitution. Moreover, one can easily confirm that $(\mathbf{X}_1, \mathbf{V}, \mathbf{X}_2)$ satisfy the covariance condition (62), i.e., the relation

$$\Sigma_{X_1 X_2}(\tilde{s}_1, \tilde{s}_2) = \Sigma_{X_1 V}(\tilde{s}_1) \Sigma_V^{-1}(\tilde{s}_1) \Sigma_{V X_2}(\tilde{s}_1, \tilde{s}_2) \quad (66)$$

holds for every $(\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2$. Note that the dependence of the covariance matrices on the states is induced by the Markov relations (61);

(c) follows from the maximum differential entropy lemma [51, Sec. 2.2] and by introducing the triplet $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$ of zero-mean jointly proper complex Gaussian random vectors with the same auto- and cross- covariance matrices as those of $(\mathbf{X}_1, \mathbf{V}, \mathbf{X}_2)$. Replacing $(\mathbf{X}_1, \mathbf{V}, \mathbf{X}_2)$ with $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$ thus increases the first entropy term. Moreover, by Lemma V-A, we conclude that the Gaussian triplet $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$, for any given $(S, \tilde{S}_1, \tilde{S}_2) = (s, \tilde{s}_1, \tilde{s}_2)$, is Markov, i.e., the relation $\mathbf{X}_1^G(\tilde{s}_1) - \mathbf{V}^G(\tilde{s}_1) - \mathbf{X}_2^G(\tilde{s}_1, \tilde{s}_2)$ holds.

(d) follows from defining $P_{1,i}(\tilde{s}_1) \triangleq \mathbb{E}[|X_{1,i}|^2 | \tilde{s}_1]$ and $P_{2,i}(\tilde{s}_1, \tilde{s}_2) \triangleq \mathbb{E}[|X_{2,i}|^2 | \tilde{s}_1, \tilde{s}_2]$ (note that these are in fact the i -th diagonal entries of the covariance matrices $\Sigma_{X_1^G X_1^G}(\tilde{s}_1)$ and $\Sigma_{X_2^G X_2^G}(\tilde{s}_1, \tilde{s}_2)$, respectively. For this reason, the constraints in (56a)-(56b) follow immediately from (54) by applying the law of total expectation);

(e) follows from defining

$$\begin{aligned} \bar{\beta}_{1,i}(\tilde{s}_1) &= \left| \frac{\mathbb{E}[V_i^G (X_{1,i}^G)^* | \tilde{s}_1]}{\sqrt{\mathbb{E}[|X_{1,i}^G|^2 | \tilde{s}_1] \mathbb{E}[|V_i^G|^2 | \tilde{s}_1]}} \right|^2 \\ &= \frac{|\mathbb{E}[V_i^G (X_{1,i}^G)^* | \tilde{s}_1]|^2}{P_{1,i}(\tilde{s}_1) \mathbb{E}[|V_i^G|^2 | \tilde{s}_1]}, \end{aligned} \quad (67)$$

where we use the notation $\bar{\alpha} = 1 - \alpha$, $\alpha \in \mathbb{R}$. We also introduce the definition

$$\begin{aligned} \bar{\beta}_{2,i}(\tilde{s}_1, \tilde{s}_2) &= \left| \frac{\mathbb{E}[V_i^G (X_{2,i}^G)^* | \tilde{s}_1, \tilde{s}_2]}{\sqrt{\mathbb{E}[|X_{2,i}^G|^2 | \tilde{s}_1, \tilde{s}_2] \mathbb{E}[|V_i^G|^2 | \tilde{s}_1]}} \right|^2 \\ &= \frac{|\mathbb{E}[V_i^G (X_{2,i}^G)^* | \tilde{s}_1, \tilde{s}_2]|^2}{P_{2,i}(\tilde{s}_1, \tilde{s}_2) \mathbb{E}[|V_i^G|^2 | \tilde{s}_1]}, \end{aligned} \quad (68)$$

that will be used to represent the additional rate constraints in (55). Note that $\bar{\beta}_{1,i}(\tilde{s}_1)$ (respectively, $\bar{\beta}_{2,i}(\tilde{s}_1, \tilde{s}_2)$) is defined to be the squared correlation coefficient between $X_{1,i}^G$ (respectively, $X_{2,i}^G$) and V_i^G for a given delayed CSI $\tilde{S}_1 = \tilde{s}_1$ (respectively, delayed CSI pair $(\tilde{S}_1, \tilde{S}_2) = (\tilde{s}_1, \tilde{s}_2)$). Accordingly, we have that $\beta_{1,i}(\tilde{s}_1), \beta_{2,i}(\tilde{s}_1, \tilde{s}_2) \in [0, 1]$ for every $i \in \{1, \dots, N\}$. The upper bounds on R_2 , and both upper bounds on the sum-rate $R_1 + R_2$, are similarly constructed.

Next, we show that the upper bounds are also achievable. We take $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$ to be zero-mean jointly proper complex Gaussian random vectors that admit the Markov relations (61) and for which the auto- and cross- covariance matrices $\Sigma_{X_1 X_1}(\tilde{s}_1)$, $\Sigma_{X_2 X_2}(\tilde{s}_1, \tilde{s}_2)$, $\Sigma_{UU}(\tilde{s}_1)$, $\Sigma_{X_1 U}(\tilde{s}_1)$ and

$\Sigma_{X_2U}(\tilde{s}_1, \tilde{s}_2)$ are diagonal for every $(\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2$. Specifically, we take

$$\Sigma_{X_1X_1}(\tilde{s}_1) = \text{diag}\left(\{P_{1,i}(\tilde{s}_1)\}_{i=1}^N\right) \quad (69a)$$

$$\Sigma_{X_2X_2}(\tilde{s}_1, \tilde{s}_2) = \text{diag}\left(\{P_{2,i}(\tilde{s}_1, \tilde{s}_2)\}_{i=1}^N\right), \quad (69b)$$

and denote the diagonal entries of the covariance matrices $\Sigma_{UU}(\tilde{s}_1)$, $\Sigma_{X_1U}(\tilde{s}_1)$ and $\Sigma_{X_2U}(\tilde{s}_1, \tilde{s}_2)$ by $\sigma_{U_i}^2(\tilde{s}_1)$, $\mathbb{E}[X_{1,i}U_i^*|\tilde{s}_1]$ and $\mathbb{E}[X_{2,i}U_i^*|\tilde{s}_1, \tilde{s}_2]$, respectively, where $i \in \{1, 2, \dots, n\}$. Moreover, $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$ are chosen to have the same entry-wise correlations as $(\mathbf{X}_1^G, \mathbf{V}^G, \mathbf{X}_2^G)$, that is

$$\frac{|\mathbb{E}[U_i X_{1,i}^*|\tilde{s}_1]|^2}{P_{1,i}(\tilde{s}_1)\sigma_{U_i}^2(\tilde{s}_1)} = \bar{\beta}_{1,i}(\tilde{s}_1) \quad (70a)$$

$$\frac{|\mathbb{E}[U_i X_{2,i}^*|\tilde{s}_1, \tilde{s}_2]|^2}{P_{2,i}(\tilde{s}_1, \tilde{s}_2)\sigma_{U_i}^2(\tilde{s}_1)} = \bar{\beta}_{2,i}(\tilde{s}_1, \tilde{s}_2). \quad (70b)$$

It can now be shown that this choice for the random vectors $(\mathbf{X}_1, \mathbf{U}, \mathbf{X}_2)$ achieves the upper bounds (55). For conciseness, we present only the calculation for R_1 and note that the proof for the remaining rate bounds is similar. As in (64), using the channel model and the Markov relations (61), we have that:

$$\begin{aligned} I(\mathbf{X}_1; \mathbf{Y}|\mathbf{X}_2, \mathbf{U}, s, \tilde{s}_1, \tilde{s}_2) &= h(\mathbf{G}_1(s)\mathbf{X}_1 + \mathbf{Z}|\mathbf{U}, s, \tilde{s}_1) - h(\mathbf{Z}) \\ &= h(\mathbf{G}_1(s)\mathbf{X}_1 + \mathbf{Z}, \mathbf{U}|s, \tilde{s}_1) - h(\mathbf{U}|\tilde{s}_1) - h(\mathbf{Z}). \end{aligned} \quad (71)$$

Clearly

$$h(\mathbf{Z}) = \log\left((\pi e)^N\right) \quad (72)$$

$$h(\mathbf{U}|\tilde{s}_1) = \log\left((\pi e)^N \prod_{i=1}^N \sigma_{U_i}^2(\tilde{s}_1)\right). \quad (73)$$

Therefore, it is left to obtain an explicit expression for

$$h(\mathbf{G}_1(s)\mathbf{X}_1 + \mathbf{Z}, \mathbf{U}|s, \tilde{s}_1) = \log\left((\pi e)^{2N} \left|\tilde{\Sigma}(s, \tilde{s}_1)\right|\right), \quad (74)$$

where $\tilde{\Sigma}(s, \tilde{s}_1)$ is a block matrix of the structure

$$\tilde{\Sigma}(s, \tilde{s}_1) = \begin{pmatrix} \mathbf{I} + \mathbf{G}_1(s)\Sigma_{X_1X_1}(\tilde{s}_1)\mathbf{G}_1^\dagger(s) & \mathbf{G}_1(s)\Sigma_{X_1U}(\tilde{s}_1) \\ \Sigma_{X_1U}^\dagger(\tilde{s}_1)\mathbf{G}_1^\dagger(s) & \Sigma_{UU}(\tilde{s}_1) \end{pmatrix}. \quad (75)$$

After some algebra it can be shown that:

$$\begin{aligned} \left|\tilde{\Sigma}(s, \tilde{s}_1)\right| &= \left(\prod_{i=1}^N \left[|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + 1\right]\right) \\ &\quad \times \left(\prod_{i=1}^N \left[\sigma_{U_i}^2(\tilde{s}_1) - \frac{|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1)\bar{\beta}_{1,i}(\tilde{s}_1)\sigma_{U_i}^2(\tilde{s}_1)}{|g_{1,i}(s)|^2 P_{1,i}(\tilde{s}_1) + 1}\right]\right) \\ &= \left(\prod_{i=1}^N \sigma_{U_i}^2(\tilde{s}_1)\right) \cdot \left(\prod_{i=1}^N \left[|g_{1,i}(s)|^2 \beta_{1,i}(\tilde{s}_1) P_{1,i}(\tilde{s}_1) + 1\right]\right). \end{aligned} \quad (76)$$

Substituting (76) along with (72), (73) and (74) into (71) and summing the mutual information terms over all state triplets $(S, \tilde{S}_1, \tilde{S}_2) = (s, \tilde{s}_1, \tilde{s}_2)$ in (63), we achieve the upper bound

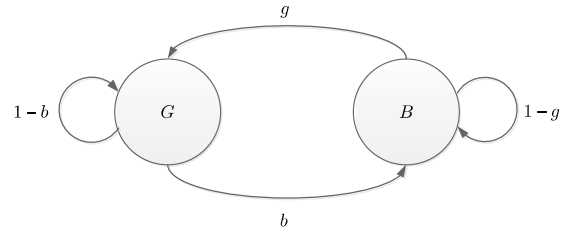


Fig. 4. Two-state AWGN channel.

for R_1 conforming with (65). In a similar manner, all other upper bounds can be shown to be achievable. This characterizes the maximization problem defining the capacity region for the diagonal vector Gaussian FSM-MAC with partially cooperative encoders and delayed CSI. Note that through this proof we have shown the optimality of the proper complex Gaussian multivariate input distribution for this model.

We note that the problem of maximizing the achievable rate region obtained using the above steps is not convex since the argument of the logarithms involve products of the optimization variables (e.g., the product $P_{1,i}(\tilde{s}_1)\bar{\beta}_{1,i}(\tilde{s}_1)$ in (76)). Thus, to convert it into a convex maximization problem we further substitute

$$\gamma_{1,i}(\tilde{s}_1) = \beta_{1,i}(\tilde{s}_1)P_{1,i}(\tilde{s}_1), \quad \forall \tilde{s}_1 \in \mathcal{S} \quad (77a)$$

$$\gamma_{2,i}(\tilde{s}_1, \tilde{s}_2) = \beta_{2,i}(\tilde{s}_1, \tilde{s}_2)P_{2,i}(\tilde{s}_1, \tilde{s}_2), \quad \forall (\tilde{s}_1, \tilde{s}_2) \in \mathcal{S}^2, \quad (77b)$$

for every $i \in \{1, \dots, N\}$. This substitution yields the rate bounds in (55) and concludes the proof. ■

B. Two-State Scalar AWGN Channel Example

To gain some intuition on the capacity region of the MAC with partially cooperative encoders and delayed CSI, we now consider the scalar Gaussian channel with only two possible states. The scalar channel corresponds to taking $N = 1$ in the diagonal vector channel definition in (53). We denote the two possible channel states by G and B (where G stands for ‘Good’ and B for ‘Bad’), thus, $\mathcal{S} = \{G, B\}$. The two states differ in their associated channel gains. When $S = G$, the gains are $g_1(s = G) = g_2(s = G) \triangleq g_G$, whereas when $S = B$ the gains are $g_1(s = B) = g_2(s = B) \triangleq g_B$. We assume without loss of generality that $g_G > g_B$. The Markov model of the state process is illustrated in Fig. 4.

The state process is specified by the the transition probability matrix:

$$K = \begin{pmatrix} P(G|G) & P(B|G) \\ P(G|B) & P(B|B) \end{pmatrix} = \begin{pmatrix} 1-b & b \\ g & 1-g \end{pmatrix}, \quad (78)$$

which induces the following stationary distribution:

$$\pi = (\pi(G) \quad \pi(B)) = \left(\frac{g}{g+b} \quad \frac{b}{g+b}\right). \quad (79)$$

We start by examining the impact of the cooperation link capacities, C_{12} and C_{21} , on the capacity regions in the particular case of symmetric CSI delays, i.e., $d_1 = d_2 \triangleq d$. Note that since $d_1 = d_2$, it immediately follows that $\tilde{S}_1 = \tilde{S}_2 \triangleq \tilde{S}$. The capacity region is presented in Fig. 5 for three different cases: (a) symmetrical capacities, represented by, $C_{12} = C_{21} \triangleq C$,

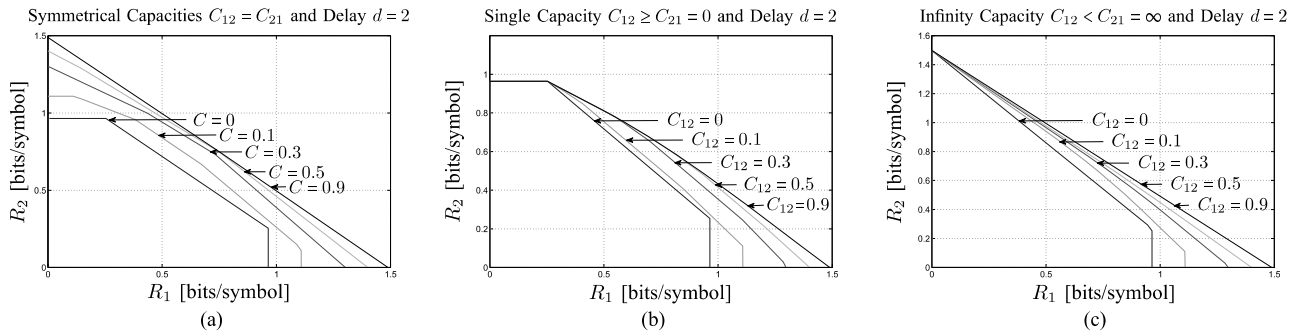


Fig. 5. Capacity region for the two-state AWGN-MAC: (a) symmetrical, $C_{12} = C_{21} \triangleq C$; (b) single cooperation link, $C_{12} \geq C_{21} = 0$; (c) infinite cooperation link, $C_{12} < C_{21} = \infty$.

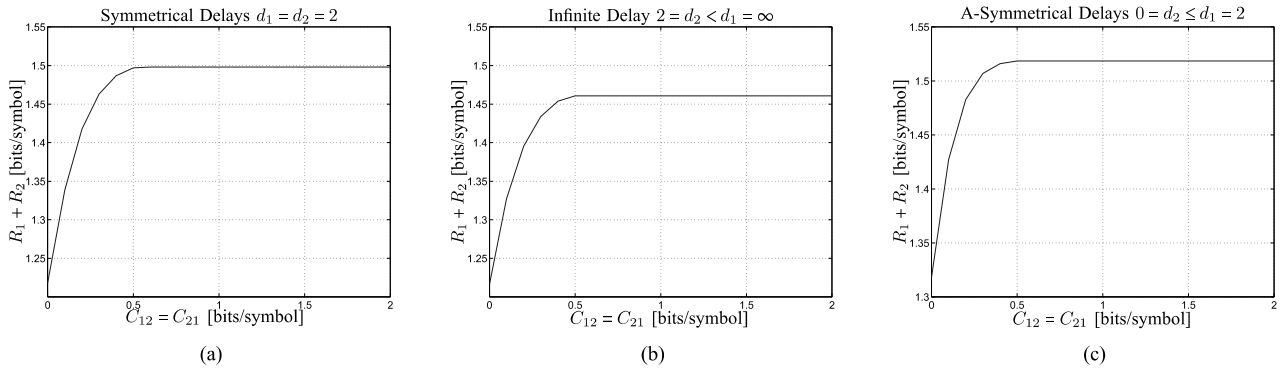


Fig. 6. The sum-rate versus the cooperation link capacities $C_{12} = C_{21}$ for three different cases of delayed CSI: (a) symmetrical delays, $d_2 = d_1 = 2$; (b) infinite delay, $2 = d_2 < d_1 = \infty$; (c) asymmetrical delays, $0 = d_2 \leq d_1 = 2$. The dashed line corresponds to the case where $C_{12} = C_{21} = \infty$.

(b) single cooperation link, represented by, $C_{12} \geq C_{21} = 0$ and (c) one infinite cooperation link, represented by, $C_{12} < C_{21} = \infty$. The capacity regions were calculated by numerically solving the optimization problem induced by Theorem 3 for the above three cases using CVX [50]. Throughout this example we assume $\bar{P}_1 = \bar{P}_2 = 10$, $g_B = 0.01$, $g_G = 1$, $g = b = 0.1$ and $d = 2$ (results of similar nature were observed for $g_B = 0.2$ and $g_B = 0.3$).

Note that in Fig. 5(a), which presents the region for the symmetrical case, as C grows without bound, the capacity region increases and eventually adopts a triangular shape. This outcome is because the first three constraints on the rates (R_1, R_2) , as given by (55a)–(55c), also grow without bound, and thus, the binding constraint is the sum-rate constraint of (55d). For the case of a single cooperation link shown in Fig. 5(b), the upper bound on R_2 remains fixed as C_{12} grows, since the constraint in (55b) does not change with C_{12} and stays fixed at approximately 0.9642. Finally, for the case of infinite cooperation link capacity $C_{21} = \infty$, as shown in Fig. 5(c), we have that the constraint on R_2 in (55b) and the first constraint on the sum-rate in (55c) are both redundant. Hence, the only meaningful constraint on R_2 is (55d), which does not involve C_{21} (or C_{12}).

Next, we demonstrate that the capacity region of this setting grows as the cooperation link capacities grow, regardless of the specific assumptions on the relation between the delays of the CSI available at the encoders. To do so, we present the maximum sum-rate versus the cooperation link capacities for

three different possible relations between the delays: (a) $d_1 = d_2 = 2$, (b) $2 = d_2 < d_1 = \infty$ and (c) $2 = d_1 \geq d_2 = 0$. For all three cases we assume $C_{12} = C_{21}$ and use the same values of the channel gains as before. The curves are shown in Fig. 6(a)–(c).

As expected, The sum-rate of case (c) (which exhibits the best CSI properties of the three) reaches the highest value as the capacities grow, whereas the sum-rate for case (b) (which exhibits the worst CSI properties) reaches the lowest value. Moreover, we note the correspondence between Fig. 6(a) and Fig. 5(a) (both corresponding to the case of symmetrical delays and equal cooperation link capacities). Evidence of this correspondence is the fact that when $C_{12} = C_{21}$ grow, the sum-rate, in both figures, approaches its maximal value, which is approximately 1.5 bits per symbol.

Another interesting aspect of the Gaussian channel example is the impact of the signal-to-noise ratio (SNR) on the correlations between the auxiliary random variable, U , and the random variables X_1 and X_2 . These correlations are associated with the level of cooperation used in the scheme. We assume that the transmit powers satisfy $P_1 = P_2 \triangleq P$ and that $g_1 = g_2 = 1$, so that the SNR, in fact, equals P , and restrict the analysis to the case where $|\mathcal{S}| = 1$, i.e., a single and constant channel state [46]. We use throughout the same notations and expressions for the rate bounds as in [46]. Note that for the case where $|\mathcal{S}| = 1$, the maximization problem in (55) turns out to be concave even without the transformation (77); thus no transformation is needed. The remaining optimization

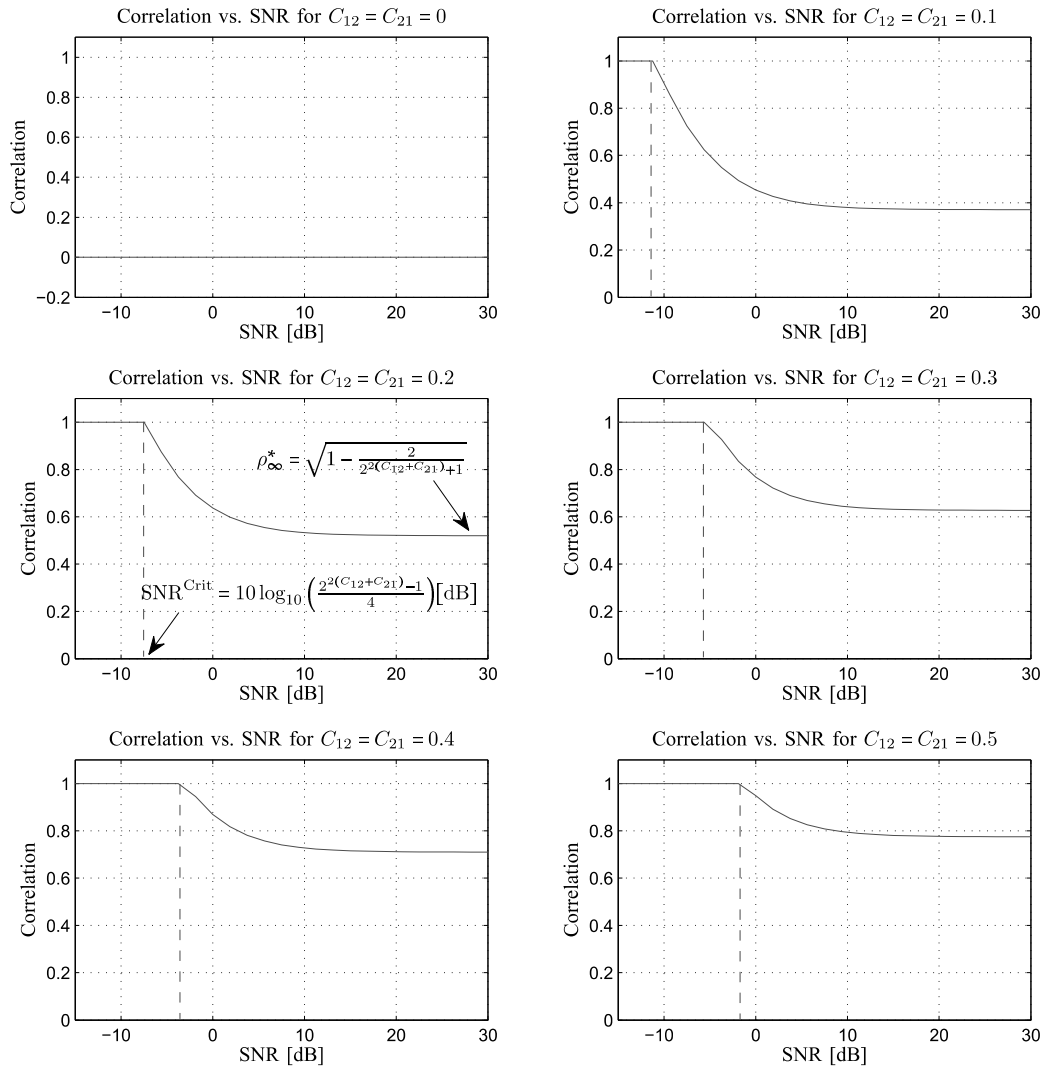


Fig. 7. Correlation as a function of SNR for different values of the capacities $C_{12} = C_{21}$.

variables are β_1 and β_2 , which are defined through (see (67)–(68))

$$\sqrt{1 - \beta_1} = \left| \frac{\mathbb{E}[UX_1]}{\sqrt{\mathbb{E}[X_1^2]\mathbb{E}[U^2]}} \right| \triangleq \rho_1 \quad (80a)$$

$$\sqrt{1 - \beta_2} = \left| \frac{\mathbb{E}[UX_2]}{\sqrt{\mathbb{E}[X_2^2]\mathbb{E}[U^2]}} \right| \triangleq \rho_2. \quad (80b)$$

We consider the case of symmetrical cooperation link capacities, i.e., $C_{12} = C_{21}$. By the symmetry of the maximization problem in (β_1, β_2) , optimality is achieved when $\beta_1 = \beta_2$. For this reason we use the notation $\beta_1 = \beta_2 \triangleq \beta$ and plot a single curve representing both correlations (which are calculated directly from β according to (80)). The numerical results are shown in Fig. 7. The dashed blue and green lines designate the asymptotic value of the correlation and the critical SNR at which the correlation drops from unity, respectively. Results are shown for six different values of $C_{12} = C_{21}$.

Although the effect of the SNR on the correlations could not be calculated analytically, we use asymptotic evaluations to gain some additional insight. Namely, we demonstrate that the optimal correlation admits

$$\rho^* = \begin{cases} 1, & \text{SNR} \leq \text{SNR}^{\text{Crit}} \\ \sqrt{1 - \frac{2}{2^{2(C_{12}+C_{21})}+1}}}, & \text{SNR} \rightarrow \infty, \end{cases} \quad (81)$$

where $\text{SNR}^{\text{Crit}} = 10 \log_{10} \left(\frac{2^{2(C_{12}+C_{21})}-1}{4} \right)$ [dB].

We start by justifying the observation that the correlation approaches 1 for small SNR values. For some positive value of $C_{12} = C_{21}$ and for $P_1, P_2 \ll 1$, consider (see (55c)–(55d)):

$$\begin{aligned} R_1 + R_2 & \leq \min \left\{ \frac{1}{2} \log \left(1 + \beta(g_1^2 P_1 + g_2^2 P_2) \right) + C_{12} + C_{21}, \right. \\ & \left. \frac{1}{2} \log \left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\beta^2 P_1 P_2} \right) \right\} \\ & = \frac{1}{2} \log \left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\beta^2 P_1 P_2} \right) \end{aligned} \quad (82)$$

Now note that the last term in (82) is maximized for $\beta^* = 0$, which, in turn, implies that the correlation is equal to unity. As shown in Fig. 7, for smaller values of SNR the correlation is indeed higher, indicating that the scheme compensates for the low SNR via cooperation.

The asymptotic evaluation for low SNRs is valid up to some critical SNR value at which the correlation drops from its maximal value of unity. We define this critical value of SNR as

$$\text{SNR}^{\text{Crit}} = \sup \{P \mid \rho^*(P) = 1\}. \quad (83)$$

To calculate SNR^{Crit} we restrict the analysis to the segment of SNRs at which the correlation is maximal (or equivalently, $\beta^* = 0$) and consider (82) taken for $P_1 = P_2 = P$ and $g_1 = g_2 = 1$. As shown in (82), when $\beta^* = 0$ and $P \rightarrow 0$, the second logarithm achieves the minimum between the two terms. Fixing $\beta^* = 0$ and increasing P increases the second logarithm in (82) while the first term remains unchanged and equals $C_{12} + C_{21}$. As long as

$$\frac{1}{2} \log \left(1 + 2P + 2\bar{\beta}P \right) \Big|_{\beta=0} < C_{12} + C_{21}, \quad (84)$$

the optimum is achieved for $\beta^* = 0$. However, when (84) is no longer valid, the optimal value of β must vary from 0. Thus, calculating SNR^{Crit} reduces to solving the following equation:

$$\frac{1}{2} \log \left(1 + 2P + 2\bar{\beta}P \right) \Big|_{\beta=0} = C_{12} + C_{21}, \quad (85)$$

yielding,

$$\text{SNR}^{\text{Crit}} = \frac{2^{2(C_{12}+C_{21})} - 1}{4}. \quad (86)$$

The value of SNR^{Crit} [dB] is represented by the perpendicular dashed green line in the plots shown in Fig. 7 and is observed to agree with the numerical results. Note that as the capacities $C_{12} = C_{21}$ grow, so does the value of SNR^{Crit} , and hence, the transition between the low- and high-SNR regimes occurs at higher SNR values.

As the SNR grows, the correlation asymptotically approaches some value in the interval $(0, 1)$; this value is denoted by ρ_∞^* . To find this asymptotic correlation, we present the following analysis for the high-SNR regime (assuming $P_1, P_2 \gg 1$). We start by excluding $\beta^* = 0$ as a possible solution for this case (a fact which will be used subsequently). Fixing $C_{12} = C_{21}$ and substituting $\beta = 0$ into the sum-rate bounds on $R_1 + R_2$ yields (see (55c)–(55d)):

$$R_1 + R_2 \leq \min \left\{ \begin{array}{l} C_{12} + C_{21}, \\ \frac{1}{2} \log \left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{P_1 P_2} \right) \end{array} \right\} \stackrel{(a)}{=} C_{12} + C_{21} \quad (87)$$

where (a) follows from the fact that $P_1, P_2 \gg 1$. We thus get that for $\text{SNR} \rightarrow \infty$, by taking $\beta = 0$, the sum-rate is bounded by the sum of the cooperation link capacities. However, since $C_{12} + C_{21}$ is a constant that does not depend on the powers P_1 and P_2 , we conclude that β^* cannot be equal to zero.

Next, assuming $\beta^* > 0$, we calculate β^* by using some approximations that are easily justified at a high SNR. First,

note that the first and second logarithms in (82) are monotonically increasing and decreasing, respectively, in β . This implies that the optimum is achieved at the value of β at which the functions intersect, that is

$$\begin{aligned} & \frac{1}{2} \log \left(1 + \beta(g_1^2 P_1 + g_2^2 P_2) \right) + C_{12} + C_{21} \\ &= \frac{1}{2} \log \left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2} \right). \end{aligned} \quad (88)$$

Using the fact that for high SNR we have:

$$\begin{aligned} & \frac{1}{2} \log \left(1 + \beta(g_1^2 P_1 + g_2^2 P_2) \right) + C_{12} + C_{21} \\ & \approx \frac{1}{2} \log \left(\beta(g_1^2 P_1 + g_2^2 P_2) \right) + C_{12} + C_{21} \\ & \frac{1}{2} \log \left(1 + g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2} \right) \\ & \approx \frac{1}{2} \log \left(g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2} \right), \end{aligned}$$

the equation in (88) reduces to:

$$\begin{aligned} & \beta(g_1^2 P_1 + g_1^2 P_2) 2^{2(C_{12}+C_{21})} \\ &= g_1^2 P_1 + g_2^2 P_2 + 2g_1 g_2 \sqrt{\bar{\beta}^2 P_1 P_2}. \end{aligned}$$

To further simplify the analysis we again assume a unit channel gain, that is, $g_1 = g_2 = 1$. After some algebra we obtain that the intersection point is given by

$$\beta^* = \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{2^{2(C_{12}+C_{21})}(P_1 + P_2) + 2\sqrt{P_1 P_2}}, \quad (89)$$

which by taking $P_1 = P_2 = P$, reduces to

$$\beta^* = \frac{2}{2^{2(C_{12}+C_{21})} + 1}. \quad (90)$$

Therefore, the optimal correlation at infinite SNR is given by

$$\rho_\infty^* = \sqrt{1 - \beta^*} = \sqrt{1 - \frac{2}{2^{2(C_{12}+C_{21})} + 1}}. \quad (91)$$

The value of ρ_∞^* , for each value of the cooperation link capacities C_{12} and C_{21} , is represented by the horizontal dashed blue line in the plots shown in Fig. 7. Note that the numerical calculations indeed meet the asymptotic results for large values of SNR.

To conclude, we interpret the numerical and analytical results in terms of the optimal transmission strategies of the users for each SNR regime. Recall that the symbols of the codewords transmitted by the users are modeled by the random variables X_1 and X_2 . The fact that for low SNR the correlation is at its maximal value of unity implies that both users tend to transmit the same codewords, which, in turn, indicates that they transmit the same message. However, the only common information the users share is the common message that they have created using the conference. Therefore, we conclude that when the channel quality is low, the best strategy for the users is to transmit the common message exclusively and to abandon their private messages (i.e., the parts of their original messages that they have not managed to share). As the SNR grows beyond SNR^{Crit} , the correlation between the code symbols decreases to some positive value $\rho^* \in (0, 1)$, asymptotically

approaching (91). This decrease in correlation is the result, when a higher quality channel is experienced, of each user transmitting not only the common (correlated) message, but also the private (uncorrelated) message.

One can also get some additional insight by examining the behavior of the correlation coefficient from the rate perspective. As long as the sum-rate falls below the sum of the cooperation link capacities, i.e., $R_1 + R_2 \leq C_{12} + C_{21}$, the transmission consists only of the correlated common message; namely, the users are fully cooperative. However, once the sum-rate crosses this threshold value, the transmitted codewords incorporate both the common and private messages, leading to a decrease of the optimal correlation coefficient.

VI. SUMMARY AND CONCLUDING REMARKS

In this paper we considered the FSM-MAC with partially cooperative encoders and delayed CSI, and derived its capacity region. The achievability proof used another result of this paper, namely, the capacity region of the FSM-MAC with a common message and delayed CSI. The latter result was obtained by providing a coding scheme that relies on strategy letters. Nonetheless, using the fact that the decoder has access to full CSI, it was also shown that optimal codes can be constructed directly over the input alphabet. Thus, a single codebook was constructed, a fact that formed the basis for simultaneous joint decoding. This approach not only successfully avoids the unnecessary complexity of a coding scheme based on rate-splitting and multiplexing (in contrast to previous works involving delayed CSI [38], [43]), but it also circumvents the need to rely on the corner points of the capacity region, which can render the analysis cumbersome and inefficient when the number of corner points is large.

The general conferencing result was then applied to the special case of the Gaussian vector MAC with diagonal channel transfer matrices, which models OFDM-based communication systems. The corresponding capacity region was given in the form of a convex optimization problem and the optimality of Gaussian Markovian inputs was established. This result serves as a generalization of [46] to the vector state-dependant case. Focusing on a two-state Gaussian FSM-MAC example, the crucial role of cooperation for low SNR values was demonstrated.

We finally note that an extension of the results to a more general state-dependant MAC with partially cooperative encoders and CSI at both transmitters and at the receiver (see [40]) is currently being investigated. Extensions of the results for the Gaussian vector FSM-MAC to general MIMO settings (see [47]) and to the ISI channel are also being considered.

APPENDIX A

PROOF OF THE MARKOV RELATION IN (28)

We prove the Markov relation (28) using the following claims. The Markov property in (28a) follows from the fact that $(M_0, S^{q-d_1-1}) - S_{q-d_1} - S_{q-d_2} - S_q, \forall q \in \{1, \dots, n\}$ and $d_1 > d_2$, and thus, due to the stationary property of the state process, also $(M_0, S_1^{Q-d_1-1}, Q) - S_{Q-d_1} - S_{Q-d_2} - S_Q$.

To show (28b) consider the following relations

$$\begin{aligned}
 & P(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) \\
 &= P(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &= \sum_{m_1 \in \mathcal{M}_1} P(m_1, x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &= \sum_{m_1 \in \mathcal{M}_1} P(m_1|s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &\quad \times P(x_{1,q}|s_q, s_{q-d_1}, s_{q-d_2}, m_0, m_1, s^{q-d_1-1}, q) \\
 &\stackrel{(a)}{=} \sum_{m_1 \in \mathcal{M}_1} P(m_1|s_{q-d_1}, m_0, s^{q-d_1-1}, q) \\
 &\quad \times P(x_{1,q}|s_{q-d_1}, m_0, m_1, s^{q-d_1-1}, q) \\
 &= \sum_{m_1 \in \mathcal{M}_1} P(m_1, x_{1,q}|s_{q-d_1}, m_0, s^{q-d_1-1}, q) \\
 &= P(x_{1,q}|s_{q-d_1}, m_0, s^{q-d_1-1}, q), \tag{92}
 \end{aligned}$$

where (a) follows from the facts that M_1 is independent of S^n given M_0 and $X_{1,q}$ is a deterministic function of $(M_0, M_1, S_{q-d_1}, S^{q-d_1-1})$. Now, since this is true for all $q \in \{1, \dots, n\}$ and because the auxiliary random variable is defined as $U = (M_0, S^{Q-d_1-1}, Q)$, we conclude that

$$P(x_1|s, \tilde{s}_1, \tilde{s}_2, u) = P(x_1|\tilde{s}_1, u). \tag{93}$$

Finally, to show (28c) we use the following relations

$$\begin{aligned}
 & P(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, u_q, q) \\
 &= P(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &= \sum_{m_2 \in \mathcal{M}_2} P(m_2, x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &= \sum_{m_2 \in \mathcal{M}_2} P(m_2|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &\quad \times P(x_{2,q}|x_{1,q}, s_q, s_{q-d_1}, s_{q-d_2}, m_0, m_2, s^{q-d_1-1}, q) \\
 &\stackrel{(a)}{=} \sum_{m_2 \in \mathcal{M}_2} P(m_2|s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &\quad \times P(x_{2,q}|s_{q-d_1}, s_{q-d_2}, m_0, m_2, s^{q-d_1-1}, q) \\
 &= \sum_{m_2 \in \mathcal{M}_2} P(m_2, x_{2,q}|s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q) \\
 &= P(x_{2,q}|s_{q-d_1}, s_{q-d_2}, m_0, s^{q-d_1-1}, q), \tag{94}
 \end{aligned}$$

where (a) follows from the facts that M_2 is independent of $(X_{1,q}, S^n)$ given M_0 and $X_{2,i}$ is independent of $(X_{1,q}, S_q)$ given $(M_0, M_2, S_{q-d_1}, S_{q-d_2}, S^{q-d_1-1})$. Again, the above holds for every $q \in \{1, \dots, n\}$, and by the definition of the random variable U , we conclude that

$$P(x_2|x_1, s, \tilde{s}_1, \tilde{s}_2, u) = P(x_2|\tilde{s}_1, \tilde{s}_2, u).$$

APPENDIX B

ERROR PROBABILITY ANALYSIS FOR THE ACHIEVABILITY PROOF OF THEOREM 1

We need to show that for the coding scheme presented in Section III-B and for a rate triplet (R_0, R_1, R_2) as given in Theorem 1, $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$. Define the event in (95) at the bottom of the next page for any $\mathbf{s} \in S^n$ (recall that a fixed

state sequence \mathbf{s} induces a fixed pair of delayed state sequences $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$). Denote the transmitted messages by (m_0, m_1, m_2) . Using (95), the probability of error, when averaged over the ensemble of codebooks, can be written as in (96) at the bottom of the page. By the union bound, (96) is further upper bounded by (97). We proceed with the following steps:

- 1) $P_e^{[1]} \rightarrow 0$ as $n \rightarrow \infty$ by the law of large numbers.
- 2) To upper bound $P_e^{[5]}$ consider the following:

$$\begin{aligned} P_e^{[5]} &\stackrel{(a)}{\leq} \sum_{(\tilde{m}_0, \tilde{m}_1, \tilde{m}_2) \neq (m_0, m_1, m_2)} 2^{-n(I(U, X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon)} \\ &\leq 2^{n(R_0 + R_1 + R_2)} 2^{-n(I(U, X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon)} \\ &= 2^{n(R_0 + R_1 + R_2 - I(U, X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2) + \delta_\epsilon)} \end{aligned} \quad (98)$$

where step (a) is proven in App. C, and $\delta_\epsilon \rightarrow 0$ as $n \rightarrow \infty$. Hence, for the probability $P_e^{[5]}$ to vanish as $n \rightarrow \infty$, the following must hold:

$$R_0 + R_1 + R_2 < I(U, X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2). \quad (99)$$

The mutual information term in (99) can be rewritten as,

$$\begin{aligned} &I(U, X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2) \\ &\stackrel{(a)}{=} I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2) + I(U; Y|X_1, X_2, S, \tilde{S}_1, \tilde{S}_2) \\ &\stackrel{(b)}{=} I(X_1, X_2; Y|S, \tilde{S}_1, \tilde{S}_2), \end{aligned} \quad (100)$$

where (a) follows from the mutual information chain rule and (b) follows from the fact that Y is independent of U given (X_1, X_2, S) , by the underlying channel model (see Section III-B).

- 3) The upper bounds on $P_e^{[2]}$, $P_e^{[3]}$ and $P_e^{[4]}$ are all observed to be redundant, since in all three types of events the codeword \mathbf{U} is assumed incorrect, which immediately

implies that the codewords \mathbf{X}_1 and \mathbf{X}_2 are also incorrect. Hence, requiring the probability of error to vanish as $n \rightarrow \infty$ produces the same upper bound as in (99) but with respect to the partial sum-rates R_0 , $R_0 + R_1$ and $R_0 + R_2$. It can therefore be concluded that the upper bound in (99) is the dominating constraint.

- 4) To upper bound $P_e^{[6]}$ consider the following steps:

$$\begin{aligned} P_e^{[6]} &\stackrel{(a)}{\leq} \sum_{\tilde{m}_1 \neq m_1} 2^{-n(I(X_1; Y|X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon)} \\ &\leq 2^{nR_1} 2^{-n(I(X_1; Y|X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon)} \\ &= 2^{n(R_1 - I(X_1; Y|X_2, U, S, \tilde{S}_1, \tilde{S}_2) + \delta_\epsilon)} \end{aligned} \quad (101)$$

where the proof of step (a) is provided in App. C, and $\delta_\epsilon \rightarrow 0$ as $n \rightarrow \infty$. It hence follows that $P_e^{[6]} \rightarrow 0$ as $n \rightarrow \infty$ as long as,

$$R_1 < I(X_1; Y|X_2, U, S, \tilde{S}_1, \tilde{S}_2). \quad (102)$$

- 5) Using similar arguments it can be shown that to guarantee that $P_e^{[7]}$ and $P_e^{[8]}$ vanish as $n \rightarrow \infty$ the following conditions must hold,

$$R_2 < I(X_2; Y|X_1, U, S, \tilde{S}_1, \tilde{S}_2) \quad (103)$$

$$R_1 + R_2 < I(X_1, X_2; Y|U, S, \tilde{S}_1, \tilde{S}_2). \quad (104)$$

Summarizing the above results, we get that the right-hand side of (97) goes to zero as the blocklength $n \rightarrow \infty$ if the rate bounds in (15) are satisfied.

APPENDIX C

PROOF OF (98) AND (101) IN APPENDIX B

For simplicity, the codewords associated with a message triplet $(m_0, m_1, m_2) \in \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2$ are denoted by

$$\mathcal{E}(i, j, k, \mathbf{s}) = \left\{ \left(\mathbf{U}(i, \tilde{\mathbf{s}}_1), \mathbf{X}_1(j, \mathbf{U}(i, \tilde{\mathbf{s}}_1), \tilde{\mathbf{s}}_1), \mathbf{X}_2(k, \mathbf{U}(i, \tilde{\mathbf{s}}_1), \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{Y}) \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y) \right) \right\} \quad (95)$$

$$\begin{aligned} P_e^{(n)} &= \mathbb{P} \left[\mathcal{E}^C(m_0, m_1, m_2, \mathbf{s}) \cup \left\{ \bigcup_{\tilde{m}_0 \neq m_0} \mathcal{E}(\tilde{m}_0, m_1, m_2, \mathbf{s}) \right\} \cup \left\{ \bigcup_{(\tilde{m}_0, \tilde{m}_1) \neq (m_0, m_1)} \mathcal{E}(\tilde{m}_0, \tilde{m}_1, m_2, \mathbf{s}) \right\} \right. \\ &\quad \cup \left\{ \bigcup_{(\tilde{m}_0, \tilde{m}_2) \neq (m_0, m_2)} \mathcal{E}(\tilde{m}_0, m_1, \tilde{m}_2, \mathbf{s}) \right\} \cup \left\{ \bigcup_{(\tilde{m}_0, \tilde{m}_1, \tilde{m}_2) \neq (m_0, m_1, m_2)} \mathcal{E}(\tilde{m}_0, \tilde{m}_1, \tilde{m}_2, \mathbf{s}) \right\} \\ &\quad \left. \cup \left\{ \bigcup_{\tilde{m}_1 \neq m_1} \mathcal{E}(m_0, \tilde{m}_1, m_2, \mathbf{s}) \right\} \cup \left\{ \bigcup_{\tilde{m}_2 \neq m_2} \mathcal{E}(m_0, m_1, \tilde{m}_2, \mathbf{s}) \right\} \cup \left\{ \bigcup_{(\tilde{m}_1, \tilde{m}_2) \neq (m_1, m_2)} \mathcal{E}(m_0, m_1, \tilde{m}_2, \mathbf{s}) \right\} \right] \quad (96) \\ &\leq \underbrace{\mathbb{P}[\mathcal{E}^C(m_0, m_1, m_2, \mathbf{s})]}_{P_e^{[1]}} + \underbrace{\sum_{\tilde{m}_0 \neq m_0} \mathbb{P}[\mathcal{E}(\tilde{m}_0, m_1, m_2, \mathbf{s})]}_{P_e^{[2]}} + \underbrace{\sum_{(\tilde{m}_0, \tilde{m}_1) \neq (m_0, m_1)} \mathbb{P}[\mathcal{E}(\tilde{m}_0, \tilde{m}_1, m_2, \mathbf{s})]}_{P_e^{[3]}} \\ &\quad + \underbrace{\sum_{(\tilde{m}_0, \tilde{m}_2) \neq (m_0, m_2)} \mathbb{P}[\mathcal{E}(\tilde{m}_0, m_1, \tilde{m}_2, \mathbf{s})]}_{P_e^{[4]}} + \underbrace{\sum_{(\tilde{m}_0, \tilde{m}_1, \tilde{m}_2) \neq (m_0, m_1, m_2)} \mathbb{P}[\mathcal{E}(\tilde{m}_0, \tilde{m}_1, \tilde{m}_2, \mathbf{s})]}_{P_e^{[5]}} \\ &\quad + \underbrace{\sum_{\tilde{m}_1 \neq m_1} \mathbb{P}[\mathcal{E}(m_0, \tilde{m}_1, m_2, \mathbf{s})]}_{P_e^{[6]}} + \underbrace{\sum_{\tilde{m}_2 \neq m_2} \mathbb{P}[\mathcal{E}(m_0, m_1, \tilde{m}_2, \mathbf{s})]}_{P_e^{[7]}} + \underbrace{\sum_{(\tilde{m}_1, \tilde{m}_2) \neq (m_1, m_2)} \mathbb{P}[\mathcal{E}(m_0, m_1, \tilde{m}_2, \mathbf{s})]}_{P_e^{[8]}} \end{aligned} \quad (97)$$

$(\mathbf{u}(m_0), \mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))$, thus omitting the functional dependence of the codewords on the delayed state sequences $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$. Moreover, when referring to the conditional ϵ -strongly typical set $\mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, Y | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$, we sometimes use the shortened notation \mathcal{T} , to save space.

A. Proof of Step (a) in (98)

Consider the equalities,

$$\begin{aligned} & \mathbb{P}\left[\left(\mathbf{U}(\tilde{m}_0), \mathbf{X}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{X}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{Y}\right) \right. \\ & \quad \left. \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y)\right] \\ &= \sum_{(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{y}) \in \mathcal{T}} P(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{y} | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &= \sum_{(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{y}) \in \mathcal{T}} P(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2). \quad (105) \end{aligned}$$

For the second term in the right-hand side (RHS) of the last equality in (105) we have (106) at the bottom of the page. Step (a) in (106) follows since given $(\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2)$, $(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))$ were drawn independently of $(\mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2))$, and (b) follows because the channel output is independent of the incorrect inputs given the correct inputs and states.

Substituting (106) into (105) we get

$$\begin{aligned} & \mathbb{P}\left[\left(\mathbf{U}(\tilde{m}_0), \mathbf{X}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{X}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{Y}\right) \right. \\ & \quad \left. \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y)\right] \\ &= \sum_{(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{y}) \in \mathcal{T}} P(\mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &\leq |\mathcal{T}| \cdot 2^{-n(H(U, X_1, X_2 | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(1)})} \cdot 2^{-n(H(Y | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(2)})} \\ &\leq 2^{n(H(U, X_1, X_2, Y | S, \tilde{S}_1, \tilde{S}_2) + \delta_\epsilon^{(3)})} \cdot 2^{-n(H(U, X_1, X_2 | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(1)})} \end{aligned}$$

$$\begin{aligned} & \times 2^{-n(H(Y | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(2)})} \\ &= 2^{-n(H(U, X_1, X_2, Y | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon)} \end{aligned}$$

where $\delta_\epsilon = \sum_{i=1}^3 \delta_\epsilon^{(i)}$. This completes the proof of step (a) in (98).

B. Proof of Step (a) in (101)

Analogous to the previous subsection we first write:

$$\begin{aligned} & \mathbb{P}\left[\left(\mathbf{U}(m_0), \mathbf{X}_1(m_0, \tilde{m}_1), \mathbf{X}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{Y}\right) \right. \\ & \quad \left. \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y)\right] \\ &= \sum_{(\mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2), \mathbf{y}) \in \mathcal{T}} P(\mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2). \quad (107) \end{aligned}$$

By similar arguments to those used to obtain (106), the second term in the RHS of (107) can be shown to satisfy

$$\begin{aligned} & P(\mathbf{y} | \mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &= P(\mathbf{y} | \mathbf{u}(m_0), \mathbf{x}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \quad (108) \end{aligned}$$

Substituting (108) into (107) yields

$$\begin{aligned} & \mathbb{P}\left[\left(\mathbf{U}(m_0), \mathbf{X}_1(m_0, \tilde{m}_1), \mathbf{X}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2, \mathbf{Y}\right) \right. \\ & \quad \left. \in \mathcal{T}_\epsilon^{(n)}(U, X_1, X_2, S, \tilde{S}_1, \tilde{S}_2, Y)\right] \\ &= \sum_{(\mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2), \mathbf{y}) \in \mathcal{T}} P(\mathbf{u}(m_0), \mathbf{x}_1(m_0, \tilde{m}_1), \mathbf{x}_2(m_0, m_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{u}(m_0), \mathbf{x}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &\leq |\mathcal{T}| \cdot 2^{-n(H(U, X_1, X_2 | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(1)})} \\ & \quad \times 2^{-n(H(Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(2)})} \\ &\leq 2^{n(H(U, X_1, X_2, Y | S, \tilde{S}_1, \tilde{S}_2) + \delta_\epsilon^{(3)})} \\ & \quad \times 2^{-n(H(U, X_1, X_2 | S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(1)})} \\ & \quad \times 2^{-n(H(Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon^{(2)})} \end{aligned}$$

$$\begin{aligned} & P(\mathbf{y} | \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &= \sum_{(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))} P(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2), \mathbf{y} | \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &= \sum_{(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))} P(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2) | \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2), \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &\stackrel{(a)}{=} \sum_{(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))} P(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ & \quad \times P(\mathbf{y} | \mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2), \mathbf{u}(\tilde{m}_0), \mathbf{x}_1(\tilde{m}_0, \tilde{m}_1), \mathbf{x}_2(\tilde{m}_0, \tilde{m}_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &\stackrel{(b)}{=} \sum_{(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2))} P(\mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2) | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) P(\mathbf{y} | \mathbf{x}_1(m_0, m_1), \mathbf{x}_2(m_0, m_2), \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \\ &= P(\mathbf{y} | \mathbf{s}, \tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2) \quad (106) \end{aligned}$$

$$= 2^{-n} \left(I(X_1; Y | X_2, U, S, \tilde{S}_1, \tilde{S}_2) - \delta_\epsilon \right)$$

where $\delta_\epsilon = \sum_{i=1}^3 \delta_\epsilon^{(i)}$. This establishes step (a) in (101).

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and the anonymous reviewers for their careful reading of the paper and their helpful comments. Especially, they would like to thank the anonymous Reviewer 1 for his suggestion to simplify the achievability proof of Theorem 1, a suggestion which led to the proof presented in the current version of the paper.

REFERENCES

- [1] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*, vol. 3. New York, NY, USA: Wiley, 2009.
- [2] I. Maric, R. D. Yates, and G. Kramer, "Capacity of interference channels with partial transmitter cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3536–3548, Oct. 2007.
- [3] O. Simeone *et al.*, "Cooperative wireless cellular systems: An information-theoretic view," *Found. Trends Commun. Inf. Theory*, vol. 8, nos. 1–2, pp. 1–177, 2011.
- [4] F. M. J. Willems, "The discrete memoryless multiple access channel with partially cooperating encoders," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 441–445, May 1983.
- [5] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, Sep. 1973, pp. 23–52.
- [6] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. Hawaii, Honolulu, HI, USA, 1972.
- [7] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 51, no. 7, pp. 1037–1076, 1973.
- [8] F. M. J. Willems and E. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.
- [9] F. M. J. Willems, "Informationtheoretical results for the discrete memoryless multiple access channel," Ph.D. dissertation, Dept. Math., Katholieke Univ. Leuven, Leuven, Belgium, Oct. 1982.
- [10] O. Simeone, D. Gündüz, H. V. Poor, A. J. Goldsmith, and S. Shamai (Shitz), "Compound multiple-access channels with partial cooperation," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2425–2441, Jun. 2009.
- [11] V. M. Prabhakaran and P. Viswanath, "Interference channels with destination cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 187–209, Jan. 2011.
- [12] V. M. Prabhakaran and P. Viswanath, "Interference channels with source cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 156–186, Jan. 2011.
- [13] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited receiver cooperation: Symmetric case," in *Proc. IEEE Inf. Theory Workshop*, Taormina, Italy, Oct. 2009, pp. 579–583.
- [14] H. Bagheri, A. S. Motahari, and A. K. Khandani, "On the symmetric Gaussian interference channel with partial unidirectional cooperation." [Online]. Available: <http://arxiv.org/pdf/0909.2777v1.pdf>
- [15] C. T. K. Ng, N. Jindal, A. J. Goldsmith, and U. Mitra, "Capacity gain from two-transmitter and two-receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3822–3827, Oct. 2007.
- [16] I. Maric, A. Goldsmith, G. Kramer, and S. Shamai (Shitz), "On the capacity of interference channels with one cooperating transmitter," *Eur. Trans. Telecommun.*, vol. 19, no. 4, pp. 405–420, 2008.
- [17] R. Dabora and S. D. Servetto, "On the role of estimate-and-forward with time sharing in cooperative communication," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4409–4431, Oct. 2008.
- [18] D. Gündüz and E. Erkip, "Source and channel coding for cooperative relaying," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3454–3475, Oct. 2007.
- [19] L. Sankar, G. Kramer, and N. B. Mandayam, "Dedicated-relay vs. user cooperation in time-duplexed multiaccess networks," *J. Commun.*, vol. 6, no. 4, pp. 330–339, Jul. 2011.
- [20] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Local base station cooperation via finite-capacity links for the uplink of linear cellular networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 190–204, Jan. 2009.
- [21] G. Kramer, I. Marić, and R. D. Yates, "Cooperative communications," *Found. Trends Netw.*, vol. 1, nos. 3–4, pp. 271–425, Oct. 2006.
- [22] A. Haghi, R. Khosravi-Farsani, M. R. Aref, and F. Marvasti, "The capacity region of p-transmitter/q-receiver multiple-access channels with common information," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7359–7376, Nov. 2011.
- [23] A. Haghi, R. Khosravi-Farsani, M. R. Aref, and F. Marvasti, "The capacity region of fading multiple access channels with cooperative encoders and partial CSIT," in *Proc. Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 485–489.
- [24] H. H. Permuter, S. Shamai (Shitz), and A. Somekh-Baruch, "Message and state cooperation in multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6379–6396, Oct. 2011.
- [25] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.
- [26] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.
- [27] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems Control Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [28] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 731–739, Sep. 1983.
- [29] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [30] A. Rosenzweig, Y. Steinberg, and S. Shamai, "On channels with partial channel state information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1817–1830, May 2005.
- [31] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5474, Dec. 2006.
- [32] M. Salehi, "Capacity and coding for memories with real-time noisy defect information at encoder and decoder," *IEE Proc. I Commun., Speech Vis.*, vol. 139, no. 2, pp. 113–117, Apr. 1992.
- [33] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3992–4003, Nov. 2005.
- [34] G. Como and S. Yüksel, "On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1267–1273, Mar. 2011.
- [35] A. Lapidoth and Y. Steinberg, "The multiple-access channel with causal side information: Common state," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 32–50, Jan. 2013.
- [36] M. Li, O. Simeone, and A. Yener, "Leveraging strictly causal state information at the encoders for multiple access channels," in *Proc. Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul./Aug. 2011, pp. 2806–2810.
- [37] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 6, pp. 445–586, Jun. 2007.
- [38] H. Viswanathan, "Capacity of Markov channels with receiver CSI and delayed feedback," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 761–771, Mar. 1999.
- [39] S. Verdú and T. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.
- [40] A. Das and P. Narayan, "Capacities of time-varying multiple-access channels with side information," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 4–25, Jan. 2002.
- [41] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [42] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
- [43] U. Basher, A. Shirazi, and H. H. Permuter, "Capacity region of finite state multiple-access channels with delayed state information at the transmitters," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3430–3452, Jun. 2012.
- [44] H. H. Permuter and O. Simeone, "Source coding when the side information may be delayed," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3607–3618, Jun. 2013.
- [45] V. Venkatesan, "Optimality of Gaussian inputs for a multi-access achievable rate-region," Semester Thesis, Signal Inf. Process. Lab., ETH Zurich Univ., Zürich, Switzerland, Jun. 2007.

- [46] S. I. Bross, A. Lapidoth, and M. A. Wigger, "The Gaussian MAC with conferencing encoders," in *Proc. Int. Symp. Inf. Theory*, Jul. 2008, pp. 2702–2706.
- [47] M. A. Wigger and G. Kramer, "Three-user MIMO MACs with cooperation," in *Proc. IEEE Inf. Theory Workshop*, Jun. 2009, pp. 221–225.
- [48] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
- [49] T. Ando and D. Petz, "Gaussian Markov triplets approached by block matrices," *Acta Sci. Math. (Szeged)*, vol. 75, nos. 1–2, pp. 329–345, 2009.
- [50] M. Grant and S. Boyd. (Sep. 2012). *CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 Beta*. [Online]. Available: <http://cvxr.com/cvx>
- [51] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [52] L. Nuaymi, *WiMAX: Technology for Broadband Wireless Access*. New York, NY, USA: Wiley, 2007.

Ziv Goldfeld (S'13) received his B.Sc. (summa cum laude) degree in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 2012. He is currently a student in the direct Ph.D. program for honor students in Electrical and Computer Engineering at that same institution.

Between 2003 and 2006, he served in the intelligence corps of the Israeli Defense Forces.

Ziv is a recipient of the Dean's List Award, the Basor Fellowship for honor students in the direct Ph.D. program, the Lev-Zion fellowship and the Minerva Short-Term Research Grant (MRG).

Haim H. Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial boards of the *IEEE TRANSACTIONS ON INFORMATION THEORY*.

Benjamin M. Zaidel (S'94–M'07) received the B.Sc. and M.Sc. degrees from Tel Aviv University, Israel, in 1990 and 1996, respectively, and the Ph.D. degree from the Technion-Israel Institute of Technology, Haifa, Israel, in 2006, all in electrical engineering.

During 1990–1997, he worked with a communications research group responsible for conducting feasibility studies of communication systems, and in particular cellular systems and other mobile communications networks. During 2001–2007 and 2008–2012, he has been with the Government Research Laboratories in the capacity of a senior research engineer. During the years 2007–2008, he held a Postdoctoral position in the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. Since 2014 he is with the Faculty of Engineering at Bar-Ilan University, Ramat-Gan, Israel.

His research interests include information-theoretic aspects of multiuser detection techniques, multi-input multi-output channels, cooperative processing in wireless networks, and the application of random matrix theory and statistical physics tools to problems in communications and information theory.