

Capacity Region of the Finite-State Multiple-Access Channel With and Without Feedback

Haim H. Permuter, *Member, IEEE*, Tsachy Weissman, *Senior Member, IEEE*, and Jun Chen, *Member, IEEE*

Abstract—The capacity region of the finite-state multiple-access channel (FS-MAC) with feedback that may be an arbitrary time-invariant function of the channel output samples is considered. We characterize both an inner and an outer bound for this region, using Massey’s directed information. These bounds are shown to coincide, and hence yield the capacity region, of indecomposable FS-MACs without feedback and of stationary and indecomposable FS-MACs with feedback, where the state process is not affected by the inputs. Though “multiletter” in general, our results yield explicit conclusions when applied to specific scenarios of interest. For example, our results allow us to do the following.

- Identify a large class of FS-MACs, that includes the additive mod2 noise MAC where the noise may have memory, for which feedback does not enlarge the capacity region.
- Deduce that, for a general FS-MAC with states that are not affected by the input, if the capacity (region) without feedback is zero, then so is the capacity (region) with feedback.
- Deduce that the capacity region of a MAC that can be decomposed into a “multiplexer” concatenated by a point-to-point channel (with, without, or with partial feedback), the capacity region is given by $\sum_m R_m \leq C$, where C is the capacity of the point to point channel and m indexes the encoders. Moreover, we show that for this family of channels source–channel coding separation holds.

Index Terms—Capacity region, causal conditioning, code-tree, directed information, feedback capacity, multiple-access channel (MAC), source–channel coding separation, sup-additivity of sets.

I. INTRODUCTION

THE multiple-access channel (MAC) has received much attention in the literature. To put our contributions in context, we begin by briefly describing some of the key results in the area. The capacity region for the memoryless MAC was derived by Ahlswede in [1]. Cover and Leung derived an achievable region for a memoryless MAC with feedback in [2]. Using block Markov encoding, superposition, and list codes, they showed that the region $R_1 \leq I(X_1; Y|X_2, U)$, $R_2 \leq I(X_2; Y|X_1, U)$, and $R_1 + R_2 \leq I(X_1, X_2; Y)$ where

Manuscript received August 02, 2007; revised August 23, 2008. Current version published May 20, 2009. This work was supported by the National Science Foundation (NSF) through the CAREER award and under Grant TFR-0729119. The material in this paper was presented in part at IEEE International Symposium on Information (ISIT), Toronto, ON, Canada, July 2008.

H. H. Permuter is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel (e-mail: haimp@bgu.ac.il).

T. Weissman is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel, on leave from Stanford University, Stanford, CA, 94305 USA (e-mail: tsachy@stanford.edu).

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 1A1, Canada (junchen@ece.mcmaster.ca).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2009.2018346

$P(u, x_1, x_2, y) = p(u)p(x_1|u)p(x_2|u)p(y|x_1, x_2)$ is achievable for a memoryless MAC with feedback. In [3], Willems showed that the achievable region given by Cover and Leung for a memoryless channel with feedback is optimal for a class of channels where one of the inputs is a deterministic function of the output and the other input. More recently, Bross and Lapidot [4] improved Cover and Leung’s region, and Wu *et al.* [5] extended Cover and Leung’s region for the case where noncausal state information is available at both encoders.

Ozarow derived the capacity of a memoryless Gaussian MAC with feedback in [6] and showed it to be achievable via a modification of the Schalkwijk–Kailath scheme [7]. In general, the capacity in the presence of noisy feedback is an open question for the point-to-point channel and *a fortiori* for the MAC. Lapidot and Wigger [8] presented an achievable region for the case of the Gaussian MAC with noisy feedback and showed that it converges to Ozarow’s noiseless-feedback sum-rate capacity as the feedback–noise variance tends to zero. Other recent variations on the Schalkwijk–Kailath scheme of relevance to the themes of our work include the case of quantization noise in the feedback link [9] and the case of interference known noncausally at the transmitter [10].

Verdú characterized the capacity region of a MAC of the form

$$P(y_i|x_1^i, x_2^i, y^{i-1}) = P(y_i|x_{1,i-m}^i, x_{2,i-m}^i)$$

without feedback in [11]. Verdú further showed in that work that in the absence of frame synchronism between the two users, i.e., there is a random shift between the users, only stationary input distributions need be considered. Cheng and Verdú built on the capacity result from [11] in [12] to show that for a Gaussian MAC there exists a water-filling solution that generalizes the point-to-point Gaussian channel.

In [13], [14], Kramer derived general capacity results for discrete memoryless networks with feedback. In particular, by using the idea of code-trees instead of codewords, Kramer derived a “multiletter” expression for the capacity of the discrete memoryless MAC. One of the main results we develop in the present paper extends Kramer’s capacity result to the case of a stationary and indecomposable finite-state MAC (FS-MAC) without ISI (intersymbol interference), to be formally defined below.

In [15], [16], Han used the information-spectrum method to derive the capacity of a general MAC without feedback, when the channel transition probabilities are arbitrary for every n symbols. Han also considered the additive mod- q MAC, which we shall use here to illustrate the way in which our general results characterize special cases of interest. In particular, our results will imply that feedback does not increase the capacity region of the additive mod- q MAC.

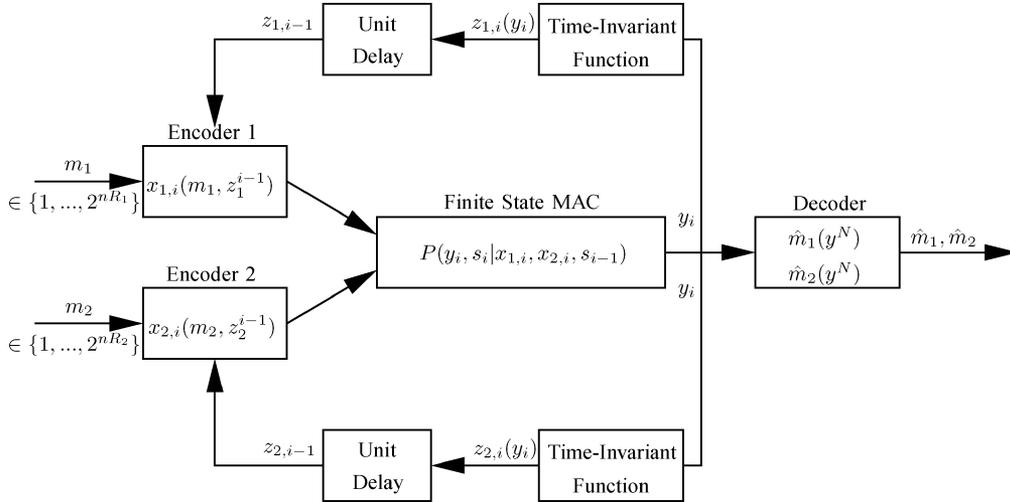


Fig. 1. Channel with feedback that is a time-invariant deterministic function of the output.

In this work, we consider the capacity region of the FS-MAC, with feedback that may be an arbitrary time-invariant function of the channel output samples. We characterize both an inner and an outer bound for this region. We show that these bounds coincide, and hence yield the capacity region, for the important subfamily of FS-MACs with states that evolve independently of the channel inputs. If feedback is not allowed the bounds coincide simply if the FS-MAC is indecomposable. Our derivation of the capacity region is rooted in the derivation of the capacity of finite-state channels in Gallager's book [17, Ch. 4,5]. More recently, Lapidoth and Telatar [18] have used it in order to derive the capacity of a compound channel without feedback, where the compound channel consists of a family of finite-state channels. In particular, they have introduced into Gallager's proof the idea of concatenating codewords, which we extend here to concatenating code trees.

Though "multiletter" in general, our results yield explicit conclusions when applied to more specific families of MACs. For example, we find that feedback does not increase the capacity of the mod- q additive noise MAC (where q is the size of the common alphabet of the input, output and noise), regardless of the memory in the noise. This result is in sharp contrast with the finding of Gaarder and Wolf in [19] that feedback can increase the capacity even of a *memoryless* MAC due to cooperation between senders that it can create. Our result should also be considered in light of Alajaji's work [20], which showed that feedback does not increase the capacity of discrete point-to-point channels with mod- q additive noise. Thus, this part of our contribution can be considered a multiterminal extension of Alajaji's result. Our results will in fact allow us to identify a class of MACs larger than that of the mod- q additive noise MAC for which feedback does not enlarge the capacity region.

Further specialization of the results will allow us to deduce that, for a general FS-MAC with states that are not affected by the input, if the capacity (region) without feedback is zero, then so is the capacity (region) with feedback. It will also allow us to identify a large class of FS-MACs for which source-channel coding separation holds.

The remainder of this paper is organized as follows. We concretely describe our channel model and assumptions in Section II. In Section III, we introduce some notation, tools, and results pertaining to directed information and convergence of sub/sup-additive regions that will be key in later sections. We state our main results in Section IV. In Section V, we apply the general results of Section IV to obtain the capacity region for several interesting classes of channels, as well as establish a source-channel separation result. The validity of our inner and outer bounds is established, respectively, in Sections VI and VII. In Section VIII, we show that our inner and outer bounds coincide, and hence yield the capacity region, when applied to the FS-MAC without feedback. This result can be thought of as a natural extension of Gallager's derivations [17] of the memoryless MAC without feedback, and Kramer's derivation [14] of memoryless MAC with feedback, to channels with states. In Section IX, we characterize the capacity region for the case of arbitrary (time-invariant) feedback and FS-MAC channels with states that evolve independently of the input, as well as the FS-MAC with limited ISI (which is the natural MAC analogue of Kim's point-to-point channel [21]), by showing that our inner and outer bounds coincide for this case. We conclude in Section X with a summary of our contribution and a related future research direction.

II. CHANNEL MODEL

In this paper, we consider a FS-MAC with a time-invariant feedback as illustrated in Fig. 1.

The MAC setting consists of two senders and one receiver. Each sender $l \in \{1, 2\}$ chooses an index m_l uniformly from the set $\{1, \dots, 2^{nR_l}\}$ and independently of the other sender. The input to the channel from encoder l is denoted by $\{X_{l1}, X_{l2}, X_{l3}, \dots\}$, and the output of the channel is denoted by $\{Y_1, Y_2, Y_3, \dots\}$. The state at time i , i.e., $S_i \in \mathcal{S}$, takes values in a finite set of possible states. The channel is homogeneous (does not change over time) and is characterized by a conditional probability $P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1})$ that satisfies

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

where the superscripts denote sequences in the following way: $x_l^i = (x_{l1}, x_{l2}, \dots, x_{li})$, $l \in \{1, 2\}$. Later in the paper, we also discuss more specific families of FS-MAC, such as indecomposable FS-MAC and FS-MAC without ISI. The definition of indecomposable FS-MAC is a straightforward extension of indecomposable FSC [17] where the effect of the initial state vanishes with time, for any given input sequence x_1^n, x_2^n . More precisely:

Definition 1: An FS-MAC is *indecomposable* if, for every $\epsilon > 0$, there exists an n_0 such that for $n \geq n_0$

$$|P(s_n|x_1^n, x_2^n, s_0) - P(s_n|x_1^n, x_2^n, s'_0)| \leq \epsilon \quad (2)$$

for all s_n, x_1^n, x_2^n, s_0 , and s'_0 .

If feedback is allowed, then the initial state s_0 may influence the future state s_n through influencing the input. Because of that reason, we also define an FS-MAC without ISI, for which the state process is not affected by the input.

Definition 2: An FS-MAC without ISI is an FS-MAC that satisfies

$$P(y_i, s_i|x_{1i}, x_{2i}, s_{i-1}) = P(s_i|s_{i-1})P(y_i|s_{i-1}, x_{1i}, x_{2i}). \quad (3)$$

Gallager also provides in [17, Theorem 4.6.3] a necessary and sufficient condition for verifying that the channel is indecomposable. For FS-MAC without ISI, the condition is simply an existence of a state s_n such that

$$P(s_n|s_0) > 0, \quad \forall s_0 \in \mathcal{S} \quad (4)$$

for some fixed $n \leq 2^{|\mathcal{S}|^2}$. This condition can be verified in a finite time, and it also implies [22, Theorem 6.3.2] that there exists a unique steady-state distribution (stationary distribution), i.e.,

$$\lim_{N \rightarrow \infty} \Pr(S_N = s|s_0) = \pi(s), \quad \forall s_0 \in \mathcal{S} \quad (5)$$

where $\pi(s)$ is the stationary distribution. Combining the definitions above we have

Definition 3: A stationary and indecomposable FS-MAC without ISI is an FS-MAC that satisfies

$$P(y_i, s_i|x_{1i}, x_{2i}, s_{i-1}) = P(s_i|s_{i-1})P(y_i|s_{i-1}, x_{1i}, x_{2i}) \quad (6)$$

and

$$P(s_0) = \pi(s_0), \quad (7)$$

where $\pi(s_0)$ is the unique stationary distribution that satisfies (5).

Throughout the paper, we assume a communication scheme with feedback z_l^i where the element z_{li} is a time-invariant function of the output y_i . For example, z_{li} could equal y_i (perfect feedback), or a quantized version of y_i , or null (no feedback). The encoders receive the feedback samples with one unit delay.

A code with feedback consists of two encoding functions $g_l : \{1, \dots, 2^{nR_1}\} \times \mathcal{Z}_l^{n-1} \rightarrow \mathcal{X}_l^n$, $l = 1, 2$, where the k th coordinate of $x_l^n \in \mathcal{X}_l^n$ is given by the function

$$x_{lk} = g_{lk}(m_l, z_l^{k-1}), \quad k = 1, 2, \dots, n, \quad l = 1, 2 \quad (8)$$

and a decoding function

$$g : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{nR_1}\} \times \{1, \dots, 2^{nR_2}\}. \quad (9)$$

The average probability of error for a $(2^{nR_1}, 2^{nR_2}, n)$ code is defined as

$$P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1, w_2} \Pr\{g(Y^n) \neq (w_1, w_2) | (w_1, w_2) \text{ sent}\}. \quad (10)$$

A rate (R_1, R_2) is said to be *achievable* for the MAC if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes with $P_e^{(n)} \rightarrow 0$. The *capacity region* of MAC, \mathcal{R} , is the closure of the set of achievable (R_1, R_2) rates.

III. PRELIMINARIES

In this section, we introduce two concepts that have a major role in characterizing the capacity of FS-MAC with feedback. The first concept is the directed information and causal conditioning idea. The second concept is the sub- and sup-additivity (a.k.a superadditive) property of sequences of regions.

A. Directed Information and Causal Conditioning

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

$$P(y^N || x^{N-d}) \triangleq \prod_{i=1}^N P(y_i | y^{i-1}, x^{i-d}) \quad (11)$$

(if $i-d \leq 0$ then x^{i-d} is set to null). In particular, we extensively use the cases where $d = 0, 1$

$$P(y^N || x^N) \triangleq \prod_{i=1}^N P(y_i | y^{i-1}, x^i) \quad (12)$$

$$Q(x^N || y^{N-1}) \triangleq \prod_{i=1}^N Q(x_i | x^{i-1}, y^{i-1}) \quad (13)$$

where the letters Q and P are both used for denoting pmfs.

Directed information $I(X^N \rightarrow Y^N)$ was defined by Massey in [23] as

$$I(X^N \rightarrow Y^N) \triangleq \sum_{i=1}^N I(X^i; Y_i | Y^{i-1}). \quad (14)$$

It has been widely used in the characterization of capacity of point-to-point channels [21], [24]–[29], compound channels [30], network capacity [13], [14], rate distortion [31]–[33], and computational biology [34], [35]. Directed information can also be expressed in terms of causal conditioning as

$$\begin{aligned} I(X^N \rightarrow Y^N) &= \sum_{i=1}^N I(X^i; Y_i | Y^{i-1}) \\ &= \mathbf{E} \left[\log \frac{P(Y^N || X^N)}{P(Y^N)} \right] \end{aligned} \quad (15)$$

where \mathbf{E} denotes expectation. The directed information from X^N to Y^N , conditioned on S , is denoted as $I(X^N \rightarrow Y^N|S)$ and is defined as

$$I(X^N \rightarrow Y^N|S) \triangleq \sum_{i=1}^N I(X^i; Y_i|Y^{i-1}, S). \quad (16)$$

Directed information between X_1^N and Y^N causally conditioned on X_2^N is defined as

$$\begin{aligned} I(X_1^N \rightarrow Y^N|X_2^N) &\triangleq \sum_{i=1}^N I(X_1^i; Y_i|X_2^i, Y^{i-1}) \\ &= \mathbf{E} \left[\log \frac{P(Y^N|X_1^N, X_2^N)}{P(Y^N|X_2^N)} \right] \end{aligned} \quad (17)$$

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

In this paper, we are using several properties of causal conditioning and directed information that follow from the definitions and simple algebra. Many of the key properties that hold for mutual information and regular conditioning carry over to directed information and causal conditioning, where $P(x^N)$ is replaced by $P(x^N|y^{N-1})$ and $P(y^N)$ is replaced by $P(y^N|x^N)$. Specifically, we have the following.

Lemma 1: (Analogue to $P(x_1^N, y^N) = P(x_1^N)P(y^N|x_1^N)$.) For arbitrary random vectors (X_1^N, X_2^N, Y^N)

$$P(x_1^N, y^N) = P(x_1^N|y^{N-1})P(y^N|x_1^N) \quad (18)$$

$$P(x_1^N, y^N|x_2^N) = P(x_1^N|y^{N-1}, x_2^N)P(y^N|x_1^N, x_2^N). \quad (19)$$

Lemma 2: (Analogue to $|I(X_1^N; Y^N) - I(X_1^N; Y^N|S)| \leq H(S)$.) For arbitrary random vectors and variables

$$|I(X_1^N \rightarrow Y^N) - I(X_1^N \rightarrow Y^N|S)| \leq H(S) \leq \log|\mathcal{S}| \quad (20)$$

$$\begin{aligned} |I(X_1^N \rightarrow Y^N|X_2^N) - I(X_1^N \rightarrow Y^N|X_2^N, S)| &\leq H(S) \\ &\leq \log|\mathcal{S}|. \end{aligned} \quad (21)$$

The proofs of Lemmas 1 and 2 can be found in [27, Sec. IV], along with some additional properties of causal conditioning and directed information. The next lemma, which is proved in Appendix I, shows that by replacing regular pmf with causal conditioning pmf we get the directed information. Let us denote the mutual information $I(X_1^N; Y^N|X_2^N)$ as a functional of $Q(x_1^N, x_2^N)$ and $P(y^N|x_1^N, x_2^N)$, i.e., $\mathcal{I}(Q(x_1^N, x_2^N); P(y^N|x_1^N, x_2^N)) \triangleq I(X_1^N; Y^N|X_2^N)$. Consider the case where the random variables X_1^N, X_2^N are independent, i.e., $Q(x_1^N, x_2^N) = Q(x_1^N)Q(x_2^N)$, then, by definition, we get (22) shown at the bottom of the page.

Lemma 3: If the random vectors X_1^N and X_2^N are causally conditionally independent given Y^{N-1} , i.e.,

$$Q(x_1^N, x_2^N|y^{N-1}) = Q(x_1^N|y^{N-1})Q(x_2^N|y^{N-1})$$

then

$$\begin{aligned} \mathcal{I}(Q(x_1^N|y^{N-1})Q(x_2^N|y^{N-1}); P(y^N|x_1^N, x_2^N)) \\ = I(X_1^N \rightarrow Y^N|X_2^N) \end{aligned} \quad (23)$$

where $\mathcal{I}()$ is defined in (22).

The next lemma, which is proved in Appendix II, shows that in the absence of feedback, mutual information becomes directed information.

Lemma 4: If $Q(x_1^N, x_2^N|y^{N-1}) = Q(x_1^N)Q(x_2^N)$ then

$$I(X_1^N; Y^N|X_2^N) = I(X_1^N \rightarrow Y^N|X_2^N). \quad (24)$$

B. Sub-, Sup-Additivity and Convergence of Finite-Dimensional Regions

In this subsection, we define basic operations (summation and multiplication by scalar), convergence, and sub/sup-additivity of regions. Furthermore, we show that the limit of a sup(sub)-additive sequence of regions converges to the union (intersection) of all regions.

Let A, B be sets in a finite-dimensional Euclidean space \mathbb{R}^d , i.e., A and B are sets of vectors in \mathbb{R}^d . The sum of two regions is denoted as $A + B$ and defined as

$$A + B = \{\mathbf{a} + \mathbf{b} : \mathbf{a} \in A, \mathbf{b} \in B\} \quad (25)$$

and multiplication of a set A with a scalar c is defined as

$$cA = \{c\mathbf{a} : \mathbf{a} \in A\}. \quad (26)$$

A sequence $\{A_n\}$, $n = 1, 2, 3, \dots$, of regions in \mathbb{R}^d is said to converge to a region A , written $A = \lim A_n$ if

$$\limsup A_n = \liminf A_n = A \quad (27)$$

where

$$\begin{aligned} \liminf A_n &= \{\mathbf{a} : \mathbf{a} = \lim \mathbf{a}_n, \mathbf{a}_n \in A_n\} \\ \limsup A_n &= \{\mathbf{a} : \mathbf{a} = \lim \mathbf{a}_{n_k}, \mathbf{a}_{n_k} \in A_{n_k}\} \end{aligned} \quad (28)$$

and n_k denotes an arbitrary increasing subsequence of the integers. In words, a point in \mathcal{R}^d is in \limsup (\liminf) if and only if the intersection of any ball around it with each set in the sequence is nonempty for infinitely (all but finitely) many of these sets. For more details on convergence of sets in finite dimensions see [36]. Let \bar{A} and \underline{A} , denote¹

$$\bar{A} = \text{cl} \left(\bigcup_{n \geq 1} A_n \right), \quad \underline{A} = \bigcap_{n \geq 1} A_n. \quad (29)$$

¹Closure is not needed in the definition of \underline{A} , since throughout the paper the sets A_n are closed and intersection of arbitrary many closed sets is closed.

$$\mathcal{I}(Q(x_1^N)Q(x_2^N); P(y^N|x_1^N, x_2^N)) \triangleq \sum_{y^N, x_1^N, x_2^N} Q(x_1^N)Q(x_2^N)P(y^N|x_1^N, x_2^N) \log \frac{P(y^N|x_1^N, x_2^N)}{\sum_{x_1^N} Q(x_1^N)P(y^N|x_1^N, x_2^N)}. \quad (22)$$

We say that a sequence $\{A_n\}_{n \geq 1}$ is *bounded*, if $\sup_{\mathbf{a} \in \bar{A}} \|\mathbf{a}\| < \infty$, where $\|\cdot\|$ denotes a norm in \mathbb{R}^d . The next two lemmas establish convergence of sub-additive and sup-additive sequences, and their proof can be found in Appendix III.

Lemma 5: Let $A_n, n = 1, 2, \dots$, be a sequence of bounded sets in \mathbb{R}^d that includes the origin, i.e., $(0, 0)$. If nA_n is sup-additive, i.e., for all $n \geq 1$ and all $N > n$

$$NA_N \supseteq nA_n + (N - n)A_{N-n} \quad (30)$$

then

$$\lim_{n \rightarrow \infty} A_n = \bar{A}. \quad (31)$$

Lemma 6: Let $A_n, n = 1, 2, \dots$, be a sequence of convex, closed, and bounded sets in \mathbb{R}^d . If nA_n is sub-additive, i.e., for all $n \geq 1$ and all $N > n$

$$NA_N \subseteq nA_n + (N - n)A_{N-n} \quad (32)$$

then

$$\lim_{n \rightarrow \infty} A_n = \underline{A}. \quad (33)$$

Corollary 7: For a sup-additive sequence, as defined in Lemma 5, the limit is convex.

This corollary follows immediately from the definition of the sup-additivity property, (32) where $n = \alpha N, 0 < \alpha < 1$, and N goes to infinity.

The (*Hausdorff*) *distance* between two sets A and B , is defined as

$$d(A, B) = \max\left\{\sup_{\mathbf{a} \in A} d(\mathbf{a}, B), \sup_{\mathbf{b} \in B} d(\mathbf{b}, A)\right\} \quad (34)$$

where the distance between a set A and a point \mathbf{b} is given by

$$d(\mathbf{b}, A) = \inf_{\mathbf{a} \in A} \|\mathbf{a} - \mathbf{b}\|. \quad (35)$$

An alternative and equivalent definition to \liminf and \limsup based on distance notation may be found in [37, Ch. 1]

$$\begin{aligned} \liminf A_n &= \left\{ \mathbf{a} : \lim_{n \rightarrow \infty} d(\mathbf{a}, A_n) = 0 \right\} \\ \limsup A_n &= \left\{ \mathbf{a} : \liminf_{n \rightarrow \infty} d(\mathbf{a}, A_n) = 0 \right\}. \end{aligned} \quad (36)$$

A straightforward consequence of this definition is the following lemma.

Lemma 8: If $\lim_{n \rightarrow \infty} d(A_n, B_n) = 0$, then $\limsup A_n = \limsup B_n$, and $\liminf A_n = \liminf B_n$.

IV. MAIN THEOREMS

We dedicate this section to a statement of our main results, proofs of which will appear in the subsequent sections (Theorem 9 in Section VI, Theorem 10 in Section VII, Theorem 11 in Section VIII-A, Theorem 12 in Section VIII-B, Theorem 13 in Section IX).

Let $\underline{\mathcal{R}}_n$ denote the following region in \mathbb{R}_+^2 (two-dimensional (2D) set of nonnegative real numbers) given in (37) at the bottom of the page, where $\text{conv}(\cdot)$ denotes the convex hull of the region. In order to avoid cases where $\underline{\mathcal{R}}_n$ is empty, we define that negative values in (37) are replaced by 0, hence, a more accurate notation of the first inequality is $R_1 \leq \max\{\min_{s_0} \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) - \frac{\log|S|}{n}, 0\}$.

The set of three inequalities is equivalent to an intersection of three regions, and \min_{s_0} is equivalent to \cap_{s_0} . Hence, an equivalent region is given in (38), also at the bottom of the page.

Theorem 9: (Inner bound.) For any FS-MAC with time-invariant feedback as shown in Fig. 1, and for any integer $n \geq 1$, the region $\underline{\mathcal{R}}_n$ is achievable, i.e., $\underline{\mathcal{R}}_n \subseteq \mathcal{R}$.

Let \mathcal{R}_n denote the following region in \mathbb{R}_+^2 :

$$\begin{aligned} \mathcal{R}_n &= \bigcup_{Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})} \left\{ \begin{array}{l} R_1 \leq \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n) \\ R_2 \leq \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n) \\ R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n) \end{array} \right\}. \end{aligned} \quad (39)$$

In the following theorem we use the standard notion of convergence of sets as defined in Section III.

Theorem 10: (Outer bound.) Let (R_1, R_2) be an achievable pair for an FS-MAC with time-invariant feedback, as shown in Fig. 1. Then, for any n there exists a distribution $Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})$ such that the following inequalities hold:

$$\begin{aligned} R_1 &\leq \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n) + \epsilon_n \\ R_2 &\leq \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n) + \epsilon_n \end{aligned}$$

$$\underline{\mathcal{R}}_n = \text{conv} \left(\bigcup_{Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})} \left\{ \begin{array}{l} R_1 \leq \min_{s_0} \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) - \frac{\log|S|}{n}, \\ R_2 \leq \min_{s_0} \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n, s_0) - \frac{\log|S|}{n}, \\ R_1 + R_2 \leq \min_{s_0} \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | s_0) - \frac{\log|S|}{n} \end{array} \right\} \right) \quad (37)$$

$$\underline{\mathcal{R}}_n = \text{conv} \left(\bigcup_{Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})} \bigcap_{s_0} \left\{ \begin{array}{l} R_1 \leq \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) - \frac{\log|S|}{n}, \\ R_2 \leq \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n, s_0) - \frac{\log|S|}{n}, \\ R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | s_0) - \frac{\log|S|}{n} \end{array} \right\} \right). \quad (38)$$

$$\overline{\mathcal{R}}_n = \text{conv} \left(\bigcup_{Q(x_1^n)Q(x_2^n)} \bigcup_{P(s_0|x_1^n, x_2^n)} \left\{ \begin{array}{l} R_1 \leq \frac{1}{n} I(X_1^n \rightarrow Y^n | X_2^n, S_0) + \frac{H(S_0)}{n}, \\ R_2 \leq \frac{1}{n} I(X_2^n \rightarrow Y^n | X_1^n, S_0) + \frac{H(S_0)}{n}, \\ R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | S_0) + \frac{H(S_0)}{n}. \end{array} \right. \right). \quad (41)$$

$$R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n) + \epsilon_n \quad (40)$$

where ϵ_n goes to zero as n goes to infinity. Moreover, the outer bound can be written as $\liminf \mathcal{R}_n$, i.e., $\mathcal{R} \subseteq \liminf \mathcal{R}_n$.

In the generality of time-invariant feedback, the outer bound in Theorem 10 is given only in terms of limit. However, if feedback is not allowed, we obtain the following outer bound which holds for any $n \geq 1$.

Let us denote $\overline{\mathcal{R}}_n$ as in (41) at the top of the page.

Theorem 11: (Outer bound for FS-MAC without feedback.)

For any FS-MAC without feedback, and for any integer $n \geq 1$, $\mathcal{R} \subseteq \overline{\mathcal{R}}_n$.

The following two theorems establish capacity for families of FS-MAC, by showing that the bounds given above are asymptotically tight.

Theorem 12: (Capacity of FS-MAC without feedback.) For any indecomposable FS-MAC without feedback, the capacity region is $\mathcal{R} = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \overline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$, and the limits exist.

Theorem 13: (Capacity of FS-MAC with feedback.) The capacity of a stationary and indecomposable FS-MAC without ISI, and with time-invariant feedback is $\mathcal{R} = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \overline{\mathcal{R}}_n$ and the limits exist.

The next theorems are proved in Section V and will be seen to be consequences of the capacity theorems given above.

Theorem 14: If the capacity without feedback of a stationary and indecomposable FS-MAC without ISI is zero, then it is also zero in the case in which there is feedback.

Corollary 15: For a memoryless MAC, the capacity with feedback is zero if and only if it is zero without feedback.

Corollary 16: Feedback does not enlarge the capacity region of a discrete additive (mod- $|\mathcal{X}|$) noise MAC.

In fact, among other results, we will see in the next section that the (mod- $|\mathcal{X}|$) noise MAC is only a subset of a larger family of MACs for which feedback does not enlarge the capacity region.

V. APPLICATIONS

The capacity formula of a FS-MAC given in Theorems 12 and 13 is a multiletter characterization. In general, it is very hard to evaluate it but, for the finite-state point-to-point channel, there are several cases where the capacity with and without feedback was found explicitly [38], [39], [26], [25], and [28].²

²For the Gaussian case without feedback there exists the water-filling solution [40], and recently the feedback capacity was found analytically, for the case where the noise is an ARMA(1)-Gaussian process (cf. [41]–[43]).

The multiletter capacity expression is also valuable for deriving useful concepts in communication. For instance, in order to show that feedback does not increase the capacity of a point-to-point memoryless channel (cf. [44]), we can use the multiletter upper bound of a channel with memory. Further, in [27] it was shown that for the cases where the capacity is given by the multiletter expression

$$C = \lim_{N \rightarrow \infty} \frac{1}{N} \max_{Q(x^N || z^{N-1})} I(X^N \rightarrow Y^N)$$

the source–channel coding separation holds. It was also shown that if the state of the channel is known at both the encoder and decoder and the channel is connected (i.e., every state can be reached with some positive probability from every other state under some input distribution), then feedback does not increase the capacity of the channel.

In this section, we use the capacity formula in order to derive three conclusions.

- 1) For stationary and indecomposable channels without ISI, the capacity is zero if and only if the capacity with feedback is zero.
- 2) Identify FS-MACs that feedback does not enlarge the capacity and show that for a MAC that can be decomposed into a “multiplexer” concatenated by a point-to-point channel (with, without, or with partial feedback), the capacity region is given by $\sum_m R_m \leq C$, where C is the capacity of the point-to-point channel.
- 3) Source–channel coding separation holds for a MAC that can be decomposed into a “multiplexer” concatenated by a point-to-point channel (with, without, or with partial feedback).

As a special case of the second concept, we show that the capacity of a binary Gilbert–Elliott MAC is $R_1 + R_2 \leq 1 - H(\mathcal{V})$, where $H(\mathcal{V})$ is the entropy rate of the hidden Markov noise that specifies the binary Gilbert–Elliott MAC.

A. Zero Capacity

The first concept is given in Theorem 14 and is proved here. The proof of Theorem 14 is based on the following lemma which is proved in Appendix IV.

Lemma 17: For a MAC described by an arbitrary causal conditioning $p(y^n | x_1^n, x_2^n)$ the following holds:

$$\begin{aligned} & \max_{Q(x_1^n || y^{n-1}) Q(x_2^n || y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) \\ & = 0 \iff \max_{Q(x_1^n) Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0 \end{aligned} \quad (42)$$

and each condition also implies that $P(y^n | x_1^n, x_2^n) = P(y^n)$ for all x_1^n, x_2^n .

Proof of Theorem 14: Since the channel is without ISI, i.e.,

$$P(y_i, s_i | x_{1,i}, x_{2,i}, s_{i-1}) = P(s_i | s_{i-1}) P(y_i | x_{1,i}, x_{2,i}, s_{i-1}) \quad (43)$$

and stationary and ergodic, its capacity region is given in Theorem 13 as $\mathcal{R} = \lim_{n \rightarrow \infty} \mathcal{R}_n$. Furthermore, since the sequence

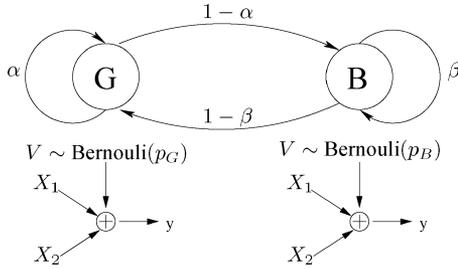


Fig. 2. Gilbert–Elliot MAC. It has two states, “Good” and “Bad” where the transition between them is according to a first-order Markov process. Given that the channel is in a “Good” (or a “Bad”) state, it behaves as binary additive noise where the noise is Bernoulli(p_G) (or Bernoulli(p_B)).

$\{\mathcal{R}_n\}$ is sup-additive (Lemma 27), then according to Lemma 5 $\mathcal{R} = \text{cl}\left(\bigcup_{n \geq 1} \mathcal{R}_n\right)$; hence, if $\mathcal{R} = 0$, then for all $n \geq 1$, $\mathcal{R}_n = 0$, i.e.,

$$\max_{Q(x_1^n)Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0. \quad (44)$$

According to Lemma 17, the maximization of the objective in (44) over the distribution $Q(x_1^n \| y^{n-1})Q(x_2^n \| y^{n-1})$ is still zero, hence, the capacity region is zero even if there is perfect feedback. \square

Corollary 15, which states that the capacity of a memoryless MAC without feedback is zero if and only if the capacity with feedback is zero, follows immediately from Theorem 14 because a memoryless MAC can be considered an FS-MAC with one state.

Clearly, Theorem 14 also holds for the case of a stationary and indecomposable FSC point-to-point channel without ISI, because a MAC is an extension of a point-to-point channel. However, it does not hold for the case of a broadcast channel. For instance, consider the binary broadcast channel given by $y_{1,i} = x \oplus n_i$ and $y_{2,i} = x \oplus n_{i-1}$, where n_i is an independent and identically distributed (i.i.d.) Bernoulli($\frac{1}{2}$) and \oplus denotes addition mod 2. The capacity without feedback is clearly zero, but if the transmitter has feedback, namely, if it knows $y_{1,i-1}$ and $y_{2,i-1}$ at time i , then it can compute the noise $n_{i-1} = y_{1,i-1} \oplus x_{i-1}$ and therefore it can transmit 1 bit per channel use to the second user.

B. Examples of Channels for Which Feedback Does Not Enlarge Capacity

1) *Gilbert–Elliot MAC*: The Gilbert–Elliot channel is a widely used example of a finite-state channel. It is often used to model wireless communication in the presence of fading [38], [39], [45]. The Gilbert–Elliot is a channel with two states without ISI, and the states are denoted as “good” and “bad.” Each state is a binary-symmetric channel and the probability of flipping the bit is lower in the “good” state. In the case of the Gilbert–Elliot MAC (Fig. 2), each state is an additive MAC with i.i.d. noise, where in the “good” channel the probability that the noise is “1” is lower than in the “bad” channel. This channel can be represented as an additive MAC as in Fig. 2, where the noise is a hidden Markov process.

Since the Gilbert–Elliot MAC is an indecomposable FS-MAC without ISI, its capacity with feedback when the initial state distribution over the states “good” and “bad” is the stationary distribution is given by $\lim_{n \rightarrow \infty} \mathcal{R}_n$ (Theorem 13). For the Gilbert–Elliot MAC, the region $\lim_{n \rightarrow \infty} \mathcal{R}_n$ reduces to the simple region

$$R_1 + R_2 \leq 1 - H(\mathcal{V}) \quad (45)$$

where $H(\mathcal{V})$ denotes the entropy rate of the hidden Markov noise. The following equalities and inequalities upper-bound the region \mathcal{R}_n and this upper bound can be achieved for any deterministic feedback by an i.i.d. input distribution $X_{1,i} \sim \text{Bernoulli}(\frac{1}{2})$ and $X_{2,i} \sim \text{Bernoulli}(\frac{1}{2})$, $i = 1, 2, \dots, n$ and X_1^n and X_2^n are independent of each other.

$$\begin{aligned} I((X_1, X_2)^n \rightarrow Y^n) &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X_1^i, X_2^i) \\ &\stackrel{(a)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | Y^{i-1}, X_1^i, X_2^i) \\ &= \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | V^{i-1}, Y^{i-1}, X_1^i, X_2^i) \\ &\stackrel{(b)}{=} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(V_i | V^{i-1}) \\ &\stackrel{(c)}{\leq} \sum_{i=1}^n \log 2 - H(V_i | V^{i-1}) \\ &= n \left(1 - \frac{H(V^n)}{n}\right). \end{aligned} \quad (46)$$

Equality (a) is due to the facts that y_i is a function of $(v_i, x_{1,i}, x_{2,i})$ and v_i is a deterministic function of $(y_i, x_{1,i}, x_{2,i})$, i.e., $y_i = x_{1,i} \oplus x_{2,i} \oplus v_i$ and $v_i = y_i \oplus x_{1,i} \oplus x_{2,i}$. Equality (b) follows from the fact that v_i is independent of the messages. Inequality (c) is due to the fact that the size of the alphabet \mathcal{V} is 2. Similarly, $\frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n) \leq 1 - \frac{H(V^n)}{n}$, and $\frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n) \leq 1 - \frac{H(V^n)}{n}$ and equality is achieved with an i.i.d. input distribution Bernoulli($\frac{1}{2}$). Finally, by dividing both sides by n and using the definition of entropy rate $H(\mathcal{V}) = \lim_{n \rightarrow \infty} \frac{1}{n} H(V^n)$ we conclude the proof.

2) *Multiplexer Followed by a Point-to-Point Channel*: Here we extend the Gilbert–Elliot MAC to the case where the discrete MAC can be decomposed into two components as shown in Fig. 3. The first component is a MAC that can behave as a multiplexer and the second component is a point-to-point channel. The definitions of those components are the following.

Definition 4: A MAC with m inputs, X_1, X_2, \dots, X_m , behaves as a *multiplexer* if the inputs and the output have common alphabets and for all $m \in 1, \dots, M$ there exists a choice of input symbols for all senders except sender m , such that the output is the m th input, i.e., $Y = X_m$.

An example of a multiplexer-MAC for the binary case is a MAC whose output is one of AND/OR/XOR of the inputs. For

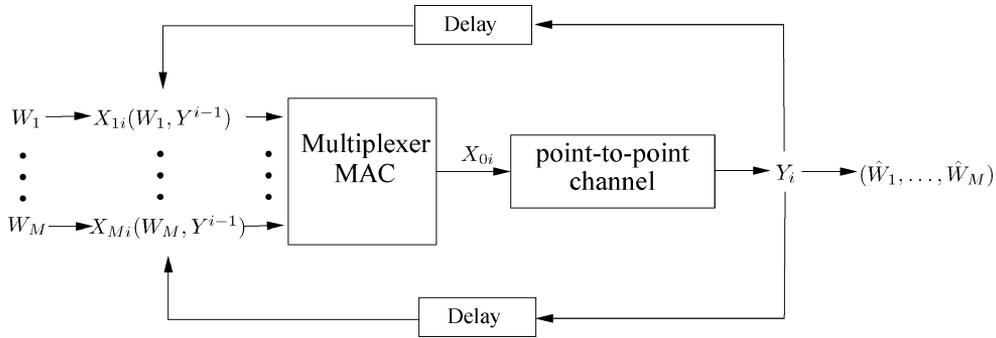


Fig. 3. Discrete MAC that can be decomposed into two parts. The first part is a MAC that behaves as a multiplexer and the second part is a point-to-point channel.

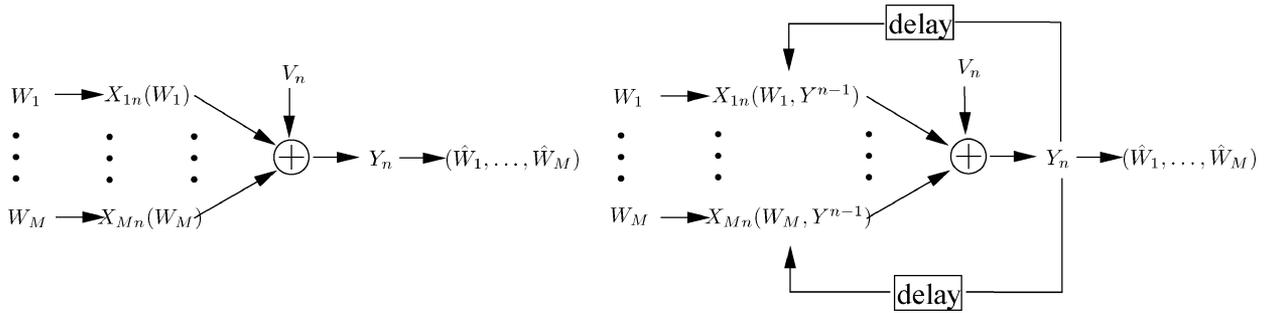


Fig. 4. Additive noise MAC with and without feedback. The random variables $X_{1n}, \dots, X_{Mn}, Y_n, V_n, n \in 1, 2, 3, \dots$, are from a common alphabet of size q , and they denote the input from sender $1, \dots, M$, the output, and the noise at time n , respectively. The relation between the random variables is given by $y_n = x_{1n} \oplus x_{2n} \cdots \oplus x_{Mn} \oplus v_n$ where \oplus denotes addition mod- q . The noise V_n , possibly with memory, is independent of the messages W_1, \dots, W_M .

a general alphabet q those operations could be max/min/addition-mod- q . For instance, if the channel is binary with two users and it is addition-mod-2, i.e., $y = x_1 \oplus x_2$, then we can ensure that $y = x_1$ by choosing $x_2 = 0$.

Theorem 18: The capacity region of a multiplexer MAC followed by a point-to-point channel with a time-invariant feedback to all encoders, as shown in Fig. 3, is

$$\sum_{m=1}^M R_m \leq C \tag{47}$$

where C is the capacity of the point-to-point channel with the time-invariant feedback $z_{i-1}(y_{i-1})$.

As shown by Dueck in [46], the capacity of a MAC, under the constraint that the maximum probability of error (among all messages) goes to zero, may be smaller than the capacity under the average error constraint. However, we show in the proof, that this theorem holds for the maximum error constraint, a fact that will be used for establishing source-channel separation.

Proof: The achievability is proved simply by time sharing. At each time, only one selected user sends information and the other users send a constant input that insures that the output is the input of the selected user. Since the time sharing is between communication over two point-to-point channel, and since for point-to-point channel the capacity is achieved with the maximum error criteria, we conclude the achievability in this theorem holds for the maximum criteria.

The converse is based on the fact that the maximum rate that can be transmitted through the point-to-point channel is C and

it is an upper bound sum-rate of multiplexer-MAC. If it had not been an upper bound for the multiplexer-MAC, we could build a fictitious multiplexer-MAC before the point-to-point channel and achieve by that a higher rate than its upper bound which would be contradiction. \square

3) *Discrete Additive MAC:* An immediate consequence of Theorem 18 is an extension of Alajaj’s result [20] to the additive MAC which is given in Corollary 16. Corollary 16 states that feedback does not enlarge the capacity region of a discrete additive (mod- $|X|$) noise MAC.

The proof of the corollary is based on the following observation. If feedback does not increase the capacity of a particular point-to-point channel then feedback also does not increase the capacity of the multiplexer (MUX) followed by the same particular channel. Specifically, feedback does not increase the achievable region of an additive MAC (Fig. 4) and the achievable region is given by

$$\sum_{m=1}^M R_m \leq \log q - H(\mathcal{V}) \tag{48}$$

where $H(\mathcal{V})$ is the entropy rate of the additive noise.

4) *Multiplexer Followed by Erasure Channel:* Consider the case of the multiplexer-erasure MAC which is a multiplexer followed by an erasure channel, possibly with memory.

Definition 5: A point-to-point channel is called *erasure* channel if the output at time n can be written as $Y_n = f(X_n, Z_n)$, and the following properties hold.

- 1) The alphabet of Z is binary and the alphabet of Y is the same as X plus one additional symbol called the erasure.

- 2) The process Z_n is stationary and ergodic and is independent of the message.
- 3) If $z_n = 0$, then $y_n = x_n$ and if $z_n = 1$, then the output is an erasure regardless of the input.

For the multiplexer–erasure channel we have the following theorem.

Corollary 19: The capacity region of the multiplexer–erasure MAC with or without feedback is

$$\sum_{m=1}^M R_m \leq (1 - p_e) \log q \quad (49)$$

where p_e is the marginal probability of having an erasure. Moreover, even if the encoder has noncausal side information, i.e., the encoders know where the erasures appear noncausally, the capacity is still given by (49).

Proof: According to Theorem 18, the capacity region is

$$\sum_{m=1}^M R_m \leq C \quad (50)$$

where C is the capacity of the erasure point-to-point channel. Diggavi and Grossglauser [47, Theorem 3.1] showed that the capacity of a point-to-point erasure channel, with and without feedback, is given by $(1 - p_e) \log q$. Since the probability of having an erasure does not depend on the input to the channel, we deduce that even in the case where the encoder knows the sequence Z^n noncausally, which is better than feedback, the transmitter can transmit only a fraction $1 - p_e$ of the time, hence, the capacity cannot exceed $(1 - p_e) \log q$. \square

5) *Multiplexer Followed by the Trapdoor Channel:* In this example, feedback increases the capacity. Based on the fact that the capacity of the trapdoor channel with feedback [28] is the logarithm of the Golden ratio, i.e., $\log \frac{\sqrt{5}+1}{2}$, the achievable region of a multiplexer followed by the trapdoor channel is

$$\sum_{m=1}^M R_m \leq \log \frac{\sqrt{5}+1}{2}. \quad (51)$$

C. Source–Channel Coding Separation

Cover, El-Gamal, and Salehi [48] showed that, in general, the source–channel separation does not hold for MACs even for a memoryless channel without feedback. However, for the case where the MAC is a discrete multiplexer followed by a channel we now show that it does hold.

We want to send the sequence of symbols U_1^n, U_2^n over the MAC, so that the receiver can reconstruct the sequence losslessly. To do this, we can use a joint source–channel coding scheme where we send through the channel the symbols $x_{1,i}(u_1^n, z^{i-1})$ and $x_{2,i}(u_2^n, z^{i-1})$. The receiver looks at his received sequence Y^n and makes an estimate \hat{U}_1^n, \hat{U}_2^n . The receiver makes an error if $\hat{U}_1^n \neq U_1^n$ or if $\hat{U}_2^n \neq U_2^n$, i.e., the probability of error $P_e^{(n)}$ is $P_e^{(n)} = \Pr((\hat{U}_1^n, \hat{U}_2^n) \neq (U_1^n, U_2^n))$.

Theorem 20: (Source–channel coding theorem for a multiplexer followed by a channel.) Let $(U_1, U_2)_{n \geq 1}$ be a finite alphabet, jointly stationary and ergodic pair of processes and let the MAC channel be a multiplexer followed by a point-to-point channel with time invariant feedback and capacity $C = \lim_{N \rightarrow \infty} \frac{1}{N} \max_{Q(x^n \| z^{n-1})} I(X^n \rightarrow Y^n)$ (e.g., a memoryless channel, an indecomposable FSC without feedback, stationary and indecomposable FSC without ISI). For the source and the MAC described above we have the following.

(direct part): There exists a source–channel code with $P_e^{(n)} \rightarrow 0$, if $H(\mathcal{U}_1, \mathcal{U}_2) < C$, where $H(\mathcal{U}_1, \mathcal{U}_2)$ is the entropy rate of the sources and C is the capacity of the point-to-point channel with a time-invariant feedback.

(converse part): If $H(\mathcal{U}_1, \mathcal{U}_2) > C$, then the probability of error is bounded away from zero (independent of the block length).

Proof: The achievability is a straightforward consequence of the Slepian–Wolf result for ergodic and stationary processes [49] and the achievability of the multiplexer followed by a point-to-point channel. First, we encode the sources by using the Slepian–Wolf achievability scheme, where we assign every u_1^n to one of 2^{nR_1} bins according to a uniform distribution on $\{1, \dots, 2^{nR_1}\}$, and, independently, we assign every u_2^n to one of 2^{nR_2} bins according to a uniform distribution on $\{1, \dots, 2^{nR_2}\}$. Second, we encode the bins as if they were messages, as shown in Fig. 5. The error is bounded by the sum of the maximum error (among all messages) at the MAC decoder and the error at the source decoder, namely, the error in the sequence decoding from the bin’s number. Both errors diminish as $n \rightarrow \infty$.

In the converse, we assume that there exists a sequence of codes with $P_e^{(n)} \rightarrow 0$, and we show that it implies that $H(\mathcal{U}_1, \mathcal{U}_2) \leq C$. Fix a given coding scheme and consider the following:

$$\begin{aligned} & H(U_1^n, U_2^n) \\ & \stackrel{(a)}{\leq} I(U_1^n, U_2^n; \hat{U}_1^n, \hat{U}_2^n) + n\epsilon_n \\ & \stackrel{(b)}{\leq} I(U_1^n, U_2^n; Y^n) + n\epsilon_n \\ & = H(Y^n) - H(Y^n | U_1^n, U_2^n) + n\epsilon_n \\ & = \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | U_1^n, U_2^n, Y^{i-1}) + n\epsilon_n \\ & \stackrel{(c)}{\leq} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | U_1^n, U_2^n, Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\ & \stackrel{(d)}{\leq} \sum_{i=1}^n H(Y_i | Y^{i-1}) - H(Y_i | Y^{i-1}, X_1^i, X_2^i) + n\epsilon_n \\ & = \sum_{i=1}^n I(X_1^i, X_2^i; Y_i | Y^{i-1}) + n\epsilon_n \\ & \stackrel{(e)}{\leq} \sum_{i=1}^n I(X_0^i; Y_i | Y^{i-1}) + n\epsilon_n \\ & = I(X_0^n \rightarrow Y^n) + n\epsilon_n \\ & \leq \max_{Q(x_0^n \| z^{n-1})} I(X_0^n \rightarrow Y^n) + n\epsilon_n. \end{aligned} \quad (52)$$

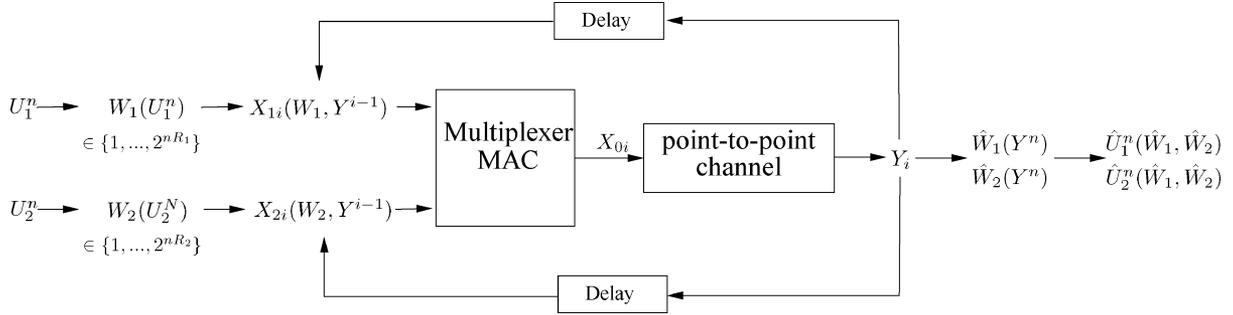


Fig. 5. Source–channel coding separation in a discrete multiplexer followed by a point-to-point channel.

Inequality (a) is due to Fano's inequality, where $n\epsilon_n = 1 + P_e^{(n)}n|\mathcal{U}_1||\mathcal{U}_2|$. Inequality (b) follows from the data processing inequality because $(U_1^N, U_2^N) - Y^N - (\hat{U}_1^N, \hat{U}_2^N)$ form a Markov chain. Equality (c) is due to the fact that, for a given code, X_1^i is a deterministic function of U_1^i, Y^{i-1} and, similarly, X_2^i is a deterministic function of U_2^i, Y^{i-1} . Equality (d) is due to the Markov chain $(U_1^N, U_2^N) - (X_1^i, X_2^i, Y^{i-1}) - Y_i$. The notation $X_{0,i}$ denotes the output of the multiplexer which is also the input to the point-to-point channel at time i . The inequality in (e) is due to the data processing inequality which can be invoked thanks to the fact that given Y^{i-1} we have the Markov chain $X_1^i, X_2^i - X_{0,i} - Y_i$.

By dividing both sides of (52) by n , taking the limit $n \rightarrow \infty$, and recalling that $C = \lim_{n \rightarrow \infty} \frac{1}{n} \max_{Q(x^n||z^{n-1})} I(X^n; Y^n)$ we have

$$H(\mathcal{U}_1, \mathcal{U}_2) = \lim_{n \rightarrow \infty} \frac{1}{n} H(U_1^n, U_2^n) \leq C. \quad (53)$$

□

We would like to emphasize that the source–channel coding separation theorem does not hold when the reconstruction of the sources is lossy³ (as opposed to the lossless case as in Theorem 20).

VI. PROOF OF ACHIEVABILITY (THEOREM 9)

The proof of achievability for the FS-MAC with feedback is similar to the proof of achievability for the point-to-point FSC given in [27, Sec. V], but there are two main differences.

- 1) In the case of FSC, only one message is sent, and in the case of FS-MAC, two independent messages are sent, which requires that we analyze three different types of errors: the first type occurs when only the first message is decoded with error, the second type occurs when only the second message is decoded with error, and the third type occurs when both messages are decoded with error.
- 2) In both cases, we generate the encoding scheme (code-trees) randomly but the distribution that is used is different. In the case of FSC we generate, for each message in $[1, \dots, 2^{NR}]$, a code-tree of length N by using the causal conditioning distribution $Q^*(x^N||z^{N-1}) = \arg \max_{Q(x^N||z^{N-1})} \min_{s_0} I(X^N \rightarrow Y^N|s_0)$, and here we

³Such an example was shown in a talk by Nazar and Gastpar [50], where the sources U_1, U_2 are i.i.d. Bernoulli($\frac{1}{2}$), independent of each other, the MAC channel is $y_i = X_{1,i} \oplus X_{2,i}$, and the reconstruction needs to be $u_{1,i} \oplus u_{2,i}$ rather than the sources themselves. It is easy to see that this can be achieved by having $x_{1,i} = u_{1,i}$ and $x_{2,i} = u_{2,i}$, but it cannot be achieved by transmitting messages through the MAC.

generate for each message in $[1, \dots, 2^{NR_l}], l = 1, 2$ a code-tree of length $N = Kn$ by concatenating K independent code-trees where each one is created with a causal conditioning distribution $Q(x_l^n||z_l^{n-1}), l = 1, 2$.

Encoding Scheme: Randomly generate for encoder $\{l \in 1, 2\}$, 2^{NR_l} code-trees of length $N = Kn$ by drawing it with the fixed distributions $Q(x_l^n||z_l^{n-1})$. In other words, given a feedback sequence z_1^{N-1} , the causal conditioning probability that the sequence x_1^N will be mapped to a given message is

$$Q(x_1^N||z_1^{N-1}) = \prod_{k=1}^K Q(x_{1,(k-1)n+1}^{kn}||z_{1,(k-1)n+1}^{kn}) \quad (54)$$

where $x_{1,(k-1)n+1}^{kn}$ denotes the vector

$$(x_{1,(k-1)n+1}, x_{1,(k-1)n+2}, \dots, x_{1,kn}).$$

Note, that such a construction is well defined, since the feedback is deterministic and time-invariant. Fig. 6 illustrates the concatenation of trees graphically. In order to shorten the notation, we will sometimes use the notation Q_N to denote $Q(x_1^N||z_1^{N-1})Q(x_2^N||z_2^{N-1})$ and we will express the concatenation of pmfs in (54) as $Q_N = \prod_{k=1}^K Q_n$.

Decoding Errors: For each code in the ensemble, the decoder uses maximum-likelihood (ML) decoding and we want to upper-bound the expected value $\mathbf{E}[P_e]$ for this ensemble. Let P_{e1}, P_{e2}, P_{e3} be defined as follows:

P_{e1} (type 1 error): probability that the decoded pair (m_1, m_2) satisfies $\hat{m}_1 \neq m_1, \hat{m}_2 = m_2$;

P_{e2} (type 2 error): probability that the decoded pair (m_1, m_2) satisfies $\hat{m}_1 = m_1, \hat{m}_2 \neq m_2$;

P_{e3} (type 3 error): probability that the decoded pair (m_1, m_2) satisfies $\hat{m}_1 \neq m_1, \hat{m}_2 \neq m_2$.

Because the error events are disjoint we have

$$P_e = P_{e1} + P_{e2} + P_{e3}. \quad (55)$$

In the next sequence of theorems and lemmas, we upper-bound the expected value of each error type and show that if (R_1, R_2) satisfies the three inequalities that define \underline{R}_n then the corresponding $\mathbf{E}[P_{ei}], i = 1, 2, 3$ goes to zero and hence $\mathbf{E}[P_e]$ goes to zero.

Theorem 21: Suppose that an arbitrary message $m_1, m_2, 1 \leq m_1 \leq M_1, 1 \leq m_2 \leq M_2$ enters the encoder with feedback and that ML decoding is employed. Let $\mathbf{E}[P_{e1}|m_1, m_2]$ denote the probability of decoding error averaged over the ensemble

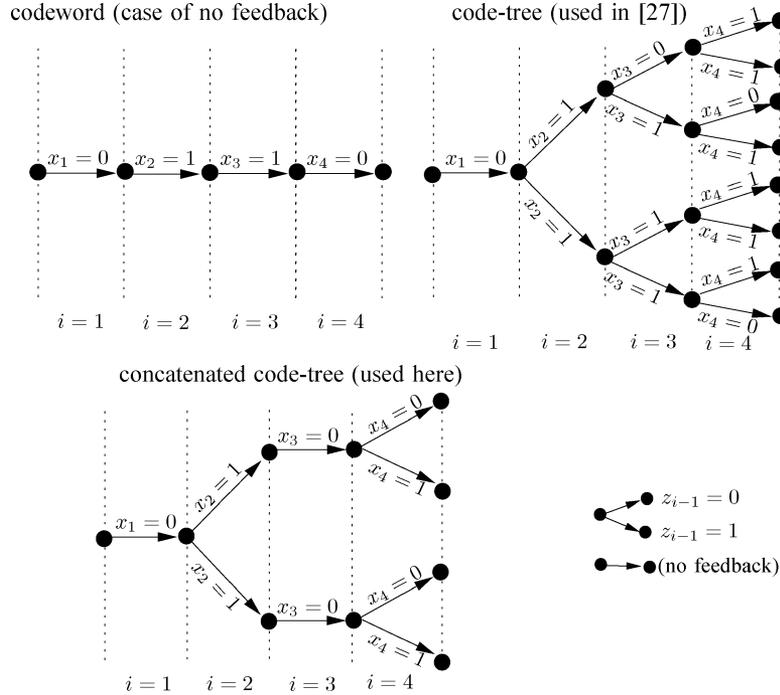


Fig. 6. Illustration of coding scheme for setting without feedback, setting with feedback as used for point-to-point channel [27], and a code-tree that was created by concatenating smaller code-trees. In the case of no feedback, each message is mapped to a codeword, and in the case of feedback each message is mapped to a code-tree. The third scheme is a code-tree of depth 4 created by concatenating two trees of depth 2.

of codes when the messages m_1, m_2 were sent. Then for any choice of $\rho, 0 < \rho \leq 1$, we get (56)–(58) shown at the bottom of the page.

The proof is given in Appendix V and is similar to [27, Theorem 9] only that here we take into account the fact that there are two encoders rather than one. For the memoryless MAC, this theorem has been proved by Kramer [14, eq. (61)], [13, eq. (5.32–5.34)]. There, the author used the fact that $P(y^{nK} \| x_1^{nK}, x_2^{nK}) = \prod_{k=1}^K P(y_{(k-1)n+1}^{kn} \| x_{1,(k-1)n+1}^{nk}, x_{2,(k-1)n+1}^{nk})$ which holds for a memoryless channel. The theorem coincides with Kramer’s result, if we consider the case $K = 1$ and $N = n$. The remainder of the proof of achievability, given in this section, diverges from the memoryless case.

Let $P_{ei}(s_0), i = 1, 2, 3$ be the probability of error of type i given that the initial state of the channel is s_0 . Also let $R_1 = \frac{1}{N} \log M_1$ and $R_2 = \frac{1}{N} \log M_2$ be the rate of the code and R_3

be the sum rate, i.e., $R_3 = R_1 + R_2$. The following theorem establishes exponential bounds on $\mathbf{E}[P_{ei}(s_0)]$.

Theorem 22: The average probability of error over the ensemble, for all initial states s_0 , and all $\rho, 0 \leq \rho \leq 1$, is bounded as

$$\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq |\mathcal{S}| 2^{\{-N[-\rho R_i + F_{N,i}(\rho, Q_N)]\}}, \quad i = 1, 2, 3 \quad (59)$$

where we get (60)–(63) at the top of the following page.

The proof is based on algebraic manipulation of the bounds given in (56)–(58). It is similar to the proof of Theorem 9 in [27] and therefore omitted. There are two differences between the proofs (and both are straightforward to accommodate): Here the input distribution $Q_N = Q(x_1^N \| z_1^N) Q(x_2^N \| z_2^N)$ is arbitrary while in [27] we chose the one that maximizes the error exponent. Second, here we bound the averaged error over the en-

$$\mathbf{E}[P_{e1} | m_1, m_2] \leq (M_1 - 1)^\rho \sum_{y^N, x_2^N} Q(x_2^N \| z^{N-1}) \left[\sum_{x_1^N} Q(x_1^N \| z_1^{N-1}) P(y^N \| x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (56)$$

$$\mathbf{E}[P_{e2} | m_1, m_2] \leq (M_2 - 1)^\rho \sum_{y^N, x_1^N} Q(x_1^N \| z_1^{N-1}) \left[\sum_{x_2^N} Q(x_2^N \| z^{N-1}) P(y^N \| x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (57)$$

$$\mathbf{E}[P_{e3} | m_1, m_2] \leq ((M_2 - 1)(M_2 - 1))^\rho \sum_{y^N} \left[\sum_{x_1^N, x_2^N} Q(x_1^N \| z_1^{N-1}) Q(x_2^N \| z^{N-1}) P(y^N \| x_1^N, x_2^N)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (58)$$

$$F_{N,i}(\rho, Q_N) = -\frac{\rho \log |\mathcal{S}|}{N} + \left[\min_{s_0} E_{N,i}(\rho, Q_N, s_0) \right], \quad i = 1, 2, 3 \quad (60)$$

$$E_{N,1}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N, x_2^N} Q(x_2^N \| z_1^{N-1}) \left[\sum_{x_1^N} Q(x_1^N \| z_1^{N-1}) P(y^N \| x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (61)$$

$$E_{N,2}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N, x_1^N} Q(x_1^N \| z_1^{N-1}) \left[\sum_{x_2^N} Q(x_2^N \| z_1^{N-1}) P(y^N \| x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho} \quad (62)$$

$$E_{N,3}(\rho, Q_N, s_0) = -\frac{1}{N} \log \sum_{y^N} \left[\sum_{x_1^N, x_2^N} Q(x_1^N \| z_1^{N-1}) Q(x_2^N \| z_1^{N-1}) P(y^N \| x_1^N, x_2^N, s_0)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \quad (63)$$

semble and in [27] we have an additional step where we claim that there exists a code that has an error that is bounded by the expression in (59). Because of this difference, the bound on the probability of error in [27] has an additional factor of 4.

The following theorem presents a few properties of the functions $E_{N,i}(\rho, Q_N, s_0)$, $i = 1, 2, 3$, such as positivity of the function and its derivative, convexity with respect to ρ , and an upper bound on the derivative which is achieved for $\rho = 0$.

Theorem 23: The term $E_{N,i}(\rho, Q_N, s_0)$ has the following properties:

$$E_{N,i}(\rho, Q_N, s_0) \geq 0; \quad \rho \geq 0, i = 1, 2, 3 \quad (64)$$

$$\begin{aligned} \frac{1}{N} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) &\geq \frac{\partial E_{N,1}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \\ \frac{1}{N} I(X_2^N \rightarrow Y^N \| X_1^N, s_0) &\geq \frac{\partial E_{N,2}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \\ \frac{1}{N} I(X_1^N, X_2^N \rightarrow Y^N | s_0) &\geq \frac{\partial E_{N,3}(\rho, Q_N, s_0)}{\partial \rho} > 0; \quad \rho \geq 0 \end{aligned} \quad (65)$$

$$\frac{\partial^2 E_{N,i}(\rho, Q_N, s_0)}{\partial \rho^2} > 0; \quad \rho \geq 0, i = 1, 2, 3. \quad (66)$$

Furthermore, equality holds in (64) when $\rho = 0$, and equality holds on the left sides of (65) when $\rho = 0$ for $i = 1, 2, 3$.

The proof of the theorem is the same as [51, eq. (2.20)], [17, Theorem 5.6.3]. In [51] the arguments Q_N of $E_{N,1}(\rho, Q_N, s_0)$ are regular conditioning i.e., $Q(x_1^N)Q(x_2^N)$, and the channel is given by $P(y^N | x_1^N, x_2^N, s_0)$, hence the derivative of $E_{N,1}(\rho, Q_N, s_0)$ with respect to ρ is upper-bounded by $I(X_1^N; Y^N | X_2^N, s_0)$. Here we replace $Q(x_1^N)Q(x_2^N)$ with $Q(x_1^N \| z_1^{N-1})Q(x_2^N \| z_1^{N-1})$ and $P(y^N | x_1^N, x_2^N, s_0)$ with $P(y^N \| x_1^N, x_2^N, s_0)$ and, according to Lemma 3, the upper bound becomes $I(X_1^N \rightarrow Y^N \| X_2^N, s_0)$. The next lemma establishes the sup-additivity of $F_{N,i}(\rho, Q_N)$, $i = 1, 2, 3$.

Lemma 24: Sup-additivity of $F_{N,i}(\rho, Q_N)$. For any finite-state channel, $F_{N,i}(\rho, Q_N)$, as given by (60), satisfies

$$\begin{aligned} &F_{n+l,i}(\rho, (Q_n Q_l)) \\ &\geq \frac{n}{n+l} F_{n,i}(\rho, Q_n) + \frac{l}{n+l} F_{l,i}(\rho, Q_l), \quad i = 1, 2, 3. \end{aligned} \quad (67)$$

The proof steps are identical to the proof of the sub-additivity for the point-to-point channel [27, Lemma 12].

Invoking this lemma on the pmf $Q_N = \prod_{k=1}^K Q_n$ where $N = nK$ we get

$$F_{N,i}(\rho, Q_N) \geq K \frac{n}{N} F_{n,i}(\rho, Q_n) = F_{n,i}(\rho, Q_n). \quad (68)$$

Let us define

$$\underline{C}_{N,1}(Q_N) = \frac{1}{N} \min_{s_0} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) \quad (69)$$

$$\underline{C}_{N,2}(Q_N) = \frac{1}{N} \min_{s_0} I(X_2^N \rightarrow Y^N \| X_1^N, s_0) \quad (70)$$

$$\underline{C}_{N,3}(Q_N) = \frac{1}{N} \min_{s_0} I(X_1^N, X_2^N \rightarrow Y^N | s_0) \quad (71)$$

where the joint distribution of X_1^N, X_2^N, Y^N conditioned on s_0 is given by

$$\begin{aligned} &P(x_1^N, x_2^N, y^N | s_0) \\ &= Q(x_1^N \| z_1^{N-1}) Q(x_2^N \| z_2^{N-1}) P(y^N \| x_1^N, x_2^N, s_0). \end{aligned}$$

Theorem 9 (inner bound) given in Section IV states that for every n and $0 \leq R_i < \underline{C}_{n,i}(Q_n) - \frac{\log |\mathcal{S}|}{n}$, $i = 1, 2, 3$ (recall, $R_3 \triangleq R_1 + R_2$) and every $\eta > 0$ there exists an N and an $(N, \lceil 2^{NR_1} \rceil, \lceil 2^{NR_1} \rceil)$ code with a probability of error $P_e(s_0)$ (averaged over the messages) that is less than η for all initial states s_0 .

Proof of Theorem 9: The proof consists of the following three steps.

- Showing that for a fixed n if $R_i < \underline{C}_{n,i}(Q_n) - \frac{\log |\mathcal{S}|}{n}$, $i = 1, 2, 3$ then there exists ρ^* such that

$$F_{n,i}(\rho^*, Q_n) - \rho^* R_i > 0, \quad i = 1, 2, 3. \quad (72)$$

- We choose $\epsilon < \min_{i \in \{1,2,3\}} F_{n,i}(\rho^*, Q_n) - \rho^* R_i$ and show that for sufficiently large N

$$\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq 2^{-N([F_{n,i}(\rho^*, Q_n) - \rho^* R_i] - \epsilon)}, \quad \forall s_0. \quad (73)$$

- From the last step we deduce the existence of a $(N, \lceil 2^{NR_1} \rceil, \lceil 2^{NR_1} \rceil)$ code s.t.

$$P_e(s_0) < \eta, \quad \forall s_0. \quad (74)$$

$$\mathcal{R}_n = \text{conv} \left(\bigcup_{Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})} \bigcap_{s_0, u_0} \left\{ \begin{array}{l} R_1 \leq \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n, s_0, u_0) - \frac{\log |\mathcal{S}| |\mathcal{U}|}{n}, \\ R_2 \leq \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n, s_0, u_0) - \frac{\log |\mathcal{S}| |\mathcal{U}|}{n}, \\ R_1 + R_2 \leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n | s_0, u_0) - \frac{\log |\mathcal{S}| |\mathcal{U}|}{n}. \end{array} \right. \right) \quad (82)$$

First Step: for any pair (R_1, R_2) , we can rewrite (59) for $i = 1, 2, 3$ as

$$\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq 2^{-N(F_{N,i}(\rho, Q_N) - \rho R_i - \frac{\log |\mathcal{S}|}{n})}. \quad (75)$$

By using (68), which states that $F_{N,i}(\rho, Q_N) \geq F_{n,i}(\rho, Q_n)$, we get

$$\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq 2^{-N(F_{n,i}(\rho, Q_n) - \rho R_i - \frac{\log |\mathcal{S}|}{n})}. \quad (76)$$

Note that $F_{n,i}(\rho, Q_n)$ and therefore $F_{n,i}(\rho, Q_n) - \rho R_i$ is continuous in $\rho \in [0, 1]$, so there exists a maximizing ρ . Let us show that if $R_1 < \underline{C}_{n,1}(Q_n) - \frac{\log |\mathcal{S}|}{n}$, then $\max_{0 \leq \rho \leq 1} [F_{n,1}(\rho, Q_n) - \rho R_1] > 0$ (the cases $i = 2, 3$ are identical to $i = 1$). Let us define $\delta \triangleq \underline{C}_{n,1} - R_1$. From Theorem 23, we have that $E_{n,1}(\rho, Q_n, s_0)$ is zero when $\rho = 0$, is a continuous function of ρ , and its derivative at zero with respect to ρ is equal or greater to $\underline{C}_{n,1}$, which satisfies $\underline{C}_{n,1} \geq R_1 + \frac{\log |\mathcal{S}|}{n} + \frac{\delta}{2}$. Thus, for each state s_0 , there is a range $\rho > 0$ such that

$$E_{n,1}(\rho, Q_n, s_0) - \rho \left(R_1 + \frac{\log |\mathcal{S}|}{n} \right) > 0. \quad (77)$$

Moreover, because the number of states is finite, there exists a $\rho^* > 0$ for which the inequality (77) is true for all s_0 . Thus, from the definition of $F_{n,1}(\rho^*, Q_n)$ given in (60) and from (77)

$$F_{n,1}(\rho^*, Q_n) = -\rho^* \frac{\log |\mathcal{S}|}{n} + \min_{s_0} E_{n,1}(\rho^*, Q_n, s_0) > \rho^* R_1, \quad \forall s_0. \quad (78)$$

Second Step: We choose a positive number ϵ such that $\epsilon < \min_{i \in \{1,2,3\}} F_{n,i}(\rho^*, Q_n) - \rho^* R_i$. It follows from (76) that for every N that satisfies $N > \frac{\log |\mathcal{S}|}{\epsilon}$

$$\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq 2^{-N(F_{n,i}(\rho^*, Q_n) - \rho^* R_i - \epsilon)} \quad (79)$$

and according to the first step of the proof, the exponent $F_{n,i}(\rho^*, Q_n, s_0) - \rho^* R_i - \epsilon$ is strictly positive.

Third Step: According to the previous step, for all $\frac{\eta}{3|\mathcal{S}|+1} > 0$ there exists an N such that $\mathbf{E}[P_{ei}(s_0) | m_1, m_2] \leq \frac{\eta}{3|\mathcal{S}|+1}$ for all $i \in 1, 2, 3$ all $s_0 \in \mathcal{S}$ and all messages. Since $P_e(s_0) = \sum_{i=1}^3 P_{ei}(s_0)$, then $\mathbf{E}[P_e(s_0) | m_1, m_2] \leq \frac{\eta}{|\mathcal{S}|+1}$; furthermore, $\mathbf{E}[P_e(s_0)] \leq \frac{\eta}{|\mathcal{S}|+1}$ for all $s_0 \in \mathcal{S}$. By using the Markov inequality, we have

$$\Pr(P_e(s_0) \geq \eta) \leq \frac{1}{|\mathcal{S}|+1} \quad (80)$$

and by using the union bound we have

$$\begin{aligned} \Pr(P_e(s_0) \geq \eta, \text{ for some } s_0 \in \mathcal{S}) \\ \leq \sum_{s_0 \in \mathcal{S}} \Pr(P_e(s_0) \geq \eta) = \frac{|\mathcal{S}|}{|\mathcal{S}|+1} < 1. \end{aligned} \quad (81)$$

Because the probability over the ensemble of codes of having a code with error probability (averaged over all messages) that is less than η for all initial states is positive, there must exist at least one code that has an error probability (averaged over all messages) that is less than η for all initial states. \square

Remark: (Extending the proof to feedback with states): In general, it is not possible to extend the proof of achievability to the general case where the feedback is any time-varying function of the channel output. The proof does not carry on because of two reasons: first, in the generation of random concatenated code-trees (54), we assumed that the feedback is time-invariant, and second we used a time-shift in the proof of Lemma 24 (subadditivity of $F_{N,i}(\rho, Q_N)$). However, it is possible to extend the proof to the case where the feedback has a finite number of states.

Assume that the feedback is of the form $z_{1,i} = f_1(y_i, u_{i-1})$, $z_{2,i} = f_2(y_i, u_{i-1})$, and $u_i = f_3(u_{i-1}, y_i)$, where $f_1(), f_2(), f_3()$ are time-invariant deterministic functions and $|\mathcal{U}| < \infty$. This, for instance, includes the case where the feedback $(z_{1,i}, z_{2,i})$ is a function of the k -tuple y_{i-k+1}^i . All the results above hold for this kind of feedback, only that the state of the system at time i is now (s_{i-1}, u_{i-1}) , and in all the places, s_{i-1} needs to be replaced with (s_{i-1}, u_{i-1}) . Hence, the inner region \mathcal{R}_n for this case is given in (82) at the top of the page.

VII. PROOF OF THE OUTER BOUND (THEOREM 10)

The first part of Theorem 10, states that for any FS-MAC, and any achievable rate-pair (R_1, R_2) , there exists a distribution $Q(x_1^n \| z_1^{n-1}) Q(x_2^n \| z_2^{n-1})$ such that the following inequalities hold:

$$\begin{aligned} R_1 &\leq \frac{1}{n} I(X_1^n \rightarrow Y^n \| X_2^n) + \epsilon_n \\ R_2 &\leq \frac{1}{n} I(X_2^n \rightarrow Y^n \| X_1^n) + \epsilon_n \\ R_1 + R_2 &\leq \frac{1}{n} I((X_1, X_2)^n \rightarrow Y^n) + \epsilon_n \end{aligned} \quad (83)$$

where ϵ_n goes to zero as n goes to infinity.

The proof is identical to the converse proof for memoryless MAC given by Kramer in [14, Sec. VII-A] or in [13, Proposition 5.1] and is therefore omitted. The proof is based on Fano's inequality and in the proof ϵ_n satisfies $h^{-1} \left(\frac{\epsilon_n}{\epsilon_n + \log |\mathcal{Y}|} \right) \leq P_e^{(n)}$, where h^{-1} is the inverse binary entropy function. Clearly, if $P_e^{(n)} \rightarrow 0$ then also $\epsilon_n \rightarrow 0$. The second part of Theorem is stated in the following corollary.

Corollary 25: The outer bound given in (83) implies that $\liminf \mathcal{R}_n$ is an outer bound for the capacity region.

Proof: Recall the definition of \mathcal{R}_n in (39). Let (R_1, R_2) be an achievable rate pair. We will create a sequence of rate pairs $(R_{1,n}, R_{2,n}) \in \mathcal{R}_n$ that converges to (R_1, R_2) and, therefore,

by the definition of \liminf of a sequence of sets $(R_1, R_2) \in \liminf \mathcal{R}_n$.

If $(R_1, R_2) \in \mathcal{R}_n$ then we choose $(R_{1,n}, R_{2,n}) = (R_1, R_2)$. Otherwise, we choose the closest point in \mathcal{R}_n to (R_1, R_2) . Because of inequality (83), the distance

$$\|(R_{1,n}, R_{2,n}) - (R_1, R_2)\| \leq 2\epsilon_n$$

and, therefore, the sequence $(R_{1,n}, R_{2,n})$ converges to (R_1, R_2) . \square

We would like to point out that if the probability of initial state, $P(s_0)$, is positive for all $s_0 \in \mathcal{S}$, then the error probability given the initial state s_0 , i.e., $P_e^{(n)}(s_0)$, needs to go to zero for all $s_0 \in \mathcal{S}$. Further, since the messages are independent of the initial state, we obtain that for any FS-MAC, and any achievable rate-pair (R_1, R_2) , there exists a distribution $Q(x_1^n \| z_1^{n-1})Q(x_2^n \| z_2^{n-1})$ such that the following inequalities hold:

$$\begin{aligned} R_1 &\leq \frac{1}{n}I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \epsilon_n \\ R_2 &\leq \frac{1}{n}I(X_2^n \rightarrow Y^n \| X_1^n, s_0) + \epsilon_n \\ R_1 + R_2 &\leq \frac{1}{n}I((X_1, X_2)^n \rightarrow Y^n, s_0) + \epsilon_n \end{aligned} \quad (84)$$

for all $s_0 \in \mathcal{S}$, and ϵ_n satisfies

$$h^{-1} \left(\frac{\epsilon_n}{\epsilon_n + \log |\mathcal{Y}|} \right) \leq \max_{s_0} P_e^{(n)}(s_0).$$

Equation (84) implies that if $P(s_0) > 0, \forall s_0 \in \mathcal{S}$, then $\lim \underline{\mathcal{R}}_n$ is an outer bound. (It is easy to observe this from the definition of $\underline{\mathcal{R}}_n$ that is given in (38).) Finally, since $\lim \underline{\mathcal{R}}_n$ is also achievable (Theorem 9), we conclude that if $P(s_0) > 0, \forall s_0 \in \mathcal{S}$, the capacity region is $\mathcal{R} = \lim \underline{\mathcal{R}}_n$.

VIII. FS-MAC WITHOUT FEEDBACK

The inner and outer bounds given in Theorems 9 and 10 specialize to the case where there is no feedback, i.e., z_1, z_2 are null. The outer bound from Theorem 10, which is given in terms of limit, can be used to establish a sequence finite-letter outer bounds for the FS-MAC without feedback. Furthermore, the inner and the outer bounds can be used in order to extend Gallager's results [17, Ch. 4] on the capacity of indecomposable FSCs to indecomposable FS-MACs.

A. Finite-Letter Outer Bound

In order to establish a finite-letter outer bound, we would like to show that the sequence $\max_{s_0} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) + \log |\mathcal{S}|$, induced by an input distribution of the form $Q(x_1^N \| z_1^N, s_0)Q(x_2^N \| z_2^N, s_0)$, is sub-additive. However, unlike the FSC case, the fact that

$$\begin{aligned} \max_{s_0} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) &\leq \max_{s_0} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) \\ &+ \max_{s_n} I(X_{1,n+1}^N \rightarrow Y_{1,n+1}^N \| X_{2,n+1}^N, s_n, s_0) + \log |\mathcal{S}| \end{aligned} \quad (85)$$

does not necessarily imply that for any input distribution $Q(x_1^N \| z_1^N, s_0)Q(x_2^N \| z_2^N, s_0)$, there exist two

input distributions $Q(x_1^n \| z_1^n, s_0)Q(x_2^n \| z_2^n, s_0)$ and $Q(x_1^l \| z_1^l, s_0)Q(x_2^l \| z_2^l, s_0)$, where $N = n + l$ that satisfies

$$\begin{aligned} \max_{s_0} I(X_1^N \rightarrow Y^N \| X_2^N, s_0) &\leq \max_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) \\ &+ \max_{s_0} I(X_1^l \rightarrow Y^l \| X_2^l, s_0) + \log |\mathcal{S}|. \end{aligned} \quad (86)$$

The Inequality (85) does not imply (86), since an input distribution of the form $Q(x_1^N \| z_1^N, s_0)Q(x_2^N \| z_2^N, s_0)$ cannot be always decomposed as

$$\begin{aligned} Q(x_{1,n+1}^N, x_{2,n+1}^N \| z_{1,n+1}^N, z_{2,n+1}^N, s_n) \\ = Q(x_{1,n+1}^N \| z_{1,n+1}^N, s_n)Q(x_{2,n+1}^N \| z_{2,n+1}^N, s_n). \end{aligned} \quad (87)$$

Even if there is no feedback, namely, $z_{1,i}$ and $z_{2,i}$ are null for all i , (87) still does not necessarily hold. Because of this fact, we introduce the idea of union over distribution $P(s_0 | x_1^n, x_2^n)$, rather than maximizing over the initial state s_0 .

Recall the definition of $\overline{\mathcal{R}}_n$ in (41). The following lemma establishes the subadditivity of $\overline{\mathcal{R}}_n$ and its proof is provided in Appendix VI.

Lemma 26: (sub-additivity of $\overline{\mathcal{R}}_n$.) For any FS-MAC, the sequence $\{\overline{\mathcal{R}}_n\}$, is sub-additive, i.e.,

$$(n+l)\overline{\mathcal{R}}_{n+l} \subseteq n\overline{\mathcal{R}}_n + l\overline{\mathcal{R}}_l. \quad (88)$$

Proof of Theorem 11: The theorem states that for any n , the region $\overline{\mathcal{R}}_n$ is an outer bound, namely, $\mathcal{R} \subseteq \overline{\mathcal{R}}_n$. The following relation proves the theorem:

$$\mathcal{R} \stackrel{(a)}{\subseteq} \liminf \mathcal{R}_n \stackrel{(b)}{\subseteq} \lim \overline{\mathcal{R}}_n \stackrel{(c)}{\subseteq} \overline{\mathcal{R}}_n. \quad (89)$$

Step (a) follows from Theorem 10. Step (b) follows from the fact that, for any n , $\mathcal{R}_n \subseteq \overline{\mathcal{R}}_n$. This is easy to observe since by Lemma 2, we have $I(X_1^n \rightarrow Y^n \| X_2^n) \leq I(X_1^n \rightarrow Y^n \| X_2^n, S_0) + H(S_0)$. Finally, Step (c) is due to the sub-additivity of $\{\overline{\mathcal{R}}_n\}$ (Lemma 26). Since $\{\overline{\mathcal{R}}_n\}$ is also convex, closed, an bounded, we can apply Lemma 6 and conclude that $\lim_n \overline{\mathcal{R}}_n = \bigcap_n \overline{\mathcal{R}}_n \subseteq \overline{\mathcal{R}}_n$. \square

B. Capacity of Indecomposable FS-MAC

Since there is no feedback, according to Lemma 4, directed information becomes mutual information and causal conditioning becomes regular conditioning in all the expressions in the inner bound (Theorem 9) and outer bound (Theorem 10).

First we establish the sup-additivity of $\underline{\mathcal{R}}_n$, which is proved in Appendix VII.

Lemma 27: (sup-additivity of $\underline{\mathcal{R}}_n$.) For any FS-MAC, the sequence $\{\underline{\mathcal{R}}_n\}$, which is defined in (37), is sup-additive, i.e.,

$$(n+l)\underline{\mathcal{R}}_{n+l} \supseteq n\underline{\mathcal{R}}_n + l\underline{\mathcal{R}}_l \quad (90)$$

and therefore $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n$ exists. Moreover, for an indecomposable FS-MAC without feedback $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$, where \mathcal{R}_n is defined (39).

Now, recall that Theorem 12 states that for any FS-MAC without feedback

$$\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \overline{\mathcal{R}}_n = \mathcal{C}. \quad (91)$$

$$\begin{aligned}
 \Delta &= \sum_{x_1^N, x_2^N, s_n, y_{n+1}^N} |\bar{P}(x_1^N, x_2^N, s_n, y_{n+1}^N) - \underline{P}(x_1^N, x_2^N, s_n, y_{n+1}^N)| \\
 &= \sum_{x_1^N, x_2^N, s_n, y_{n+1}^N} P(y_{n+1}^N | x_{1,n+1}^N, x_{2,n+1}^N, s_n) |\bar{P}(x_1^N, x_2^N, s_n) - \underline{P}(x_1^N, x_2^N, s_n)| \\
 &= \sum_{x_1^N, x_2^N, s_n} |\bar{P}(x_1^N, x_2^N, s_n) - \underline{P}(x_1^N, x_2^N, s_n)|. \tag{98}
 \end{aligned}$$

$$\begin{aligned}
 &\sum_{x_1^N, x_2^N, s_n} |\bar{P}(x_1^N, x_2^N, s_n) - \underline{P}(x_1^N, x_2^N, s_n)| \\
 &\stackrel{(a)}{=} \sum_{x_1^N, x_2^N, s_n} Q(x_1^N)Q(x_2^N) \left| \sum_{s_0} \bar{P}(s_0 | x_1^N, x_2^N) P(s_n | x_1^n, x_2^n, s_0) - \sum_{s_0} \underline{P}(s_0 | x_1^N, x_2^N) P(s_n | x_1^n, x_2^n, s_0) \right| \\
 &\stackrel{(b)}{\leq} \sum_{x_1^N, x_2^N, s_n} Q(x_1^N)Q(x_2^N) \left(\left| \sum_{s_0} \bar{P}(s_0 | x_1^N, x_2^N) P(s_n | x_1^n, x_2^n) - \sum_{s_0} \underline{P}(s_0 | x_1^N, x_2^N) P(s_n | x_1^n, x_2^n) \right| + 2\epsilon \right) \\
 &= 2\epsilon |\mathcal{S}| \tag{99}
 \end{aligned}$$

Proof of Theorem 12: Since

$$\min_{s_0} I(X_1^N; Y^N | X_2^N, s_0) \geq \min_{P(s_0 | x_1^n, x_2^n)} I(X_1^N; Y^N | X_2^N, S_0)$$

it is enough to show that if the channel is indecomposable, then in the limit $P(s_0 | x_1^n, x_2^n)$ does not influence the objective, and hence, $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \bar{\mathcal{R}}_n$.

Let us denote

$$\begin{aligned}
 \bar{P}(s_0, x_1^N, x_2^N) &\triangleq Q(x_1^N)Q(x_2^N)\bar{P}(s_0 | x_1^N, x_2^N) \\
 \underline{P}(s_0, x_1^N, x_2^N) &\triangleq Q(x_1^N)Q(x_2^N)\underline{P}(s_0 | x_1^N, x_2^N). \tag{92}
 \end{aligned}$$

Let $\bar{I}(X_1^N; Y^N | X_2^N, S_0)$ and $\underline{I}(X_1^N; Y^N | X_2^N, S_0)$ be the mutual information induced by $\bar{P}(s_0, x_1^N, x_2^N)$ and $\underline{P}(s_0, x_1^N, x_2^N)$, respectively.

We are proving the theorem by showing that exists a function $\delta(N)$ that depends on N but not on $\bar{P}(s_0, x_1^N, x_2^N)$, $\underline{P}(s_0, x_1^N, x_2^N)$ such that

$$0 \leq \frac{1}{N} \bar{I}(X_1^N; Y^N | X_2^N, S_0) - \frac{1}{N} \underline{I}(X_1^N; Y^N | X_2^N, S_0) \leq \delta(N) \tag{93}$$

and $\lim_{N \rightarrow \infty} \delta(N) = 0$. We have

$$\begin{aligned}
 &\bar{I}(X_1^N; Y^N | X_2^N, S_0) \\
 &\stackrel{(a)}{=} \bar{I}(X_1^N; Y^n | X_2^N, S_0) + \bar{I}(X_1^N; Y_{n+1}^N | X_2^N, Y^n, S_0) \\
 &\stackrel{(b)}{\leq} n \log |\mathcal{Y}| + \bar{I}(X_1^N; Y_{n+1}^N | X_2^N, S_n) + 2 \log |S| \tag{94}
 \end{aligned}$$

and

$$\begin{aligned}
 &\underline{I}(X_1^N \rightarrow Y^N | X_2^N, S_0) \\
 &\stackrel{(a)}{=} \underline{I}(X_1^N; Y^n | X_2^N, S_0) + \underline{I}(X_1^N; Y_{n+1}^N | X_2^N, Y^n, S_0) \\
 &\stackrel{(b)}{\geq} 0 + \underline{I}(X_1^N; Y_{n+1}^N | X_2^N, S_n) - 2 \log |S| - n \log |\mathcal{Y}| \tag{95}
 \end{aligned}$$

where equalities (94(a)) and (95 (a)) are due to the chain rule of mutual information and equalities (94 (b)) and (95 (b)) are

obtained by first conditioning on S_n and then using [17, Lemma 4A.1, p. 112] and the facts that $0 \leq H(Y^n) \leq n \log |\mathcal{Y}|$. In (94 (b)) we also use the fact that $\bar{I}(X_1^N; Y_{n+1}^N | X_2^N, Y^n, S_n) \leq \bar{I}(X_1^N; Y_{n+1}^N | X_2^N, S_n)$, which holds because for any FS-MAC, $H(Y_{n+1}^N | X_2^N, X_1^N, Y^n, S_n) = H(Y_{n+1}^N | X_2^N, X_1^N, S_n)$.

Entropy is uniformly continuous in distribution [52, Theorem 2.7, p. 33], i.e.,

$$\begin{aligned}
 &|\bar{H}(X) - \underline{H}(X)| \\
 &\leq - \sum_x |\bar{P}(x) - \underline{P}(x)| \log \frac{\sum_x |\bar{P}(x) - \underline{P}(x)|}{|\mathcal{X}|} \tag{96}
 \end{aligned}$$

where $\bar{H}(X)$ and $\underline{H}(X)$ denote the entropy induced by $\bar{P}(x)$ and $\underline{P}(x)$, respectively. Hence, we have

$$\begin{aligned}
 &\frac{1}{N} |\bar{I}(X_1^N; Y_{n+1}^N | X_2^N, S_n) - \underline{I}(X_1^N; Y_{n+1}^N | X_2^N, S_n)| \\
 &\leq \frac{3}{N} \Delta \log \frac{3\Delta}{|\mathcal{X}|^N |\mathcal{Y}|^N} \tag{97}
 \end{aligned}$$

where Δ is the total variation distance between

$$\bar{P}_{X_1^N, X_2^N, S_n, Y_{n+1}^N} \quad \text{and} \quad \underline{P}_{X_1^N, X_2^N, S_n, Y_{n+1}^N}.$$

Now, we show that Δ is arbitrarily small for large n . See (98) at the top of the page. Moreover, for any $\epsilon > 0$ there exists an $n_0(\epsilon)$ such that for all $N > n > n_0$, we get the expression (99) also shown at the top of the page, where in inequality (a), we used the fact that $P(s_n | x_1^N, x_2^N, s_0) = P(s_n | x_1^n, x_2^n, s_0)$ (there is no feedback), and in inequality (b), we used the indecomposable assumption, which is given in (2). Combining (94)–(95) and (97)–(99), we obtain (93), where

$$\delta(N) = \frac{1}{N} \left(4 \log |S| + n \log |\mathcal{Y}| + 6\epsilon |S| \log \frac{6\epsilon |S|}{|\mathcal{X}|^N |\mathcal{Y}|^N} \right).$$

Since ϵ can be arbitrary small for n large enough, it follows that $\delta(N)$ can be arbitrary small for N large enough. \square

IX. SUFFICIENT CONDITIONS FOR THE INNER AND OUTER BOUNDS TO COINCIDE FOR GENERAL FEEDBACK

A. Stationary and Indecomposable FS-MAC Without ISI

A stationary and indecomposable FS-MAC without ISI satisfies

$$P(y_i, s_i | x_{1i}, x_{2i}, s_{i-1}) = P(s_i | s_{i-1})P(y_i | s_{i-1}, x_{1i}, x_{2i}) \quad (100)$$

where the initial state distribution is the stationary distribution $P(s_0) = \pi(s_0)$. In words, the states are not affected by the channel inputs.

For the stationary FS-MAC, the sequence $\{\mathcal{R}_n\}$ is sup-additive. It follows from the fact that if we concatenate two input distributions $Q_{n+k} = Q_n Q_k$, then

$$I(X_1^{n+k} \rightarrow Y^{n+k} \| X_2^{n+k}) = I(X_1^n \rightarrow Y^n \| X_2^n) + I(X_{1,n+1}^{n+k} \rightarrow Y_{n+1}^{n+k} \| X_{2,n+1}^{n+k})$$

hence $(n+k)\mathcal{R}_{n+k} \supseteq n\mathcal{R}_n + k\mathcal{R}_k$. According to Lemma 5, the limit exists and is equal to

$$\lim_{n \rightarrow \infty} \mathcal{R}_n = \text{cl} \left(\bigcup_{n \geq 1} \mathcal{R}_n \right). \quad (101)$$

Next, we prove Theorem 13 that states that for a stationary and indecomposable FS-MAC without ISI, the inner bound (Theorem 9) and the outer bound (Theorem 10) coincide and therefore the capacity region is given by $\lim_{n \rightarrow \infty} \mathcal{R}_n$.

For the case that there is no feedback, it was enough to assume that the FS-MAC is indecomposable (Theorem 12), but for the case where feedback is allowed we need a stronger condition—that the states evolve independently of the input. To show that the inner and the outer regions coincide we need to show

that the influence of the initial state vanishes as $n \rightarrow \infty$. If feedback is permitted, the initial state influences the evolution of the channel state even when $n \rightarrow \infty$, through the input. Hence, in order to avoid the influence of the initial state as $n \rightarrow \infty$, we either avoid the influence of the initial state on the input by not having feedback (Theorem 12), or we need to avoid the influence of the input on the evolution of the channel state (Theorem 13).

Proof of Theorem 13: Recall that the inner bound is given in Theorem 9 as $\underline{\mathcal{R}}_N$ and the outer bound given in Theorem 10 and in Corollary 25 as $\liminf \mathcal{R}_N$. Next we show that the distance between $\underline{\mathcal{R}}_N$ and \mathcal{R}_N goes to zero which implies by Lemma 8 that both limits are equal and therefore the capacity region can be written as $\lim \mathcal{R}_N$.

Let us consider a specific input distribution denoted by $\overline{Q}(x_1^N \| z^{N-1})\overline{Q}(x_2^N \| z^{N-1})$ corresponding to the region of the outer bound \mathcal{R}_N . Let us now consider an input distribution \underline{Q} for $n+N$ inputs corresponding to the inner bound $\underline{\mathcal{R}}_N$, such that it is arbitrary for the first n inputs and then it is $\overline{Q}(x_1^N \| z^{N-1})\overline{Q}(x_2^N \| z^{N-1})$.

Now let us show that the term of the inner bound, i.e., $I_{\underline{Q}}(X_1^N \rightarrow Y^N \| X_2^{N+n}, s_0)$ and the term of the outer bound $I_{\overline{Q}}(X_1^N \rightarrow Y^N \| X_2^N)$ are arbitrarily close to each other. This can be seen in (102) shown at the bottom of the page, where

- (a) follows from Lemma 2 that states that conditioning on S_n reduces or increases the directed information by at most $\log |\mathcal{S}|$;
- (b) follows from omitting the first n elements in the sum that defines directed information;
- (c) follows from the fact that conditioning decreases entropy;
- (d) follows from the fact that the Markov chain is indecomposable, hence for any $\delta > 0$, there exists an n such that $|P(s_n | s_0) - \pi(s_n)| \leq \delta$ for any $s_0 \in \mathcal{S}$ and $s_n \in \mathcal{S}$, where $\pi(s_n)$ is the stationary distribution of s_n ;

$$\begin{aligned}
& I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} \| X_2^{N+n}, s_0) \\
& \stackrel{(a)}{\geq} I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} \| X_2^{N+n}, S_n, s_0) - \log |\mathcal{S}| \\
& \stackrel{(b)}{\geq} \sum_{i=n+1}^{N+n} H_{\underline{Q}}(Y_i | Y^{i-1}, X_2^i, S_n, s_0) - H_{\underline{Q}}(Y_i | Y^{i-1}, X_2^i, X_1^i, S_n, s_0) - \log |\mathcal{S}| \\
& \stackrel{(c)}{\geq} \sum_{i=n+1}^{N+n} H_{\underline{Q}}(Y_i | Y_{n+1}^{i-1}, X_{2,n+1}^i, S_n, s_0) - H_{\underline{Q}}(Y_i | Y_{n+1}^{i-1}, X_{2,n+1}^i, X_{1,n+1}^i, S_n, s_0) - \log |\mathcal{S}| \\
& = I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} \| X_{2,n+1}^{N+n}, S_n, s_0) - H(S_n) \\
& \stackrel{(d)}{\geq} I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} \| X_{2,n+1}^{N+n}, S_n)(1 - \delta) - \log |\mathcal{S}| \\
& \geq I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} \| X_{2,n+1}^{N+n}, S_n) - \delta(N+n) \log |\mathcal{Y}| - \log |\mathcal{S}| \\
& \stackrel{(e)}{\geq} I_{\underline{Q}}(X_{1,n+1}^{N+n} \rightarrow Y_{n+1}^{N+n} \| X_{2,n+1}^{N+n}) - \delta(N+n) \log |\mathcal{Y}| - 2 \log |\mathcal{S}| \\
& \stackrel{(f)}{\geq} I_{\overline{Q}}(X_1^N \rightarrow Y^N \| X_2^N) - \delta(N+n) \log |\mathcal{Y}| - 2 \log |\mathcal{S}| \tag{102}
\end{aligned}$$

- (e) follows from Lemma 2 that states that conditioning on S_n can differ by at most $\log |\mathcal{S}|$;
 (f) follows from the stationarity of the channel.

Dividing both sides by $N + n$ we get that for any s_0

$$\begin{aligned} & \frac{1}{N+n} I_{\underline{Q}}(X_1^{N+n} \rightarrow Y^{N+n} \| X_2^{N+n}, s_0) \\ & - \frac{1}{N+n} I_{\overline{Q}}(X_1^N \rightarrow Y^N \| X_2^N) \\ & \geq -\delta(1 + \frac{n}{N}) \log |\mathcal{Y}| - 2 \frac{\log |\mathcal{S}|}{N+n}. \end{aligned} \quad (103)$$

Inequality (103) shows that the difference between the upper bound region and the lower bound is arbitrarily small for N large enough and, hence, in the limit the regions coincide. \square

B. Indecomposable FS-MAC With Limited ISI

In this subsection, we consider a MAC inspired by Kim's point-to-point channel [21]. The conditional probability of the MAC is given by

$$\begin{aligned} & P(y_i, z_i | x_1^i, x_2^i, z_{i-1}) \\ & = P(z_i | z_{i-1}) P(y_i | z_{i-1}, x_{1,i-m}^i, x_{2,i-m}^i), \quad i = 1, 2, 3, \dots \end{aligned} \quad (104)$$

where the distribution of Z_0 is the stationary distribution $P(z_0)$, and there is also some initial distribution $P(x_{-m+1}, \dots, x_0)$.

This channel is an FS-MAC where the state at time i is $(z_{i-1}, x_{1,i-m}^{i-1}, x_{2,i-m}^{i-1})$ and therefore the inner bound (Theorem 9) and the outer bound (Theorem 10) apply to this channel. Theorem 13 also holds for this kind of channels, namely, the capacity region is given by $\lim_{n \rightarrow \infty} \mathcal{R}_n$. The proof is very similar, the only difference being that the input \underline{Q} for $n + N$ inputs is constructed slightly differently: it is arbitrary for the first $n - m$ inputs, then it is as the initial distribution $P(x_{-m+1}, \dots, x_0)$, and then it is $\overline{Q}(x_1^N \| z^{N-1}) \overline{Q}(x_2^N \| z^{N-1})$.

It is also possible to represent the channel with an alternative law, identical to the law of the channel given in (104) for $i \geq m + 1$ but for $i \leq m$ the output y_i is not influenced by the input and is, with probability 1, a particular output $\phi \in \mathcal{Y}$. Let us define \mathcal{R}_n^ϕ similarly as \mathcal{R}_n but with the alternative law for the channel. On one hand, it is clear that $\mathcal{R}_n^\phi \subseteq \mathcal{R}_n$ for all n , and on the other hand the difference between \mathcal{R}_n^ϕ and \mathcal{R}_n is at most $m \log \mathcal{Y}$ because it is possible to use the distribution of the first m inputs, $Q(x_1^m)$, to create a desired initial distribution and then use the same input as in \mathcal{R}_n . Hence

$$\lim_{n \rightarrow \infty} \mathcal{R}_n^\phi = \lim_{n \rightarrow \infty} \mathcal{R}_n. \quad (105)$$

The advantage of analyzing \mathcal{R}_n^ϕ rather than analyzing \mathcal{R}_n is that the sequence $n\mathcal{R}_n^\phi$ is sup-additive, i.e., $(n + l)\mathcal{R}_{n+l}^\phi \supseteq n\mathcal{R}_n^\phi + l\mathcal{R}_l^\phi$, and according to Lemma 5, $\lim_{n \rightarrow \infty} \mathcal{R}_n^\phi = \text{cl} \left(\bigcup_{n \geq 1} \mathcal{R}_n^\phi \right)$. Hence, we can conclude that Theorem 14 holds for this channel as well, namely, if the capacity of a stationary and indecomposable FS-MAC with limited ISI is zero without feedback then it is zero also in the presence of feedback.

X. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we have shown that directed information and causal conditioning emerge naturally in characterizing the capacity region of FS-MACs in the presence of a time-invariant feedback. The capacity region is given as a ‘‘multiletter’’ expression and it is a first step toward deriving useful concepts in communication. For instance, we use this characterization to show that for a stationary and indecomposable FS-MAC without ISI, the capacity is zero if and only if the capacity with feedback is zero. Further, we identify FS-MACs for which feedback does not enlarge the capacity region and for which source–channel separation holds.

For the point-to-point channel with feedback, recent work has shown that, for some families of channels such as unifilar channels [26], [28] or the additive Gaussian where the noise is ARMA [41], [42], the directed information formula can be computed and, further, can lead to the development of capacity-achieving coding schemes. One future direction is to use the characterizations developed in this paper to explicitly compute the capacity regions of classes of MACs with memory and feedback (other than the multiplexer followed by a point-to-point channel), and to find optimal coding schemes. An additional future direction is to use the multiletter capacity region in order to gain insight for designing good communication schemes, such as identifying weaker sufficient conditions and, ideally, fully characterize the conditions under which source–channel separation holds as for the point-to-point channel without feedback [53], [15, Ch. 3].

APPENDIX I

PROOF OF LEMMA 3

Recall that Lemma 3 states that if

$$Q(x_1^N, x_2^N \| y^{N-1}) = Q(x_1^N \| y^{N-1}) Q(x_2^N \| y^{N-1}) \quad (106)$$

then

$$\mathcal{I}(Q(x_1^N, x_2^N \| y^{N-1}); P(y^N \| x_1^N, x_2^N)) = I(X_1^N \rightarrow Y^N \| X_2^N). \quad (107)$$

Proof: The sequence of equalities in (108), given at the top of the following page, proves the lemma, where

- follows from the assumption given in (106);
- follows from the definition of the functional $\mathcal{I}(Q; P)$ given in (22);
- follows from Lemma 1 that states that $P(x_1^N, x_2^N, y^N) = Q(x_1^N, x_2^N \| y^{N-1}) P(y^N \| x_1^N, x_2^N)$ and the assumption given in (106);
- follows from the definition of directed information. \square

APPENDIX II

PROOF OF LEMMA 4

Lemma 4 states that if

$$Q(x_1^N, x_2^N \| y^{N-1}) = Q(x_1^N) Q(x_2^N) \quad (109)$$

then

$$I(X_1^N; Y^N | X_2^N) = I(X_1^N \rightarrow Y^N | X_2^N). \quad (110)$$

$$\begin{aligned}
& \mathcal{I}(Q(x_1^N, x_2^N \| y^{N-1}); P(y^N \| x_1^N, x_2^N)) \\
& \stackrel{(a)}{=} \mathcal{I}(Q(x_1^N \| y^{N-1})Q(x_2^N \| y^{N-1}); P(y^N \| x_1^N, x_2^N)) \\
& \stackrel{(b)}{=} \sum_{y^N, x_1^N, x_2^N} Q(x_1^N \| y^{N-1})Q(x_2^N \| y^{N-1})P(y^N \| x_1^N, x_2^N) \log \frac{P(y^N \| x_1^N, x_2^N)}{\sum_{x_1^N} Q(x_1^N \| y^{N-1})P(y^N \| x_1^N, x_2^N)} \\
& \stackrel{(c)}{=} \sum_{y^N, x_1^N, x_2^N} P(x_1^N, x_2^N, y^N) \log \frac{P(y^N \| x_1^N, x_2^N)}{\sum_{x_1^N} Q(x_1^N \| y^{N-1})P(y^N \| x_1^N, x_2^N)} \\
& = \mathbf{E} \left[\log \frac{P(y^N \| x_1^N, x_2^N)}{\sum_{x_1^N} Q(x_1^N \| y^{N-1}, x_2^N)P(y^N \| x_1^N, x_2^N)} \right] \\
& = \mathbf{E} \left[\log \frac{Q(x_2^N \| y^{N-1})P(y^N \| x_1^N, x_2^N)}{Q(x_2^N \| y^{N-1}) \sum_{x_1^N} Q(x_1^N \| y^{N-1}, x_2^N)P(y^N \| x_1^N, x_2^N)} \right] \\
& = \mathbf{E} \left[\log \frac{Q(x_2^N \| y^{N-1})P(y^N \| x_1^N, x_2^N)}{\sum_{x_1^N} P(y^N, x_1^N, x_2^N)} \right] \\
& = \mathbf{E} \left[\log \frac{Q(x_2^N \| y^{N-1})P(y^N \| x_1^N, x_2^N)}{P(x_2^N, y^N)} \right] \\
& = \mathbf{E} \left[\log \frac{P(y^N \| x_1^N, x_2^N)}{P(y^N \| x_2^N)} \right] \\
& \stackrel{(d)}{=} I(X_1^N \rightarrow Y^N | X_2^N)
\end{aligned} \tag{108}$$

Proof: The following sequence of equalities proves the lemma:

$$\begin{aligned}
& I(X_1^N; Y^N | X_2^N) \\
& = \mathbf{E} \left[\log \frac{P(Y^N, X_1^N | X_2^N)}{P(Y^N | X_2^N)Q(X_1^N | X_2^N)} \right] \\
& \stackrel{(a)}{=} \mathbf{E} \left[\log \frac{P(Y^N, X_1^N, X_2^N)}{P(Y^N, X_2^N)Q(X_1^N | X_2^N)} \right] \\
& \stackrel{(b)}{=} \mathbf{E} \left[\log \frac{Q(X_1^N, X_2^N | Y^{N-1})P(Y^N | X_1^N, X_2^N)}{P(Y^N | X_2^N)Q(X_2^N | Y^{N-1})Q(X_1^N | X_2^N)} \right] \\
& \stackrel{(c)}{=} \mathbf{E} \left[\log \frac{Q(X_1^N)Q(X_2^N)P(Y^N | X_1^N, X_2^N)}{P(Y^N | X_2^N)Q(X_2^N)Q(X_1^N)} \right] \\
& = \mathbf{E} \left[\log \frac{P(Y^N | X_1^N, X_2^N)}{P(Y^N | X_2^N)} \right] \\
& = I(X_1^N \rightarrow Y^N | X_2^N).
\end{aligned} \tag{111}$$

where

- (a) follows from multiplying the numerator and denominator by $P(x_2^N)$;
- (b) follows from decomposing the joint distributions $P(y^N, x_1^N, x_2^N)$ and $P(Y^N, X_2^N)$ into causal conditioning distribution by using Lemma 1;
- (c) follows from the fact that the assumption of the lemma given in (109) implies that $Q(X_1^N, X_2^N) = Q(X_1^N)Q(X_2^N)$. This can be obtained by multiplying both sides of (109) by $P(y^N | x_1^N, x_2^N)$ and then summing over all $y^N \in \mathcal{Y}^n$. \square

APPENDIX III

PROOF OF CONVERGENCE OF SUB-, SUP- ADDITIVITY

Lemma 5, given in Section III-B, states that if $\{nA_n\}$ is a sup-additive sequence of regions, then $\lim_n A_n = \bar{A} = \text{cl} \left(\bigcup_{n \geq 1} A_n \right)$.

Proof of Lemma 5: From the definitions we have $\bar{A} \supseteq \limsup A_n \supseteq \liminf A_n$. Hence, it is enough to show that $\bar{A} \subseteq \liminf A_n$.

Let \mathbf{a} be a point in \bar{A} . Then for every $\epsilon > 0$ there exists an n and a point \mathbf{a}_ϵ such that $\mathbf{a}_\epsilon \in A_n$ and $\|\mathbf{a} - \mathbf{a}_\epsilon\| \leq \epsilon$. By induction we prove that for any integer $m \geq 2$, $A_n \subseteq A_{mn}$, and this implies that $\mathbf{a}_\epsilon \in A_{mn}$. For $m = 2$ we choose $N = 2n$ and we get that

$$A_{2n} \supseteq \frac{A_n}{2} + \frac{A_n}{2} \supseteq A_n. \tag{112}$$

Now assume that it holds for $m - 1$ and let us show that it holds for m .

$$A_{mn} \supseteq \frac{A_n}{m} + \frac{(m-1)A_{(m-1)n}}{m} \supseteq \frac{A_n}{m} + \frac{(m-1)A_n}{m} \supseteq A_n. \tag{113}$$

Now, for any $N > n$, we can represent N as $mn + j$ where $0 \leq j \leq n - 1$, hence

$$A_{mn+j} \supseteq \frac{j}{mn+j} A_j + \frac{mn}{mn+j} A_{mn}. \tag{114}$$

Because \mathbf{a}_ϵ is in A_n , then it implies that it is in A_{mn} as well. Following (114) and the fact that $(0, 0) \in A_j$ we obtain

$$\frac{mn}{mn+j} \mathbf{a}_\epsilon \in A_{mn+j}. \tag{115}$$

For any $\delta > 0$ and for any $N \geq \frac{n}{\delta}$ we conclude the existence of an element in A_N for which the distance from \mathbf{a} can be upper-bounded by

$$\begin{aligned} \left\| \frac{mn}{mn+j} \mathbf{a}_\epsilon - \mathbf{a} \right\| &= \left\| \mathbf{a}_\epsilon - \mathbf{a} - \frac{j}{mn+j} \mathbf{a}_\epsilon \right\| \\ &\leq \|\mathbf{a}_\epsilon - \mathbf{a}\| + \delta \|\mathbf{a}_\epsilon\| \leq \epsilon + \delta \|\mathbf{a}_\epsilon\|. \end{aligned} \quad (116)$$

Because ϵ and δ are arbitrarily small, we can find a sequence of points $\mathbf{a}_n \in A_n$ that converges to \mathbf{a} and therefore $\mathbf{a} \in \liminf A_n$, which implies that $\bar{A} \subseteq \liminf A_n$. \square

Lemma 6, given in Section III-B, states that if $\{nA_n\}$ is a sub-additive sequence of regions, convex and closed then $\lim_n A_n = \underline{A} = \left(\bigcap_{n \geq 1} A_n \right)$.

Proof of Lemma 6: From the definitions we have $\limsup A_n \supseteq \liminf A_n \supseteq \underline{A}$. Hence, it is enough to show that $\underline{A} \supseteq \limsup A_n$. We prove the theorem by showing that if $\mathbf{a} \in \limsup A_n$, then $\mathbf{a} \in A_n$ for all $n \geq 1$.

We observe that for all $n \geq 1$

$$A_{2n} \stackrel{(a)}{\subseteq} \frac{1}{2} A_n + \frac{1}{2} \stackrel{(b)}{\subseteq} A_n \subseteq A_n, \quad (117)$$

where (a) is due to the sub-additivity of $\{A_n\}$ and (b) due to the convexity of A_n . By induction, it follows that $A_{mn} \subseteq A_n$ for all $m, n \geq 1$.

Let $\mathbf{a} \in \limsup A_n$. In the reminder of the proof, we show that $\mathbf{a} \in A_n$ for any $n \geq 1$. Since $\mathbf{a} \in \limsup A_n$, then for any $\epsilon > 0$ there exists an increasing sequence of indices (n_1, n_2, \dots) and a sequence of points in \mathbb{R}^d , $(\mathbf{a}_1, \mathbf{a}_2, \dots)$, such that $\mathbf{a}_k \in A_{n_k}$ and $\|\mathbf{a} - \mathbf{a}_k\| \leq \epsilon$ for all $k \geq 1$. This implies that for any $m \geq 1$, there exists an index $n_k > nm$, such that $\|\mathbf{a} - \mathbf{a}_k\| \leq \epsilon$. Let m' be the largest index that $nm' \leq n_k$. Hence, we have $n_k = nm' + j$, where $0 \leq j < n$. Due to the sub-additivity assumption we have

$$A_{n_k} \subseteq \frac{nm'}{nm'+j} A_{nm'} + \frac{j}{nm'+j} A_j. \quad (118)$$

Equation (118) implies that there exists a point $\mathbf{a}' \in A_{nm'}$ such that

$$\left\| \mathbf{a}_k - \frac{nm'}{nm'+j} \mathbf{a}' \right\| \leq \frac{j}{nm'+j} b \quad (119)$$

where b is the maximum norm of a point in the set A_j . Since the sequence of sets is bounded then b is finite (because of the sub-additivity, one can choose $b = \max_{\mathbf{b} \in A_1} \|\mathbf{b}\|$). Equation (119) implies that

$$\|\mathbf{a}_k - \mathbf{a}'\| \leq \frac{2j}{nm'+j} b \leq \frac{2b}{m} \quad (120)$$

and since $A_n \supseteq A_{nm'}$, then $\mathbf{a}' \in A_n$. The distance between \mathbf{a}' and \mathbf{a} is bounded by

$$\begin{aligned} \|\mathbf{a} - \mathbf{a}'\| &\leq \|\mathbf{a} - \mathbf{a}_k + \mathbf{a}_k - \mathbf{a}'\| \\ &\leq \|\mathbf{a} - \mathbf{a}_k\| + \|\mathbf{a}_k - \mathbf{a}'\| \\ &\leq \epsilon + \frac{2b}{m}, \end{aligned} \quad (121)$$

were $\epsilon > 0$ and m is an arbitrary positive integer. Finally, since A_n is closed, then (121) implies that $\mathbf{a} \in A_n$. \square

APPENDIX IV PROOF OF LEMMA 17

Lemma 17 states that

$$\begin{aligned} \max_{Q(x_1^n \| y^{n-1}) Q(x_2^n \| y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) \\ = 0 \iff \max_{Q(x_1^n) Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n) = 0 \end{aligned} \quad (122)$$

and each condition also implies that $P(y^n | x_1^n, x_2^n) = P(y^n)$ for all x_1^n, x_2^n .

Proof: Proving the direction \implies is trivial since

$$\begin{aligned} \max_{Q(x_1^n \| y^{n-1}) Q(x_2^n \| y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) \\ \geq \max_{Q(x_1^n) Q(x_2^n)} I(X_1^n, X_2^n \rightarrow Y^n). \end{aligned} \quad (123)$$

For the other direction, \Leftarrow , we have the assumption that $I(X_1^n, X_2^n \rightarrow Y^n) = 0$ for all input distributions $Q(x_1^n)Q(x_2^n)$, and in particular for the case that X_1^n and X_2^n are uniformly distributed over their alphabets. Directed information can be written as a Kullback–Leibler divergence, as shown in (124) at the bottom of the page, and by using the fact that if the Kullback–Leibler divergence $D(P\|Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}$ is zero, then $P(x) = Q(x)$ for all $x \in \mathcal{X}$, we conclude that (124) implies that $P(y^n | x_1^n, x_2^n) = P(y^n)$ for all $x_1^n \in \mathcal{X}_1^n$ and all $x_2^n \in \mathcal{X}_2^n$. It follows that

$$\begin{aligned} \max_{Q(x_1^n \| y^{n-1}) Q(x_2^n \| y^{n-1})} I(X_1^n, X_2^n \rightarrow Y^n) \\ = \max_{Q(x_1^n \| y^{n-1}) Q(x_2^n \| y^{n-1})} \mathbf{E} \left[\log \frac{P(Y^n | X_1^n, X_2^n)}{P(Y^n)} \right] \\ = \max_{Q(x_1^n \| y^{n-1}) Q(x_2^n \| y^{n-1})} \mathbf{E}[0] = 0. \end{aligned} \quad (125)$$

\square

APPENDIX V PROOF OF THEOREM 21

See equation (126) at the bottom of the page, where $P[\text{error} | m_1, m_2, x^N, y^N]$ is the error probability of decoding

$$\sum_{x_1^n, x_2^n, y^n} Q(x_1^n) Q(x_2^n) P(y^n | x_1^n, x_2^n) \log \frac{Q(x_1^n) Q(x_2^n) P(y^n | x_1^n, x_2^n)}{P(y^n) Q(x_1^n) Q(x_2^n)} = 0 \quad (124)$$

m_1 given that m_2 is decoded correctly. Throughout the remainder of the proof we fix the message m_1, m_2 . For a given tuple $(m_1, m_2, x_1^N, x_2^N, y^N)$ define the event $A_{m'_1}$, for each $m'_1 \neq m_1$, as the event that the message m'_1 is selected in such a way that $P(y^N | m'_1, m_2) > P(y^N | m, m_2)$ which is the same as $P(y^N | x_1^N, x_2^N) > P(y^N | x_1^N, x_2^N)$, where x_1^N is a shorthand notation for $x_1^N(m'_1, z^{N-1}(y^{N-1}))$ and x_2^N is a shorthand notation for $x_2^N(m_l, z_l^{N-1}(y^{N-1}))$ for $l = 1, 2$. From the definition of $A_{m'_1}$ we have (127), also shown at the bottom of the page, where $\mathbf{1}(x)$ denotes the indicator function.

$$\begin{aligned}
& P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N] \\
&= P\left(\bigcup_{m'_1 \neq m} A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N\right) \\
&\leq \min \left\{ \sum_{m'_1 \neq m} P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N), 1 \right\} \\
&\leq \left[\sum_{m'_1 \neq m_1} P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N) \right]^\rho; \quad \text{any } 0 \leq \rho \leq 1 \\
&\leq \left[(M_1 - 1) \sum_{x_1^N} Q(x_1^N | z^{N-1}) \left[\frac{P(y^N | x_1^N, x_2^N)}{P(y^N | x_1^N, x_2^N)} \right]^s \right]^\rho, \\
&\quad 0 \leq \rho \leq 1, s > 0 \quad (128)
\end{aligned}$$

where the last inequality is due to inequality (127). By substituting inequality (128) in (126) we obtain the last equation at the bottom of the page. By substituting $s = 1/(1 + \rho)$, and recognizing that x' is a dummy variable of summation, we obtain (56) and complete the proof of the bound on $\mathbf{E}[P_{e1}]$.

The proof for bounding $\mathbf{E}[P_{e2}]$ is identical to the proof that is given here for $\mathbf{E}[P_{e1}]$, up to exchanging the indices. For $\mathbf{E}[P_{e3}]$, the upper bound is identical to the case of the point-to-point channel with an input x_1^N, x_2^N , as proven in [27] where the union bound which appears here in (128) consists of $(M_1 - 1)(M_2 - 1)$ terms. \square

APPENDIX VI

PROOF OF LEMMA 26

Recall the definition of $\bar{\mathcal{R}}_n$ in (41). Lemma 26 states, that for any FS-MAC

$$(n + l)\bar{\mathcal{R}}_{n+l} \subseteq n\bar{\mathcal{R}}_n + l\bar{\mathcal{R}}_l. \quad (129)$$

Proof: We denote $N = n + l$. Let $Q(x_1^N)Q(x_2^N)P(s_0|x_1^N, x_2^N)$ be a fixed distribution. This distribution induces the distributions $Q(x_1^n)Q(x_2^n)P(s_0|x_1^n, x_2^n)$ and $Q(x_1^l)Q(x_2^l)P(s_0|x_1^l, x_2^l)$ as follows:

$$\begin{aligned}
& Q(x_1^n)Q(x_2^n)P(s_0|x_1^n, x_2^n) \\
&= \sum_{x_{n+1}^N, x_{n+1}^N} Q(x_1^N)Q(x_2^N)P(s_0|x_1^N, x_2^N) \quad (130)
\end{aligned}$$

and

$$\begin{aligned}
& Q(x_1^l)Q(x_2^l)P(s_0|x_1^l, x_2^l) \\
&= Q(x_{1,n+1}^N)Q(x_{2,n+1}^N)P(s_{n+1}|x_{n+1}^N, x_{n+1}^N) \quad (131)
\end{aligned}$$

$$= \sum_{s_0, x_1^{n+1}, x_2^{n+1}} Q(x_1^N)Q(x_2^N)P(s_0, s_{n+1}|x_1^N, x_2^N). \quad (132)$$

We establish the sub-additivity by showing that

$$\begin{aligned}
& I(X_1^N \rightarrow Y^N \| X_2^N, S_0) \\
&\leq I(X_1^n \rightarrow Y^n \| X_2^n, S_0) + I(X_1^l \rightarrow Y^l \| X_2^l, S_0) + \log |\mathcal{S}| \\
&\quad (133)
\end{aligned}$$

where $I(X_1^N \rightarrow Y^N \| X_2^N, S_0)$, $I(X_1^n \rightarrow Y^n \| X_2^n, S_0)$, and $I(X_1^l \rightarrow Y^l \| X_2^l, S_0)$ results from the distributions $Q(x_1^N)Q(x_2^N)P(s_0|x_1^N, x_2^N)$, $Q(x_1^n)Q(x_2^n)P(s_0|x_1^n, x_2^n)$, and $Q(x_1^l)Q(x_2^l)P(s_0|x_1^l, x_2^l)$, respectively, and the channel.

We have

$$\begin{aligned}
& I(X_1^N \rightarrow Y^N \| X_2^N, S_0) \\
&= I(X_1^n \rightarrow Y^n \| X_2^n, S_0) + \sum_{j=n+1}^{n+l} I(Y_j; X_1^j | Y^{j-1}, X_2^j, S_0), \\
&\quad (134)
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}[P_{e1}] &= \sum_{y^N} \sum_{x_1^N, x_2^N} P(x_1^N, x_2^N, y^N) P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N] \\
&= \sum_{y^N} \sum_{x_1^N, x_2^N} Q(x_1^N | z^{N-1}) Q(x_2^N | z^{N-1}) P(y^N | x_1^N, x_2^N) P[\text{error1} | m_1, m_2, x_1^N, x_2^N, y^N] \quad (126)
\end{aligned}$$

$$\begin{aligned}
P(A_{m'_1} | m_1, m_2, x_1^N, x_2^N, y^N) &= \sum_{x^N} Q(x^N | z^{N-1}) \cdot \mathbf{1}[P(y^N | x^N, x_2^N) > P(y^N | x_1^N, x_2^N)] \\
&\leq \sum_{x^N} Q(x^N | z^{N-1}) \left[\frac{P(y^N | x^N, x_2^N)}{P(y^N | x_1^N, x_2^N)} \right]^s; \quad \text{any } s > 0 \quad (127)
\end{aligned}$$

$$\mathbf{E}[P_{e1}] \leq (M - 1)^\rho \sum_{y^N, x_2^N} Q(x_2^N | z^{N-1}) \left[\sum_{x^N} Q(x^N | z^{N-1}) P(y^N | x_1^N, x_2^N)^{1-s\rho} \right] \left[\sum_{x_1^N} Q(x_1^N | z^{N-1}) P(y^N | x_1^N, x_2^N)^s \right]^\rho.$$

and the last term can be further bounded as shown in (135) at the bottom of the page. Inequality (a) is due to Lemma 2. Inequality (b) is because conditioning reduces entropy and because

$$P(y_j|x_1^j, x_2^j, y^{j-1}, s_n, s_0) = P(y_j|x_{1,n+1}^j, x_{2,n+1}^j, y_{n+1}^{j-1}, s_n) \quad (136)$$

for $j > n$, and any FSC with and without feedback. Equality (c) is due to the stationarity of the channel and the construction of $Q(x_1^l)Q(x_2^l)P(s_0|x_1^l, x_2^l)$, which is given in (131).

We have shown the sub-additivity property holds for the term in the first inequality of $\{\overline{\mathcal{R}}_n\}$. However, by identical arguments, one can show that the sub-additivity property holds for the terms in the last two inequalities, as well. Since the construction of $Q(x_1^n)Q(x_2^n)P(s_0|x_1^n, x_2^n)$ and $Q(x_1^l)Q(x_2^l)P(s_0|x_1^l, x_2^l)$ is identical for all cases, it implies that (88) holds. \square

APPENDIX VII

PROOF OF LEMMA 27

Recall the definition of $\underline{\mathcal{R}}_n$ and \mathcal{R}_n in (37) and (39), respectively. Lemma 27 states that

$$(n+l)\underline{\mathcal{R}}_{n+l} \supseteq n\underline{\mathcal{R}}_n + l\underline{\mathcal{R}}_l \quad (137)$$

and for an indecomposable FS-MAC without feedback $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$.

Proof of Lemma 27: We notice that if a sequence of sets is sup-additive then the sequence of the convex hull of the sets is also sup-additive. Hence, it is enough to prove the sup-additivity of the sequence $\underline{\mathcal{R}}_n$ without taking the convex hull.

The set $\underline{\mathcal{R}}_n$ is defined by three expressions that involve directed information. Because each expression is sup-additive, the whole set is sup-additive. We prove that the first expression, i.e., $\min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) - \log |\mathcal{S}|$, is sup-additive (the proofs of the sup-additivity of the other expressions are similar and therefore omitted).

Let Q_n, Q_l be the input distribution that induces the expression

$$\min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) \quad \text{and} \quad \min_{s_0} I(X_1^l \rightarrow Y^l \| X_2^l, s_0)$$

respectively. Let us choose, $Q_{n+l} = Q_n Q_l$ to be the input distribution that induces $\min_{s_0} I(X_1^{n+l} \rightarrow Y^{n+l} \| X_2^{n+l}, s_0)$. We show next, that under this choice of input distribution, we obtain the sup-additivity property given in (138) at the top of the following page, where

- (a) follows the definition of the directed information the fact that $\min_s [f(s) + g(s)] \geq \min_s f(s) + \min_s g(s)$;
- (b) follows the fact that $I(X; Y, Z) \geq I(X; Y)$;
- (c) follows Lemma 2 that states that conditioning by S_n can differ by at most $\log |\mathcal{S}|$;
- (d) follows from the stationarity of the channel and the fact that the input distribution satisfies $Q_{n+l} = Q_n Q_l$.

According to Lemma 5, since the sequence $\{\underline{\mathcal{R}}_n\}$ is sup-additive, the limit exists. In the rest of the proof we show that $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \mathcal{R}_n$. Let us denote $\underline{\mathcal{R}}'_n$ the same region as $\underline{\mathcal{R}}_n$, which is given in (37), but without taking the convex hull. We now show that $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}'_n$. In the first half of the proof we showed that $\underline{\mathcal{R}}'_n$ is sub-additive. Using this fact, we show now, that any convex combination with rational weights $(\frac{l}{k}, \frac{k-l}{k})$ of any two points from $\underline{\mathcal{R}}'_n$ is in $\underline{\mathcal{R}}'_{kn}$

$$\underline{\mathcal{R}}'_{kn} \supseteq \frac{l}{k} \underline{\mathcal{R}}'_l n + \frac{k-l}{k} \underline{\mathcal{R}}'_{(k-l)n} \supseteq \frac{l}{k} \underline{\mathcal{R}}'_n + \frac{k-l}{k} \underline{\mathcal{R}}'_n. \quad (139)$$

The left and the right inclusions in (139) are due to the sup-additivity of $\underline{\mathcal{R}}'_n$. The left inclusion is from the definition of the sup-additivity, and the right inclusion is due to the fact that sup-additivity of $\underline{\mathcal{R}}'_n$ implies that for any two positive integers m, n , $\underline{\mathcal{R}}'_{mn} \supseteq \underline{\mathcal{R}}'_n$. (This is shown by induction in (112), (113).) From (139) we can deduce that for any $\epsilon > 0$ we can find a $k(\epsilon)$ such that $\underline{\mathcal{R}}_n \subseteq \underline{\mathcal{R}}'_{nk} + \epsilon$. This fact, together with the trivial fact that $\underline{\mathcal{R}}_n \supseteq \underline{\mathcal{R}}'_n$, and the fact that the limits of both sequences exist, allow us to deduce that the limits are the same, i.e., $\lim_{n \rightarrow \infty} \underline{\mathcal{R}}_n = \lim_{n \rightarrow \infty} \underline{\mathcal{R}}'_n$.

We conclude the proof by showing that, for any input distribution $Q(x_1^n)Q(x_2^n)$, the difference between the terms in the inequalities of $\{\underline{\mathcal{R}}'_n\}$ and $\{\mathcal{R}_n\}$ goes to zero as $n \rightarrow \infty$, hence the distance between the sets of the sequences goes to zero as

$$\begin{aligned} & \sum_{j=n+1}^N I(Y_j; X_1^j | Y^{j-1}, X_2^j, S_0) \\ & \stackrel{(a)}{\leq} \sum_{j=n+1}^N I(Y_j; X_1^j | Y^{j-1}, X_2^j, S_n, S_0) + \log |\mathcal{S}| \\ & = \sum_{j=n+1}^N H(Y_j | X_2^j, Y^{j-1}, S_n, S_0) - H(Y_j | X_1^j, X_2^j, Y^{j-1}, S_n, S_0) + \log |\mathcal{S}| \\ & \stackrel{(b)}{\leq} \sum_{j=n+1}^N H(Y_j | X_{2,n+1}^j, Y_{n+1}^{j-1}, S_n) - H(Y_j | X_{1,n+1}^j, X_{2,n+1}^j, Y_{n+1}^{j-1}, S_n) + \log |\mathcal{S}| \\ & = I(X_{1,n+1}^N \rightarrow Y_{n+1}^N \| X_{2,n+1}^N, S_n) + \log |\mathcal{S}| \\ & \stackrel{(c)}{=} I(X_1^l \rightarrow Y^l \| X_2^l, S_0) + \log |\mathcal{S}|. \end{aligned} \quad (135)$$

$$\begin{aligned}
& \min_{s_0} I(X_1^{n+l} \rightarrow Y^{n+l} \| X_2^{n+l}, s_0) \\
& \stackrel{(a)}{\geq} \min_{s_0} \sum_{i=1}^n I(Y_i; X_1^i | Y^{i-1}, X_2^i, s_0) + \min_{s_0} \sum_{j=n+1}^{n+l} I(Y_j; X_1^j | Y^{j-1}, X_2^j, s_0) \\
& \stackrel{(b)}{\geq} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_2^j, s_0) \\
& \stackrel{(c)}{\geq} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_2^j, S_n, s_0) - \log |\mathcal{S}| \\
& = \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \min_{s_0} \sum_{s_n} P(s_n | s_0) \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y^{j-1}, X_{2,n+1}^j, s_n) - \log |\mathcal{S}| \\
& \geq \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \min_{s_n} \sum_{j=n+1}^{n+l} I(Y_j; X_{1,n+1}^j | Y_{n+1}^{j-1}, X_{2,n+1}^j, s_n) - \log |\mathcal{S}| \\
& \stackrel{(d)}{=} \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \min_{s_0} I(X_1^l \rightarrow Y^l \| X_2^l, s_0) - \log |\mathcal{S}|. \tag{138}
\end{aligned}$$

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{n} \left| I(X_1^n \rightarrow Y^n \| X_2^n) - \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \log |\mathcal{S}| \right| \\
& \stackrel{(a)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} \left| I(X_1^n \rightarrow Y^n \| X_2^n, S_0) - \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) + \log |\mathcal{S}| \right| + \log |\mathcal{S}| \\
& = \lim_{n \rightarrow \infty} \frac{1}{n} \left[I(X_1^n \rightarrow Y^n \| X_2^n, S_0) - \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) \right] \\
& \stackrel{(b)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} \left[\max_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) - \min_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) \right] \\
& \stackrel{(c)}{=} 0 \tag{140}
\end{aligned}$$

$n \rightarrow \infty$, and, by Lemma 8, the limits of the sequences are the same as shown in (140) at the top of the page, where

- (a) follows from Lemma 2 and the triangle inequality;
(b) follows from the fact that

$$\max_{s_0} I(X_1^n \rightarrow Y^n \| X_2^n, s_0) \geq I(X_1^n \rightarrow Y^n \| X_2^n, S_0);$$

- (c) follows from the same arguments as the proof of Theorem 4.6.4 in [17] that states this equality for indecomposable FSC without feedback (recall also that directed information equals mutual information in the absence of feedback). The extension of Theorem 4.6.4 in [17] to FS-MAC is straightforward. \square

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and anonymous referees for their comments, which significantly helped to improve the content and the organization of the paper.

REFERENCES

- [1] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory (Thakadisor, Armenian SSR, Sept. 1971)*. Budapest, Hungary: Hungarian Acad. Sci., 1973, pp. 23–52.
- [2] T. M. Cover and C. S. K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 3, pp. 292–298, May 1981.
- [3] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple access channels," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 93–95, Jan. 1982.
- [4] S. Bross and A. Lapidoth, "An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 811–833, Mar. 2005.
- [5] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of multiple access channels with state information and feedback," *IEEE Trans. Inf. Theory*, submitted for publication.
- [6] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 623–628, Jul. 1984.
- [7] J. P. M. Schalkwijk and T. Kailath, "Coding scheme for additive noise channels with feedback I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.
- [8] A. Lapidoth and M. A. Wigger, "On the Gaussian MAC with imperfect feedback," in *Proc. 24th IEEE Conv. Electrical and Electronics Engineers in Israel (IEEEI06)*, Eilat, Israel, Nov. 2006, pp. 203–207.
- [9] N. Martins and T. Weissman, "Coding schemes for additive white noise channels with feedback corrupted by quantization or bounded noise," *IEEE Trans. Inf. Theory*, submitted for publication.
- [10] N. Merhav and T. Weissman, "Coding for the feedback Gel'fand-Pinsker channel and the feedforward Wyner-Ziv source," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4207–4211, Sep. 2006.
- [11] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 605–619, May 1989.
- [12] R. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 773–785, May 1993.
- [13] G. Kramer, "Directed Information for Channels With Feedback," Ph.D. dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1998.
- [14] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

- [15] T. S. Han, *Information-Spectrum Method in Information Theory*. New York: Springer, 2003.
- [16] T. S. Han, "An information-spectrum approach to capacity theorems for the general multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2773–2795, Nov. 1998.
- [17] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [18] A. Lapidoth and I. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 973–983, May 1998.
- [19] N. T. Gaarder and J. K. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 100–102, Jan. 1975.
- [20] F. Alajaji, "Feedback does not increase the capacity of discrete channels with additive noise," *IEEE Trans. Inf. Theory*, vol. 41, no. 2, pp. 546–549, Mar. 1995.
- [21] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. IT-54, no. 4, pp. 1488–1499, Apr. 2008.
- [22] R. Ash, *Information Theory*. New York: Wiley, 1965.
- [23] J. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Information Theory Its Applic. (ISITA-90)*, Waikiki, HI, Nov. 1990, pp. 303–305.
- [24] S. Tatikonda, "Control Under Communication Constraints," Ph.D. dissertation, MIT, Cambridge, MA, 2000.
- [25] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780–789, Mar. 2005.
- [26] S. Yang, A. Kavčić, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.
- [27] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.
- [28] H. H. Permuter, P. W. Cuff, B. Van-Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.
- [29] S. C. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, Sep. 2006 [Online]. Available: arxiv.org/cs.IT/0609139, submitted for publication
- [30] B. Shrader and H. H. Permuter, "On the compound finite state channel with feedback," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 396–400.
- [31] S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 212.
- [32] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion function for general sources," in *IEEE Information Theory Workshop (ITW)*, San Antonio, TX, Oct. 2004, pp. 330–335.
- [33] R. Zamir, Y. Kochman, and U. Erez, "Achieving the Gaussian rate-distortion function by prediction," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3354–3364, Jul. 2008.
- [34] A. Rao, A. Hero, D. States, and J. Engel, "Inference of biologically relevant gene influence networks using the directed information criterion," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2006)*, Toulouse, France, May 2006.
- [35] P. Mathai, N. C. Martins, and B. Shapiro, "On the detection of gene network interconnections using directed mutual information," in *Proc. Conf. Information Theory and Application (ITA)*, San Diego, CA, Jan. 2007.
- [36] G. Salinetti and R. Wets, "On the convergence of sequences of convex sets in finite dimensions," *SIAM Rev.*, vol. 21, no. 1, pp. 18–33, 1979.
- [37] J. P. Aubin and H. Frankowska, *Set-Valued Analysis (Systems and Control: Foundations and Applications)*. Boston, MA: Birkhauser, 1990.
- [38] M. Mushkin and I. Bar-David, "Capacity and coding for the Gilbert–Elliott channel," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.
- [39] A. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.
- [40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
- [41] Y.-H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3063–3079, Jul. 2006.
- [42] Y. H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory* [Online]. Available: arxiv.org/pdf/cs.IT/0602091", submitted for publication
- [43] S. Yang, A. Kavčić, and S. Tatikonda, "On the feedback capacity of power constrained Gaussian channels with memory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 929–954, Mar. 2007.
- [44] C. E. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-2, pp. 8–19, 1956.
- [45] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, 1963.
- [46] G. Dueck, "Maximal error capacity regions are smaller than average. error capacity regions for multi-user channels," *Probl. Contr. Inf. Theory*, vol. 7, pp. 409–413, 1978.
- [47] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, Mar. 2006.
- [48] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.
- [49] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 3, pp. 226–228, Mar. 1975.
- [50] B. Nazer and M. Gastpar, "The case for structured random codes in network communication theorems," in *Proc. 2007 Information Theory Workshop (ITW)*, Lake Tahoe, CA, Sep. 2007, pp. 138–143.
- [51] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, Mar. 1985.
- [52] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [53] S. Vembu, S. Verdú, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, Jan. 1995

Haim H. Permuter (S'08–M'09) received the B.Sc. (*summa cum laude*) and M.Sc. (*summa cum laude*) degrees in electrical and computer engineering from the Ben-Gurion University, Be'er-Sheva, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2008.

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

Prof. Permuter is a recipient of the Fullbright Fellowship and the Stanford Graduate Fellowship (SGF).

Tsachy Weissman (S'99–M'02–SM'07) received the B.Sc. and Ph.D. degrees in electrical engineering from the Technion–Israel Institute of Technology, Haifa, Israel.

He has held postdoctoral appointments with the Statistics Department at Stanford University, Stanford, CA, and at Hewlett-Packard Laboratories, San Jose, CA. Currently he is with the Departments of Electrical Engineering at Stanford University and at the Technion. His research interests span information theory and its applications, and statistical signal processing. His papers thus far have focused mostly on data compression, communications, prediction, denoising, and learning. He is also inventor or coinventor of several patents in these areas and has been involved in several high-tech companies as a researcher or advisor.

Prof. Weissman received the NSF CAREER award and a Horev fellowship for leaders in Science and Technology. He has been a Robert N. Noyce Faculty Scholar of the School of Engineering at Stanford, and is a recipient of the 2006 IEEE joint IEEE IT/COM Societies best paper award.

Jun Chen (S'03–M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2003 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Josef Raviv Memorial Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Assistant Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.