On the Capacity of Indecomposable Finite-State Channels With Feedback

Ron Dabora, *Member, IEEE*, and Andrea J. Goldsmith, *Fellow, IEEE*

***Abstract*—We study the capacity of indecomposable finite-state channels (IFSCs) with feedback. It is first shown that the capacity-achieving input distribution for IFSCs with feedback is independent of the initial channel state, even though the capacity depends on the worst-case channel state. In addition, it is shown that for a large class of IFSCs for which the channel state is a deterministic function of a finite number of the most recent channel inputs and outputs, the feedback capacity depends only on the best-case channel state. This result is obtained by a novel transmission strategy whereby feedback is used to synchronize the beginning of the codeword transmission to be at the best-case channel state.**

***Index Terms*—Feedback, finite-state channels (FSCs), frame synchronization, indecomposable channels.**

I. Introduction

CAPACITY analysis for time-varying channels with memory has been the focus of considerable interest for many decades, motivated by the proliferation of wireless systems. The finite-state channel (FSC) [1] is one approach for modeling time-varying channels whereby the state of the channel varies according to a given process over a finite set of possible states. The process defines the transition probabilities between the states, which may depend on past states and past inputs. In this model, the impact of the channel memory on the output is captured by the state of the channel at the end of the previous symbol transmission.

In this work, we focus on a class of FSCs called indecomposable FSCs (IFSCs) [2, Ch. 4.6]. Loosely speaking, for IFSCs *without feedback*, the effect of the initial channel state on the state transition probabilities becomes negligible as time evolves. A consequence of the indecomposability property is that it is possible for the channel state process to go from any state to any other state in a finite number of steps.[1] IFSCs are very common in digital communication. Applications of this model include mobile and fixed wireless communications [3], digital subscriber line communications [4, Ch. 11.2.2], and magnetic recording [5]. The prevalence of IFSCs in communication systems makes understanding the fundamental limits for this class of channels of particular importance.

IFSCs without feedback were initially studied by Shannon in 1957 [6]. Specifically, Shannon analyzed the capacity of IFSCs subject to the assumptions that 1) the initial channel state is known at the transmitter, 2) the transmitter can calculate the state sequence (i.e., the current state is a function of the previous state and the current channel input), and 3) the channel output depends only on the previous state and the channel input. We note that multiple channels fall into this category. Shannon proved that for such channels, capacity is independent of the initial channel state, by showing that prefixing and appending the codewords with appropriately selected sequences can drive the channel to any desired state from any given state. In a following work, Blackwell *et al.* derived the capacity of a class of IFSCs in which the channel output at every time interval depends only on the channel state [7].

A major contribution to the study of IFSCs without feedback was made by Gallager in [2, Ch. 4.6, Ch. 5.9]. Gallager used error-exponent analysis to derive the capacity without feedback of general FSCs as well as of IFSCs, and showed that for IFSCs, knowledge of the initial channel state at the transmitter does not affect the achievable rate. In [8], an algorithm for computing the capacity of IFSCs with feedback under the assumption that the channel states can be computed from the channel inputs (e.g., channels with intersymbol interference (ISI)) was developed. In [9], Tatikonda and Mitter studied the computation of the capacity of FSCs with feedback in the framework of Markov decision problems. The work in [9] also demonstrates simplifications of the capacity computation for FSCs in which the channel state can be computed from the channel inputs and outputs. We also note that the capacity of Gaussian channels with stationary noise and feedback was obtained by Kim [10]. The result proved in [10] was originally conjectured by Yang *et al.* in [11].

Recently, the capacity of FSCs with feedback has been studied [12]. Note that the general capacity result in [12] applies to decomposable as well as to IFSCs. In particular, it was

Manuscript received September 13, 2010; revised January 05, 2012; accepted July 11, 2012. Date of publication August 17, 2012; date of current version December 19, 2012. This work was supported in part by the Defense Advanced Research Projects Agency Information Theory for Mobile Ad-Hoc Networks program under Grant 1105741-1-TFIND, in part by the Army Research Office under Multidisciplinary Research Initiative Award W911NF-05-1-0246, and in part by the National Science Foundation Center for Science of Information. R. Dabora was supported in part by the European Commission's Marie Curie International Reintegration Grants Fellowship PIRG05-GA-2009-246657 under the Seventh Framework Programme. This paper was presented in part at the 46th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2008, and in part at the 2009 IEEE Information Theory Workshop.

R. Dabora is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel (e-mail: ron@ee.bgu.ac.il).

A. J. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu).

Communicated by M. Gastpar, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2012.2214023

[1]To be precise, an indecomposable channel may have a set of inessential states (see [20, Ch. 4]) that are noncommunicating. The asymptotic probability of being in such states goes to zero and therefore they do not affect the capacity of the channel. We thus ignore such situations, which is equivalent to assuming that the channel has been operating for a sufficiently long time.

shown that feedback can increase the capacity in some cases. Capacity regions of multiuser FSCs have also been recently derived. These include finite-state multiple-access channels with and without feedback [13], as well as finite-state broadcast channels without feedback [14] and with feedback [15]. In [16], FSCs with state known at the receiver, and with channel and state feedback were studied. These works, however, did not extend Gallager's results on IFSCs without feedback (namely, that knowledge of the initial channel state at the nodes does not affect the rate) to IFSCs with feedback. The main obstacle lies in the fact that Gallager's method relies on the fact that without feedback, the initial state does not impact the (capacity-achieving) input distribution. However, when feedback is present, this property does not hold. For this reason, recent work on the capacity of point-to-point (PtP) FSCs with feedback [12] restricted the treatment of indecomposable channels to the class of finite-state Markov channels (FSMCs). FSMCs are a subset of IFSCs for which the state transitions do not depend on the channel inputs. Thus, FSMCs cannot be used to model multipath channels or other practical channels of interest within the broad class of IFSCs.

In this work, we aim at filling this gap. In particular, we study the feedback capacity of IFSCs in their most general form, and present a class of IFSCs for which the feedback capacity is achieved using a new transmission strategy we refer to as *Tx-Rx state synchronization*.

The main contributions of this paper are as follows:

1) *We show that the capacity-achieving input distribution for IFSCs with feedback is independent of the initial channel state.* This implies that when using feedback in IFSCs, the capacity-achieving input distribution can be found by fixing any arbitrary initial channel state, similar to the situation without feedback. This is surprising as feedback makes the channel decomposable, i.e., its capacity depends on the initial channel state. Therefore, this result provides insight regarding the design of the capacity-achieving scheme. Additionally, this saves a factor of $|\mathcal{S}|$ in the number of evaluations needed to compute the capacity compared to searching over all initial channel states. We expand on these ideas in Section IV.
2) *We introduce Tx-Rx state synchronization combined with random coding.* We consider the class of IFSCs in which the channel state is a deterministic function of the inputs and outputs. This class generalizes the multipath channel model and characterizes a large family of practical communication scenarios, as explained in Section V. For this class of IFSCs, we show that knowledge of the initial channel state at the transmitter does not increase the achievable rate. This is shown using the novel transmission strategy of *Tx-Rx state synchronization*, which is used to coordinate the transmitter and receiver such that the transmission of information is delayed until the channel enters the best initial state. The error events now consist of two types of errors: failing to synchronize the transmitter and receiver such that transmission of information begins at the best initial state, and failing to correctly decode the message given that synchronization is achieved. This technique provides new insights into practical code designs for channels with memory and feedback, as detailed in the discussion in Section V, right after the derivation of the results.

The rest of this paper is organized as follows: Section II presents the relevant definitions and notations. Section III reviews the capacity of IFSCs without feedback, and discusses the effect of feedback on the indecomposability property and its implications. Section IV focuses on evaluating the general capacity expression for IFSCs with feedback. Section V considers IFSCs in which the state is a deterministic function of a finite number of channel inputs and outputs, and derives their feedback capacity using Tx-Rx state synchronization. Finally, Section VI presents concluding remarks.

II. Notations and Definitions

In the following, we denote random variables (RVs) with upper case letters, e.g., X and Y, and their realizations with lower case letters, x and y. An RV X takes values in a set $\mathcal{X}$. We use $|\mathcal{X}|$ to denote the cardinality of a finite, discrete set $\mathcal{X}$, $\mathcal{X}^n$ to denote the n-fold Cartesian product of $\mathcal{X}$, and $p_X(x)$ to denote the probability mass function (pmf) of a discrete RV X on $\mathcal{X}$. For brevity, we may omit the subscript X when it is the upper case version of the realization symbol x. We use $p_{X|Y}(x|y)$ to denote the conditional pmf of X given $Y=y$. We denote vectors with boldface letters, e.g., $\mathbf{x}, \mathbf{y}$; the ith element of a vector $\mathbf{x}$ is denoted with x_i, and we use x_i^j where $i<j$ to denote the vector $(x_i, x_{i+1}, \ldots, x_{j-1}, x_j)$; x^j is short form notation for x_1^j, and $\mathbf{x} \equiv x^n$. We use $I(\cdot;\cdot)$ to denote the mutual information between two RVs, as defined in [2, Ch. 2], and $I(\cdot;\cdot)_q$ to denote the mutual information evaluated with a pmf q on the channel inputs.

Throughout this paper, we shall use the definitions of directed mutual information and causally conditioned distribution as in [13], [17], and [18, Sec. II-C]. Directed mutual information between two random vectors X^n and Y^n, given a third random vector Z^n, is defined as

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

At every time instant i, the mutual information $I(X^i;Y_i|Y^{i-1},Z^n) = H(X^i|Y^{i-1},Z^n) - H(X^i|Y^i,Z^n)$ can be viewed as the incremental "decrease in uncertainty" about the sequence X^i at the receiver due to the reception of Y_i. Summing over all time indices, we obtain the total "decrease in uncertainty" for the n symbol transmissions. Directed mutual information arises naturally in the analysis of causal channels with memory.

The probability distribution of X^n causally conditioned on Y^n is defined as

$$Q(x^n\|y^{n-1}) \triangleq \prod_{i=1}^{n} p(x_i|x^{i-1}, y^{i-1}).$$

This is a compact way for describing a signal X^n in which every symbol X_i is generated using causal knowledge of the signal Y^n, i.e., X_i is a random function of X^{i-1} and Y^{i-1}. Causal conditioning is widely used in describing input distributions under feedback scenarios. A codeword x^n generated as a causal function of y^n is referred to as a *codetree*, see [18] for a detailed explanation. We also define the shifted causally conditioned distribution

$$Q_k(x^{n_2}\|y^{n_2-1}) \triangleq \prod_{i=k}^{k+n_2-1} p(x_i|x_k^{i-1}, y_k^{i-1}).$$

Finally, we let $[a]$ denote the integer part of a real number $a \in \mathfrak{R}$, and denote $\Gamma(a) = \Pr(N < a)$, where $N \sim \mathcal{N}(0,1)$ is a real Gaussian RV with zero mean and unit variance.

Definition 1: A *finite-state channel* is defined by the triplet $\{\mathcal{X} \times \mathcal{S}, p(y,s|x,s'), \mathcal{Y} \times \mathcal{S}\}$ where X is the input symbol, Y is the output symbol, S' is the channel state at the beginning of the symbol transmission, and S is the channel state at the end of the symbol transmission. $\mathcal{S}$, $\mathcal{X}$, and $\mathcal{Y}$ are discrete alphabets of finite cardinalities. At every time $i > 0$, the pmf satisfies

$$p(y_i, s_i|x^i, y^{i-1}, s^{i-1}, s_0) = p(y_i, s_i|x_i, s_{i-1})$$

where s_0 is the initial channel state.

The pmf of a block of n transmissions, when feedback is present, is given by

$$\begin{aligned} p(y^n, s^n, x^n|s_0) &= \prod_{i=1}^{n} p(y_i, s_i, x_i|y^{i-1}, s^{i-1}, x^{i-1}, s_0) \\ &\overset{(a)}{=} \prod_{i=1}^{n} p(x_i|y^{i-1}, x^{i-1}) p(y_i, s_i|y^{i-1}, s^{i-1}, x^i, s_0) \\ &\overset{(b)}{=} \prod_{i=1}^{n} p(x_i|y^{i-1}, x^{i-1}) \prod_{i=1}^{n} p(y_i, s_i|x_i, s_{i-1}) \end{aligned}$$

where (a) follows as the transmitter is oblivious of the channel states, and (b) follows from Definition 1.

Comment 1: Some of the previous works which analyzed channels with memory and feedback can be generalized to include channel models in which the input and output sets have infinite cardinalities, e.g., [9] (see [9, Footnote 1]). However, these works assume that the channel state is fixed at the beginning of each message block. Therefore, in these works, the channel is memoryless in-between message blocks, and has memory only during transmission of the message block. For this reason, the rate expressions in these works, e.g., [9, Th. 5.1], are independent of the initial channel state s_0. Clearly, this does not accurately represents channels with memory, as memory exists at all times. The FSC model accurately models the situation by letting transmissions that took place at previous blocks affect the signals at the current block through the initial channel state s_0.

Definition 2: An (R, n) *code* for the FSC with feedback consists of a message set $\mathcal{M} \triangleq \{1, 2, \ldots, 2^{nR}\}$, a collection of mappings $(\{f_i\}_{i=1}^n, g)$ such that $f_i : \mathcal{M} \times \mathcal{Y}^{i-1} \mapsto \mathcal{X}$ is the encoding function at time i, $i = 1, 2, \ldots, n$, and $g : \mathcal{Y}^n \mapsto \mathcal{M}$, is the decoder. It is assumed that the transmitter and receiver do not know the channel states.

Definition 3: The *average probability of error* of a code of blocklength n is defined as $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0)$, where $P_e^{(n)}(s_0) \triangleq \Pr\left(g(Y^n) \neq M \middle| S_0 = s_0\right)$, and the message M is selected independently and uniformly from $\mathcal{M}$.

Definition 4: A rate R is called *achievable* for the FSC if for every $\epsilon > 0$ and $\delta > 0$, there exists an integer $n(\epsilon, \delta)$ such that for all $n > n(\epsilon, \delta)$ it is possible to construct an $(R - \delta, n)$ code with $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0) \leq \epsilon$. The *capacity* is the supremum of all achievable rates.

III. IFSCs

We begin by stating the formal definition of IFSCs introduced by Gallager:

Definition 5 ([2, Ch. 4.6]): An FSC is called indecomposable if for any $\epsilon > 0$, there exists a time index $r(\epsilon)$ such that for all $k > r(\epsilon)$, any channel states s_k, s_0, s'_0, and input sequence x^k, it holds that

$$|p(s_k|x^k, s_0) - p(s_k|x^k, s'_0)| < \epsilon. \tag{1}$$

Intuitively, this definition implies that for *indecomposable* channels, the effect of the initial channel state s_0 on the state transition probabilities decays as time evolves.

As an example of an IFSC, consider a digital communication system operating over a finite-duration ISI channel (i.e., a multipath channel). Let x_i denote the channel input (belonging to a set of finite cardinality), y_i denote the channel output after a K level A/D conversion at the receiver, and n_i denote a bandlimited, additive, white Gaussian noise sample, all at time i. Letting $\{h_j\}_{j=0}^J$ ($J > 0$, $h_0, h_J \neq 0$) be the channel coefficients, known at the receiver, and $\Upsilon_K[\cdot]$ be a quantizer with K levels, the relationship between the channel input and its output at time i is given by

$$y_i = \Upsilon_K \left[h_0 x_i + \sum_{j=1}^{J} h_j x_{i-j} + n_i \right]. \tag{2}$$

From (2), it is evident that prior to the A/D, this channel has a memory that consists of the last J channel input symbols $(x_{i-J}, \ldots, x_{i-1})$. Since quantization is memoryless, the overall memory can be represented by a finite space $\mathcal{S}$ with cardinality $|\mathcal{S}| = |\mathcal{X}|^J$, and s_{i-1}, the channel state at time $i-1$, is simply the last J channel inputs, $s_{i-1} = (x_{i-J}, \ldots, x_{i-1})$. This channel clearly satisfies Definition 5.

A. IFSCs Without Feedback

In his work on PtP-FSCs with no feedback (NFB) [2, Ch. 4.6, Ch. 5.9], Gallager showed that the capacity (of decomposable and indecomposable channels) is given by

$$C_{NFB} = \lim_{n \to \infty} \max_{p(x^n)} \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n|s_0). \tag{3}$$

This result leads to two important implications.

A1. Due to the lack of knowledge of the initial channel state at the source and destination, the capacity is determined by the initial state that results in the smallest rate for the maximizing input distribution. Thus, knowledge of the initial channel state at the transmitter and receiver *may* *increase* the achievable rate (i.e., the rate will be higher than the minimum over all initial channel states), see also [2, p. 178]. Generally speaking, knowledge of the initial state at the nodes will change the expression in (3) from "$\max_{p(x^n)} \min_{s_0 \in \mathcal{S}}$" to "$\min_{s_0 \in \mathcal{S}} \max_{p(x^n|s_0)}$," which is at least as large as the former.

A2. Evaluation of the achievable rate (3) for a fixed blocklength n requires a search over all input distributions $p(x^n)$ *and, for each of these distributions, a search over all channel states* $s_0 \in \mathcal{S}$. In other words, evaluation of the achievable rate for a given input distribution $p(x^n)$ requires evaluating $|\mathcal{S}|$ mutual information expressions.

When the FSC is indecomposable (ID), Gallager showed that the capacity (3) also satisfies (see [2, Ch. 4.6])

$$C_{\text{NFB-ID}} \overset{(3)}{=} \lim_{n\to\infty} \max_{p(x^n)} \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | s_0) \tag{4a}$$

$$= \lim_{n\to\infty} \max_{p(x^n)} \max_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n; Y^n | s_0). \tag{4b}$$

Two basic conclusions can be drawn from (4):

B1. Since the rates for the maximizing initial state and for the minimizing initial state are equal, we conclude that knowledge of the initial channel state at the transmitter and at the receiver *will not increase the achievable rate.*

B2. Evaluation of the capacity and finding the capacity-achieving distribution of IFSCs without feedback can be done *by fixing the initial state* to some arbitrary channel state $s_0' \in \mathcal{S}$ and optimizing the input distribution only for that s_0'

$$C_{\text{NFB-ID}} = \lim_{n\to\infty} \max_{p(x^n)} \frac{1}{n} I(X^n; Y^n | s_0'). \tag{5}$$

Therefore, the capacity-achieving input distribution is independent of the initial channel state. This saves a factor of $|\mathcal{S}|$ in the evaluation of (3) for large n.

Comment 2: The effects of indecomposability can be observed by comparing A1 to B1 and A2 to B2.

B. Introducing Feedback Into the Channel

In Gallager's analysis of IFSCs in [2, Th. 4.6.4], the actual condition that needs to be verified in order to show (4) is

$$|p(s_k, x^k | s_0) - p(s_k, x^k | s_0')| < \epsilon, \qquad \forall s_0, s_0', s_k, x^k. \tag{6}$$

Gallager showed that by expanding each distribution using $p(s_k, x^k | s_0) = p(x^k | s_0) p(s_k | x^k, s_0)$, and utilizing the fact that without feedback $p(x^k | s_0) = p(x^k)$, it follows from the indecomposability definition (1) that (6) is satisfied. However, when feedback is present, $p(x^k | s_0) \neq p(x^k)$. This follows from the fact that feedback lets past channel outputs affect the present output through the selection of the channel inputs. Therefore, when feedback is present, it is not possible to follow the steps in [2, Th. 4.6.4] and conclude that the initial channel state does not affect the asymptotic rate of general IFSCs in the sense of (4).

Due to this dependence on the initial state, recent works on the capacity of PtP-FSCs and multiple-access FSCs with feedback [12], [13] restricted the treatment of indecomposable channels to the class of FSMCs. The state of FSMCs evolves according to a Markov chain independent of the current channel input; hence, $p(s_i | x_i, s_{i-1}) = p(s_i | s_{i-1})$[19]. Note that the finite-ISI channel (2) is an indecomposable channel [1] but it is not an FSMC since the transition from state s_{i-1} to state s_i depends on the new channel input symbol x_i, i.e., $p(s_i | x_i, s_{i-1}) \neq p(s_i | s_{i-1})$.

This discussion provides the motivation for studying the effect of feedback for IFSCs, reported in the following sections.

C. An Upper Bound on the Feedback Capacity

Using Fano's inequality it can be shown that the feedback capacity of any FSC is upper bounded by (see [12, Th. 15])

$$C_{FB} \le \lim_{n\to\infty} \max_{Q(x^n \| y^{n-1})} \max_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n | s_0). \tag{7}$$

We refer to this expression as "lim-max-max." If the "lim-max-max" (7) is achievable, it is concluded that the achievable rate is *the same for all initial states.* This is the case for IFSCs without feedback [2, Sec. 4.6], and for stationary FSMCs with feedback [12, Sec. VII].

IV. Achieving the Feedback Capacity of IFSCs

When feedback is present, the capacity of the general FSC is given by (see [12])

$$C_{FB} = \lim_{n\to\infty} \max_{Q(x^n \| y^{n-1})} \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n | s_0). \tag{8}$$

We refer to this expression as the "lim-max-min."[2] Evaluating the capacity expression (8) requires calculating a series of achievable rates with increasing blocklength. Given a blocklength n, each achievable rate is given by (see [12, Sec. III])

$$\max_{Q(x^n \| y^{n-1})} \min_{s_0 \in \mathcal{S}} \frac{1}{n} I(X^n \to Y^n | s_0) - \frac{\log_2 |\mathcal{S}|}{n}.$$

The evaluation of the achievable rate requires searching over all codetree distributions, and for every codetree distribution, $|\mathcal{S}|$ directed mutual information expressions need to be evaluated in order to find the minimum over all initial states. In this section, we show how the evaluation of the $|\mathcal{S}|$ directed mutual information expressions for each codetree distribution can be eliminated, thus reducing the number of calculations by a factor of $|\mathcal{S}|$. This is particularly useful when one is interested in determining the achievable rate for fixed blocklength and codetree, as we discuss in more detail in the following. We next define the functions $k(n)$ and $\tilde{I}_{k(n)}(X^n \to Y^n | s_0)$ as follows: let $k(n)$ be some positive, monotone nondecreasing integer-valued function of n that satisfies

$$\lim_{n\to\infty} k(n) = \infty, \qquad \text{and} \qquad \lim_{n\to\infty} \frac{k(n)}{n} = 0 \tag{9}$$

(e.g., set $k(n) = [\sqrt{n}]$), and denote

$$\tilde{I}_{k(n)}(X^n \to Y^n | s_0) \triangleq \sum_{i=k(n)+1}^{n} I(X^i; Y_i | Y^{i-1}, s_0). \tag{10}$$

[2]Note that this result relies on the fact that the cardinality of the state space is finite. If the state space is countable, the derivation in [12], and its consequence, the capacity result in (8), which is the starting point for our derivations, do not hold. See also Comment 1 regarding the implication of general alphabets.

In the following, we let $p_{k(n)}Q_{k(n)+1}$ be a short notation for $p(x^{k(n)})Q_{k(n)+1}(x^{n-k(n)}\|y^{n-k(n)-1})$. We now state the main theorem.

Theorem 1: Fix $k(n)$ to be some positive, monotone nondecreasing integer-valued function that satisfies (9). The feedback capacity of IFSCs is given by

$$C_{FB} = \lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \min_{s_0\in\mathcal{S}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0)$$
$$= \lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \max_{s_0\in\mathcal{S}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0). \quad (11)$$

Proof: see the Appendix. ■

Comment 3: As $k(n)$ is an increasing function of n, taking n large enough, indecomposability guarantees that after $k(n)$ symbols, the channel state process arrives arbitrarily close to the asymptotic distribution of the state transitions, which is independent of the initial channel state. This is because the first $k(n)$ symbols are generated without feedback. When the use of feedback begins at the $(k(n)+1)$th symbol, the transmitter observes the same channel state distribution, irrespective of the channel state at time $i=0$. Clearly, the situation is no worse than for the worst-case initial state (i.e., the initial state that achieves the minimal rate), as the transmitter does not know the channel states. This observation eventually leads to the equality in (11). At the same time, since $\frac{k(n)}{n}\to 0$ as $n\to\infty$, the rate loss associated with such operation is asymptotically negligible. Therefore, only the two properties in (9) are needed from $k(n)$, and there is no need to optimize over $k(n)$. A class of functions that satisfies (9) is $k(n)=[n^\alpha]$ with $\alpha<1$. For example, $k(n)=[\sqrt{n}]$ belongs to this class.

From Theorem 1, we conclude the following:

Corollary 1: The feedback capacity of IFSCs can be evaluated by fixing an arbitrary initial state

$$C_{FB} = \lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0') \quad (12)$$

for an arbitrary $s_0'\in\mathcal{S}$, and $k(n)=[\sqrt{n}]$.

Proof: Clearly, for any fixed $s_0'\in\mathcal{S}$

$$\lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \min_{s_0\in\mathcal{S}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0)$$
$$\le \lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0')$$
$$\le \lim_{n\to\infty} \max_{p_{k(n)}Q_{k(n)+1}} \max_{s_0\in\mathcal{S}} \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0).$$

The corollary follows from the property of limits [21, Th. 7.2.2] as by Theorem 1 the outer limits are equal to each other and to C_{FB}. ■

A. Discussion

1) The result of (11) on the capacity of IFSCs with feedback is parallel to Gallager's result regarding the capacity of IFSCs without feedback given in (4) (cf., (4) and (11)). It is noted that equality here holds, despite the fact that, contrary to the NFB case, *the channel may not be indecomposable with feedback*, and in general $p(x^k|s_0)\neq p(x^k), \forall k$, when feedback is present. Thus, equality (11) holds even though

$$C_{FB} = \lim_{n\to\infty} \max_{Q(x^n\|y^{n-1})} \min_{s_0\in\mathcal{S}} \frac{1}{n}I(X^n \to Y^n|s_0)$$
$$\le \lim_{n\to\infty} \max_{Q(x^n\|y^{n-1})} \max_{s_0\in\mathcal{S}} \frac{1}{n}I(X^n \to Y^n|s_0). \quad (13)$$

In other words, Gallager showed that for the NFB case, the "lim-max-min" is equal to the "lim-max-max," from which it is concluded that the "lim-max-min" in the NFB case can be computed without searching over all initial states. Here, we showed in Corollary 1 that although *with* feedback, the "lim-max-min" *is not necessarily equal to the* "lim-max-max," still, the "lim-max-min" can be computed without searching over all initial channel states. Therefore, the capacity-achieving input distribution is independent of the initial channel state *although* the feedback capacity depends on the initial channel state.

2) Comparing (12) with the general expression for evaluating $C_{\rm FB}$ in (8) we observe that *the search over the input distributions in* (12) *has the same complexity* as in (8) (see also next item), but in (12) the search over all initial states is not needed. This saves a factor $|\mathcal{S}|$ in the number of calculations.

3) Note also that for ϵ arbitrarily small, one can pick some fixed k and a deterministic sequence x^k such that $\left|p(S_k|x^k,s_0)-p(S_k|x^k,s_0')\right|\le\epsilon$ for all s_0, s_0'. This fixes $p(x^k,y^k|s_0)$ and $p(S_k|x^k,s_0)$ for all $s_0\in\mathcal{S}$. Thus, for any blocklength $n+k$, an achievable rate can be calculated without searching over all initial states *and without having to recalculate the channel state distribution when feedback begins at time* $k+1$. This provides an achievable rate to within $\frac{\log_2|\mathcal{S}|}{n+k}+\epsilon\log_2|\mathcal{Y}|$ for any arbitrary initial state.

V. Deterministic-State FSCs (DS-FSCs)

In this section, we consider a class of FSCs for which the state is a deterministic function of a finite number of the most recent channel inputs and outputs. The most general form of such channels is obtained by defining the state S_i as

$$S_i = \big(X_{i-N_X+1}, X_{i-N_X+2},\ldots,X_i \quad Y_{i-N_Y+1}, Y_{i-N_Y+2},\ldots,Y_i\big) \quad (14)$$

for some integers N_X and N_Y. We refer to such FSCs as DS-FSCs. In [2, Ch. 5.9], Gallager studied FSCs for which the state is a deterministic function of the previous state and the channel output, i.e., $s_n=g(y_n,s_{n-1})$. The present model generalizes Gallager's model by letting also the current input affect the channel state. However, in this work, the state is restricted to consist only of the channel inputs and outputs while Gallager considered a general state that satisfies the deterministic relationship. We note that the model (14) applies, in particular, to linear filter channels, including channels in which the filter coefficients are taken from a countable set. This point is elaborated in the examples in Section V-A and the discussion in Section V-C.

In this section, we show that the "lim-max-min" is equal to the "lim-max-max" for indecomposable DS-FSCs *with feedback*. The proof technique introduces a new element of *Tx-Rx state synchronization* into the coding scheme: the transmitter and the receiver coordinate the transmission of information to begin at the "best" initial state, namely, at the initial channel state that maximizes the capacity upper bound (7). This implies that the capacity upper bound "lim-max-max" is achievable.

Comment 4: There is a fundamental difference between the result proved in this section and Theorem 1: Theorem 1 states that when the transmitter and the receiver are unaware of the initial state, the "lim-max-min" can be achieved by fixing the initial state. However, as "lim-max-min" $\leq$ "lim-max-max," knowledge of the initial channel state at the transmitter and at the receiver *can increase the achievable rate*. In this section, we show that for DS-FSCs, "lim-max-min" = "lim-max-max," thus knowledge of the initial state *does not increase the achievable rate*. This can affect system design considerations as in general, providing knowledge of the initial channel state to the transmitter and to the receiver is expected to increase the achievable rate, contrary to what happens here.

A. Channel Transition Function for DS-FSCs

For DS-FSCs, the channel state is given by (14) and the pmf satisfies

$$\begin{aligned} &p(y_i, s_i|x^i, y^{i-1}, s^{i-1}, s_0) \\ &= p\Big(y_i, \{x_j\}_{j=i-N_X+1}^{i}, \{y_j\}_{j=i-N_Y+1}^{i}\Big| x_i, \{x_j\}_{j=i-N_X}^{i-1}, \\ &\qquad\qquad \{y_j\}_{j=i-N_Y}^{i-1}\Big) \\ &= p\Big(y_i\Big|\{x_j\}_{j=i-N_X}^{i}, \{y_j\}_{j=i-N_Y}^{i-1}\Big). \end{aligned} \tag{15}$$

Note that both the channel inputs and the outputs affect the state transitions

$$p(s_i|y_i, x_i, s_{i-1}) \neq p(s_i|x_i, s_{i-1}). \tag{16}$$

From (15), it follows that for every set of $N_X + 1$ inputs, x^{N_X+1}, there is a $|\mathcal{Y}|^{N_Y} \times |\mathcal{Y}|^{N_Y}$ stochastic matrix mapping from the y-part of the previous state to the y-part of the current state. These matrices correspond to the pmf $p(s_i|s_{i-1}, x_i)$. For a given input sequence of length k, the state distribution after the transmission of the entire sequence x^k can be obtained by multiplying the corresponding transition matrices. Since for each input sequence of length $N_X + 1$ the transition matrix is different, the corresponding Markov chain is called inhomogeneous. In general, indecomposability can be verified by checking Gallager's conditions [2, Th. 4.6.5]. The formulation of regular matrices [20] provides an alternative (sometimes easy) way to verify that a channel is indecomposable by checking its transition matrices. We now provide two examples of indecomposable DS-FSCs:

1) Discrete Linear Channels With Rational Transfer Functions: A large class of FSCs of interest are channels characterized by rational transfer functions (followed by a quantizer):

$$y_i = \Upsilon_K\left[a_0x_i + \sum_{j=1}^{N_X} a_jx_{i-j} + \sum_{j=1}^{N_Y} b_jy_{i-j} + n_i\right] \tag{17}$$

where the a_j's and b_j's are real constants, $n_i \sim \mathcal{N}(0, \sigma_n^2)$ i.i.d., and $\Upsilon_K[\cdot]$ is a quantization function. This model represents a multipath propagation channel and also captures memory introduced by components at the receiver (e.g., filters). From [20, Th. 4.10], it is straightforward to conclude that DS-FSCs of the type (17) are weakly ergodic. Hence, DS-FSCs of the type (17) are indecomposable.

A related model to (17) is the ISI channel with autoregressive noise. This model can be obtained from the representation (17) by replacing $y_{i-N_Y}^{i-1}$ with $n_{i-N_Y}^{i-1}$. This model is applicable to storage applications, and, as long as it is possible to whiten the noise using an invertible linear filter, which is approximately a finite impulse response filter, this model can also be analyzed in the framework of DS-FSCs.

2) General FSCs With State Known at the Receiver and With State and Channel Feedback: In [16], FSCs with state known at the receiver and with state and channel feedback were studied. It is easy to show that such channels are DS-FSC, where the equivalent channel output is $\tilde{Y}_k = (Y_k, S_k)$ and the equivalent state is $\tilde{S}_k = \tilde{Y}_k$, see [16].

B. Using Tx-Rx State Synchronization With Random Coding to Achieve the Capacity of Indecomposable DS-FSCs

First, we recall that indecomposability implies that every channel state can be reached from any other channel state within a finite number of steps (see also Footnote 1). Letting $s' = (x'^{N_X}, y'^{N_Y})$ be some state of the DS-FSC, then indecomposability implies that for every initial state $s_0 \in \mathcal{S}$ there exists some input sequence $x^{L_X(s_0)}$ whose last N_X symbols equal x'^{N_X} such that $p\big(Y_{L_X(s_0)-N_Y+1}^{L_X(s_0)} = y'^{N_Y}\big|x^{L_X(s_0)}, s_0\big) > 0$. The feedback capacity of indecomposable DS-FSCs is characterized by the following theorem

Theorem 2: The feedback capacity of indecomposable DS-FSCs is given by

$$C_{DS,FB} = \lim_{n\to\infty} \max_{Q(x^n\|y^{n-1})} \max_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n|s_0). \tag{18}$$

Comment 5: The importance of indecomposability in Theorem 2 follows as it guarantees that it is possible to (eventually) transition from any initial state to the channel state $s_{0,n}^* \triangleq \big(x_*^{N_X}, y_*^{N_Y}\big)$ such that $\big(Q^*(x^n\|y^{n-1}), s_{0,n}^*\big)$ is the maximizing pair for $\max_{Q(x^n\|y^{n-1})} \max_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n|s_0)$ when n is finite and given, i.e., there exists some $k > 0$ and a sequence x^k such that

$$p(S_k = s_{0,n}^*|x^k, s_0) \geq \Delta > 0, \qquad \text{for every } s_0 \in \mathcal{S}. \tag{19}$$

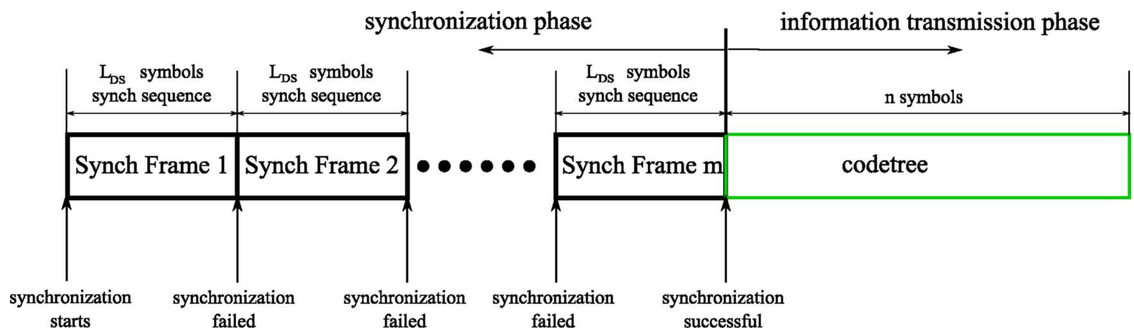

Fig. 1. A schematic description of the message transmission process. In this figure, synchronization was achieved after $m \le N_\Delta(n)$ frames.

However, in order to show that (18) is achievable, the transmission of information has to start at the maximizing initial state $s^*_{0,n}$ *despite the fact that at the end of transmission, the channel may not be at that state*. Note also that contrary to previous work (e.g., [6]) in DS-FSCs, there is no deterministic scheme that the transmitter can use to drive the channel state into $s^*_{0,n}$. The required Tx-Rx coordination is achieved by the novel *state synchronization mechanism* introduced in the proof.

Proof: Fix the blocklength $n > 2^{|\mathcal{S}|^2+1}$ and let $\big(Q^*(x^n\|y^{n-1}), s^*_{0,n}\big)$, with $s^*_{0,n} = \big(x_*^{N_X}, y_*^{N_Y}\big)$, be the input distribution and the initial channel state that maximize (18) for the value of n fixed above (i.e., when the limit in (18) is dropped). From [2, Th. 4.6.3], it follows that for some fixed $k \le 2^{|\mathcal{S}|^2}$, there exists a channel state s'' and an input sequence x^k such that, when operating without feedback, $p(S_k = s''|x^k, s_0) \ge \delta > 0$ for all initial states $s_0 \in \mathcal{S}$. Finally, we append the input sequence x^k with the appropriate $x^{L_X(s'')}$ for which $s^*_{0,n}$ can be reached from s'' with a positive probability. This sequence must consist of at most $2^{|\mathcal{S}|^2}$ symbols. The last N_X symbols of $x^{L_X(s'')}$ must be $x_*^{N_X}$. The probability of arriving at the maximizing initial state $s^*_{0,n}$ is lower bounded by

$$\begin{aligned} p(S_{k+L_X(s'')} &= s^*_{0,n}|x^{k+L_X(s'')}, s_0) \ge \Delta \\ &\ge \delta \cdot p\Big(Y^{k+L_X(s'')}_{k+L_X(s'')-N_Y+1} = y_*^{N_Y}\,\Big|\,X^{k+L_X(s'')}_{k+1} = x^{L_X(s'')}, \\ &\qquad S_k = s''\Big) > 0, \quad \forall s_0 \in \mathcal{S} \end{aligned}$$

where $p\Big(Y^{k+L_X(s'')}_{k+L_X(s'')-N_Y+1} = y_*^{N_Y}\Big|X^{k+L_X(s'')}_{k+1} = x^{L_X(s'')}, S_k = s''\Big)$ is the probability of getting from s'' to $s^*_{0,n}$ by transmitting $x^{L_X(s'')}$. Let $L_{DS} = k + L_X(s'')$ denote the overall length of the sequence constructed as above, and $x^{L_{DS}}$ denote the corresponding input sequence used for arriving from the initial channel state s_0 to $s^*_{0,n}$ with a probability lower bounded by Δ. Then, from the aforementioned discussion, it follows that

$$p\big(S_{L_{DS}} = s^*_{0,n}\big|x^{L_{DS}}, s_0\big) \ge \Delta > 0, \qquad \forall s_0 \in \mathcal{S}.$$

We refer to $x^{L_{DS}}$ as a *synchronization frame*. For a blocklength n, set $k(n)$ to be a fixed function of n that satisfies (9). For each of the 2^{nR} messages, generate a codetree of length n according to $Q^*(x^n\|y^{n-1})$. Define $N_\Delta(n) = \left\lfloor \frac{k(n)}{L_{DS}} \right\rfloor$. This is the number of $x^{L_{DS}}$ synchronization frames that can be transmitted during the first $k(n)$ symbol times.

Transmission of a message is carried out in two phases: the first phase is the state synchronization phase and the second phase is the codetree transmission phase. The entire message transmission process is schematically depicted in Fig. 1.

Phase 1: Tx-Rx State Synchronization Phase: In order to synchronize the transmitter and receiver to the best initial state, the transmitter sends $x^{L_{DS}}$. If the channel state after the transmission of an $x^{L_{DS}}$ frame is $s^*_{0,n}$, then phase two begins. The transmitter knows it is $s^*_{0,n}$ due to the feedback, and the receiver knows synchronization was achieved once it observes $y_*^{N_Y}$ at the end of a synchronization frame of L_{DS} symbols. If $s^*_{0,n}$ is not achieved, $x^{L_{DS}}$ is retransmitted. The maximum number of retransmissions is $N_\Delta(n) - 1$. If $s^*_{0,n}$ was not observed at the end of none of the $N_\Delta(n)$ synchronization frames, then message transmission has failed. Since the probability of achieving $s^*_{0,n}$ after a single frame is at least Δ, irrespective of the initial channel state, the probability of failure to achieve synchronization is upper bounded by

$$P_{e,\text{synch}} \le (1-\Delta)^{N_\Delta(n)}.$$

This probability can be made arbitrarily small by taking n large enough.

Phase 2: Information Transmission Phase: If synchronization was successful, the transmission of the codetree representing the message begins and lasts for n symbols. At the end, the receiver uses maximum-likelihood decoding [12] to extract the message.

The maximum achievable rate *given that synchronization is successful is*

$$\begin{aligned} &R_{DS,FB} \\ &\quad = \frac{n}{n+k(n)}\left(\frac{1}{n}I(X^n \to Y^n|s^*_{0,n})_{Q^*(x^n\|y^{n-1})} - \frac{\log_2|\mathcal{S}|}{n}\right) \\ &\quad = \max_{Q(x^n\|y^{n-1})}\max_{s_0\in\mathcal{S}} \frac{n}{n+k(n)}\left(\frac{1}{n}I(X^n \to Y^n|s_0) - \frac{\log_2|\mathcal{S}|}{n}\right), \end{aligned}$$

which asymptotically achieves the capacity upper bound (18):

$$\begin{aligned} \lim_{n\to\infty}\max_{Q(x^n\|y^{n-1})}\max_{s_0\in\mathcal{S}} \frac{n}{n+k(n)}\left(\frac{1}{n}I(X^n \to Y^n|s_0) - \frac{\log_2|\mathcal{S}|}{n}\right) \\ = C_{DS,FB}. \end{aligned}$$

■

Equation (8) and Theorem 2 lead to the following corollary:

Corollary 2: The feedback capacity of indecomposable DS-FSCs can be computed as

$$C_{DS,FB} = \lim_{n\to\infty} \max_{Q(x^n\|y^{n-1})} \frac{1}{n} I(X^n \to Y^n|s_0')$$

where s_0' is an arbitrary initial state, $s_0' \in \mathcal{S}$.

C. Discussion

We make the following remarks.

1) As noted earlier, in [2, Ch. 5.9], Gallager considered FSCs for which $s_n = g(y_n, s_{n-1})$. For such channels, the receiver knows the instantaneous state (assuming it knew the state at some point in the past). In the present model, the receiver does not know the instantaneous state as it also depends on the channel input. The transmitter knows the channel state due to the feedback.
2) Tx-Rx state synchronization is a specific structure of a codetree (the causally conditioned distribution $Q(x^n\|y^{n-1})$ is commonly referred to as a codetree [18]). The name "codetree" follows as the evolution of the transmitted symbols over time can conceptually be viewed as a tree whose root is the first transmitted symbol, and for every possible feedback sequence, the channel inputs correspond to a branch of the tree. During the synchronization phase, the tree consists of a single branch as the channel inputs are independent of the feedback symbols. Once synchronization is achieved, the codetree corresponding to the received feedback sequence (see, e.g., [12]) is transmitted.
3) Note that DS-FSCs may not be indecomposable with feedback.
4) As follows from Definition 1 and (14), the model (14) applies also to linear filter channels with finite memory in which the *filter coefficients belong to a finite state space*. In particular, the model represents (without any modifications) linear time-invariant filter channels (i.e., multipath channels) and also linear filter channels in which the filter coefficients vary i.i.d. between time indices, as long as the memory is bounded by some predefined size. Note that both Theorems 1 and 2 hold in this case.

   For indecomposable FSCs for which the channel state is a function of some "partial" arbitrarily varying state parameters $H_1^{L_H}$, L_H is fixed, i.e., $S_i = \left(X_{i-N_X+1}^i, Y_{i-N_Y+1}^i, H_1^{L_H}(i)\right)$, and the receiver has knowledge of $H_1^{L_H}(i) = (H_{1,i}, H_{2,i}, \ldots, H_{L_H,i})$, $H_1^{L_H}(i) \in \mathcal{H}$, $|\mathcal{H}| < \infty$; then, Theorems 1 and 2 still hold. Here, $H_1^{L_H}(i)$ may represent slowly-varying multipath coefficients that the receiver can estimate. Note that for such a filter, then during the synchronization phase in the proof of Theorem 2, the receiver need not send feedback on the Y_i symbol at every symbol time. Instead, at the end of a synchronization frame, the receiver sends a binary signal indicating only whether synchronization was achieved or not. Finally, note that if for this case the receiver cannot estimate $H_1^{L_H}(i)$ but the channel is indecomposable, then Theorem 1 still holds, but Theorem 2 does not.
5) As long as *the channel input and output sets have finite cardinalities*, e.g., due to quantization, then (14) represents also finite-memory, time-invariant linear filter channels, as well as finite-memory, i.i.d. varying, linear filter channels, with *the filter coefficients belonging to a countable space*. Then, Theorems 1 and 2 apply as well. If, however, the filter coefficients are taken from a countable set and the channel is arbitrarily varying then, as the filter coefficients must be included in the state variable to satisfy Definition 1 (see previous item), the channel is not an FSC and its analysis is not treated in this paper.
6) Note that Tx-Rx state synchronization requires that the receiver be able to know the channel state. Therefore, synchronization cannot be done if the state variable is hidden.

VI. Conclusion

In this work, we investigated the feedback capacity of IFSCs. We first showed that the capacity-achieving input distribution for IFSCs with feedback, subject to the worst-case definition of the average probability of error (Definition 3), can be found *without searching over all initial channel states*. This is surprising as the capacity expression depends on the initial channel state. This represents a saving by a factor of $|\mathcal{S}|$ in the number of calculations needed for computing the achievable rate for large blocklengths. We also explained why the capacity-achieving scheme does not need to use feedback at the beginning of transmission.

We then considered the class of IFSCs in which the state is a deterministic function of a finite number of the most recent channel inputs and outputs. For this class of channels, we showed that the upper bound on the feedback capacity, given by the "lim-max-max" of the directed mutual information (7), is achievable. The proof introduced the novel transmission technique of *Tx-Rx state synchronization* between the transmitter and the receiver. Thus, two conditions have to be satisfied for successful communication: one is the standard condition of correctly decoding the message assuming Tx-Rx state synchronization was achieved, and the second is that indeed this synchronization was achieved. This shows that for indecomposable DS-FSCs with feedback, the "lim-max-min" equals to the "lim-max-max," i.e., the worst-case and the best-case initial channel states result in the same rate. This implies that *letting the transmitter and the receiver know the initial channel state will not increase the rate*, and capacity can be evaluated without searching over all initial channel states.

Appendix
Proof of Theorem 1

1) Codebook Generation and Achievable Rate: Let $m = (m_1, m_2)$, $m_q \in \mathcal{M}_q \stackrel{\triangle}{=} \{1, 2, \ldots, 2^{n_q R}\}$, $q = 1, 2$, $n_1 = k$. Fix the pmfs $p(x^k)$ and $Q_{k+1}(x^{n_2}\|y^{n_2-1})$. For each $m_1 \in \mathcal{M}_1$, the encoder generates a codeword $x^k(m_1)$ according to the pmf $p(x^k)$. For each message $m_2 \in \mathcal{M}_2$ and feedback sequence $\left(y_{k+1}, y_{k+2}, \ldots, y_{k+n_2-1}\right)$, the encoder generates a codeword

$x^{n_2}(m_2; y_{k+1}^{k+n_2-1})$ according to $Q_{k+1}(x^{n_2}\|y^{n_2-1})$. For transmission of the message $m=(m_1,m_2)$, the encoder first outputs $x^k(m_1)$ and starting from the $(k+1)$th symbol it outputs $\left\{x_i\left(m_2; y_{k+1}^{i-1}\right)\right\}_{i=k+1}^{n}$.

As this is a special case of the scheme used in [12] to derive the capacity expression (8), we can write the achievable rate of this scheme for a given length-k initial sequence and an overall blocklength $n=k+n_2$ as

$$\underset{\sim}{\mathcal{R}}_n(k) = \max_{p(x^k)Q_{k+1}(x^{n_2}\|y^{n_2-1})} \min_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n|s_0) - \frac{\log_2|\mathcal{S}|}{n}$$
$$\overset{(a)}{=} \max_{p(x^k)Q_{k+1}(x^{n_2}\|y^{n_2-1})} \min_{s_0\in\mathcal{S}} \frac{1}{n}\Bigg(I(X^k;Y^k|s_0) + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)\Bigg) - \frac{\log_2|\mathcal{S}|}{n} \quad \text{(A.1)}$$

where in (a) we used the fact that without feedback $\sum_{i=1}^{k} H(Y_i|Y^{i-1},X^i,s_0) = H(Y^k|X^k,s_0)$, see [17].

2) Bounding the Expression in (A.1): Define first

$$\underline{\mathcal{R}}_n \triangleq \max_{Q(x^n\|y^{n-1})} \min_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n|s_0) - \frac{\log_2|\mathcal{S}|}{n}.$$

In [12], it was established that $\lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}$ exists and is finite. Let $k(n)$ be a monotone nondecreasing integer function of n that satisfies (9). This can be satisfied, for example, by setting $k(n)=[\sqrt{n}]$. We note the obvious fact that

$$\lim_{n\to\infty}\underline{\mathcal{R}}_{n-k(n)} = \lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}. \quad \text{(A.2)}$$

This follows from the fact that $n-k(n) = n\left(1-\frac{k(n)}{n}\right) \to n$ as $n\to\infty$. We now have the following lemma:

Lemma A.1: For every $0<k<n$, it holds that $\frac{n-k}{n}\underline{\mathcal{R}}_{n-k} - \frac{\log_2|\mathcal{S}|}{n} \le \underset{\sim}{\mathcal{R}}_n(k) \le \underline{\mathcal{R}}_n$.

Before proving Lemma A.1, we note that as $\lim_{n\to\infty}\frac{n-k(n)}{n}\underline{\mathcal{R}}_{n-k(n)} - \frac{\log_2|\mathcal{S}|}{n} = \lim_{n\to\infty}\underline{\mathcal{R}}_{n-k(n)}$, then from (A.2)

$$\lim_{n\to\infty}\frac{n-k(n)}{n}\underline{\mathcal{R}}_{n-k(n)} - \frac{\log_2|\mathcal{S}|}{n} = \lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}.$$

Combined with Lemma A.1, this implies that

$$\lim_{n\to\infty}\underset{\sim}{\mathcal{R}}_n(k(n)) = \lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}. \quad \text{(A.3)}$$

Finally, let s_0' minimize $\frac{k(n)\log_2|\mathcal{X}|}{n} + \frac{1}{n}\tilde{I}_{k(n)}(X^n\to Y^n|s_0)_p$ and s_0'' minimize $\frac{1}{n}I(X^{k(n)};Y^{k(n)}|s_0)_p + \frac{1}{n}\tilde{I}_{k(n)}(X^n \to Y^n|s_0)_p$. We now have for the same input distribution p

$$\min_{s_0\in\mathcal{S}}\frac{k(n)\log_2|\mathcal{X}|}{n} + \frac{1}{n}\tilde{I}_{k(n)}(X^n\to Y^n|s_0)_p \ge \frac{1}{n}\tilde{I}_{k(n)}(X^n\to Y^n|s_0'')_p \ge \min_{s_0\in\mathcal{S}}\frac{1}{n}\tilde{I}_{k(n)}(X^n\to Y^n|s_0)_p.$$

We also note that if $\forall x,\ a(x)>b(x)$, then $\max_x a(x) \ge \max_x b(x)$[3]. Therefore

$$\lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg(I(X^{k(n)};Y^{k(n)}|s_0) + \sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0)\Bigg) - \frac{\log_2|\mathcal{S}|}{n}$$
$$\ge \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0) \quad \text{(A.4)}$$

and also

$$\lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg(I(X^{k(n)};Y^{k(n)}|s_0) + \sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0)\Bigg) - \frac{\log_2|\mathcal{S}|}{n}$$
$$\le \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg(k(n)\log_2|\mathcal{X}| + \sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0)\Bigg) - \frac{\log_2|\mathcal{S}|}{n}$$
$$= \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0). \quad \text{(A.5)}$$

Combining (A.4) and (A.5), we conclude that

$$\lim_{n\to\infty}\underset{\sim}{\mathcal{R}}_n(k(n)) = \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg(I(X^{k(n)};Y^{k(n)}|s_0) + \sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0)\Bigg) - \frac{\log_2|\mathcal{S}|}{n}$$
$$= \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}\|y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k(n)+1}^{n}I(X^i;Y_i|Y^{i-1},s_0)$$

which, as $n_2=n-k(n)$, gives the first mutual information expression used in (11).

We now return to the proof of Lemma A.1: in the proof of Lemma A.1, we use [13, Lemma 2][4], restated here for convenience:

Lemma A.2 [13, Lemma 2]: Let (Z^n,U^n,S) be a joint ensemble of RVs such that $|\mathcal{S}|$ is finite. For $0<i_0<n$, it holds that

$$\left|\sum_{i=i_0}^{n}I(U^i;Z_i|Z^{i-1}) - \sum_{i=i_0}^{n}I(U^i;Z_i|Z^{i-1},S)\right| \le \log_2|\mathcal{S}|.$$

Proof of Lemma A.1: Let $(\underset{\sim}{p}_n,\underset{\sim}{s}_{0,n})$ be the pair that achieves the max-min solution for $\underset{\sim}{\mathcal{R}}_n(k)$ and let $(p^*_{n-k},s^*_{0,n-k})$ achieve the max-min solution for $\underline{\mathcal{R}}_{n-k}$. We write $\underset{\sim}{p}_n = \underset{\sim}{q}_k\underset{\sim}{Q}_{k+1}$, where $\underset{\sim}{q}_k = \underset{\sim}{q}(x^k)$ is the component without feedback of the optimal distribution for $\underset{\sim}{\mathcal{R}}_n(k)$,

[3]Otherwise for some x, $b(x)>a(x)$.

[4]Lemma A.2 is a slight variation of [13, Lemma 2].

and $\underset{\sim}{Q}_{k+1} = \underset{\sim}{Q}_{k+1}(x^{n-k}\|y^{n-k-1})$ is the component with feedback of the optimal distribution for $\underset{\sim}{\mathcal{R}}_n(k)$. We also write $p^*_{n-k} = Q^*(n-k) = Q^*(x^{n-k}\|y^{n-k-1})$, the optimal causally conditional distribution for $\underset{\sim}{\mathcal{R}}_{n-k}$. Then

$$\begin{aligned}
\underset{\sim}{\mathcal{R}}_n(k) &= \max_{p_n}\min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg[I(X^k;Y^k|s_0)_{p_n} + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{p_n}\Bigg] - \frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(a)}{\geq} \min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg[I(X^k;Y^k|s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}} + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}}\Bigg] - \frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(b)}{\geq} \min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}} - 2\frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(c)}{=} \min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k+1}^{n} I(X^i_{k+1};Y_i|Y^{i-1}_{k+1},S_k,s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}} - 2\frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(d)}{=} \min_{s_0\in\mathcal{S}}\frac{n-k}{n}\sum_{\tilde{s}_k\in\mathcal{S}}\tilde{p}(\tilde{s}_k|s_0)\times \Bigg[\frac{1}{n-k}\sum_{i=k+1}^{n} I(X^i_{k+1};Y_i|Y^{i-1}_{k+1},S_k=\tilde{s}_k)_{\underset{\sim}{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] - \frac{\log_2|\mathcal{S}|}{n}\\
&\geq \min_{s_0\in\mathcal{S}}\frac{n-k}{n}\sum_{\tilde{s}_k\in\mathcal{S}}\tilde{p}(\tilde{s}_k|s_0)\times \min_{s_k\in\mathcal{S}}\Bigg[\frac{1}{n-k}\sum_{i=k+1}^{n} I(X^i_{k+1};Y_i|Y^{i-1}_{k+1},S_k=s_k)_{\underset{\sim}{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] - \frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(e)}{=} \frac{n-k}{n}\underset{\sim}{\mathcal{R}}_{n-k} - \frac{\log_2|\mathcal{S}|}{n}
\end{aligned}$$

where in (a) we set the distribution of x^n_{k+1} for the feedback sequence y^{n-1}_{k+1} to be the optimal distribution for $\underset{\sim}{\mathcal{R}}_{n-k}$, i.e., $\underset{\sim}{Q}^*_{k+1} = Q^*(n-k)$ with the appropriate index shift; (b) follows from Lemma A.2 and the relationship $\sum_{i=k+1}^n I(X^i;Y_i|Y^{i-1},s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}} \geq \min_{s_0\in\mathcal{S}}\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,s_0)_{\underset{\sim}{q}_k \underset{\sim}{Q}^*_{k+1}} - \log_2|\mathcal{S}|$; (c) follows from

$$\begin{aligned}
H(Y_i|X^i,Y^{i-1},S_k,s_0) &= \sum_{\mathcal{X}^i\times\mathcal{Y}^{i-1}\times\mathcal{S}} p(x^i,y^{i-1},s_k|s_0)H(Y_i|x^i,y^{i-1},s_k,s_0)\\
&= \sum_{\mathcal{X}^{i-k}\times\mathcal{Y}^{i-k-1}\times\mathcal{S}} p(x^i_{k+1},y^{i-1}_{k+1},s_k|s_0)H(Y_i|x^i_{k+1},y^{i-1}_{k+1},s_k)\\
&= H(Y_i|X^i_{k+1},Y^{i-1}_{k+1},S_k,s_0)
\end{aligned}$$

and recalling that by the structure of $\underset{\sim}{p}_n$, when $i \geq k+1$, then $p(x_i|x^{i-1},y^{i-1}) = p(x_i|x^{i-1}_{k+1},y^{i-1}_{k+1})$; (d) follows from the structure of the FSC and the fact that only feedback from Y^n_{k+1} is used. Finally, (e) is because in $\underset{\sim}{\mathcal{R}}_{n-k}$ the minimizing initial state is used.

The inequality $\underset{\sim}{\mathcal{R}}_n(k) \leq \underset{\sim}{\mathcal{R}}_n$ is obvious: let $(\underset{\sim}{p}_n, \underset{\sim}{s}_{0,n})$ be the pmf-state pair that optimizes $\underset{\sim}{\mathcal{R}}_n(k)$. Then, as for $\underset{\sim}{\mathcal{R}}_n$, the search for the maximizing probability distribution is carried over a larger class of input distributions which includes $\underset{\sim}{p}_n$, the rate $\underset{\sim}{\mathcal{R}}_n$ cannot be less than $\underset{\sim}{\mathcal{R}}_n(k)$. ■

3) Asymptotic Expression for (A.1): We prove the following lemma:

Lemma A.3: Let $p_n = q_k Q_{k+1}$, where $k < n$, $q_k = q(x^k)$ is a pmf over x^k and $Q_{k+1} \overset{\triangle}{=} Q_{k+1}(x^{n-k}\|y^{n-k-1})$, and define

$$\widetilde{\mathcal{R}}_n(k) \overset{\triangle}{=} \max_{p_n}\max_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg[I(X^k;Y^k|s_0)_{p_n} + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{p_n}\Bigg].$$

For every $\epsilon > 0$, there exists k large enough such that $\lim_{n\to\infty}\left|\widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k)\right| \leq \epsilon|\mathcal{S}|\log_2|\mathcal{Y}|$.

Proof: Define first

$$\underset{\sim}{\mathcal{R}}_n(k;p_n) \overset{\triangle}{=} \min_{s_0\in\mathcal{S}}\frac{1}{n}\Bigg[I(X^k;Y^k|s_0)_{p_n} + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{p_n}\Bigg].$$

Clearly, $\underset{\sim}{\mathcal{R}}_n(k) \geq \underset{\sim}{\mathcal{R}}_n(k;p_n)$. Let $(\tilde{p}_n, \tilde{s}_{0,n})$ be the maximizing pair for $\widetilde{\mathcal{R}}_n(k)$ and let $\underset{\sim}{s}_{0,n}$ minimize $\underset{\sim}{\mathcal{R}}_n(k;\tilde{p}_n)$. Then, with $\tilde{p}_n = \tilde{q}_k\tilde{Q}_{k+1}$ (where $\tilde{q}_k = \tilde{q}(x^k)$ and $\tilde{Q}_{k+1} = \tilde{Q}_{k+1}(x^{n-k}\|y^{n-k-1})$ denote, respectively, the NFB and the causally conditional components of the pmf $\tilde{p}_n$, parallel to p_n in the statement of the lemma), we have

$$\begin{aligned}
&\left|\widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k)\right|\\
&\leq \left|\widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k;\tilde{p}_n)\right|\\
&\leq 2\frac{\log_2|\mathcal{S}|}{n} + \Bigg|\frac{1}{n}\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,\tilde{s}_{0,n})_{\tilde{p}_n} - \frac{1}{n}\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,\underset{\sim}{s}_{0,n})_{\tilde{p}_n}\Bigg| + \frac{k}{n}\log_2|\mathcal{X}|\\
&\leq \frac{k}{n}\log_2|\mathcal{X}| + 2\frac{\log_2|\mathcal{S}|}{n} + \Bigg|\frac{1}{n}\sum_{s_k\in\mathcal{S}}\left|\tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n})\right|\times \sum_{i=k+1}^{n} I(X^i_{k+1};Y_i|Y^{i-1}_{k+1},s_k)_{\tilde{Q}_{k+1}}\Bigg|\\
&\leq \sum_{s_k\in\mathcal{S}}\left|\tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n})\right|\frac{n-k}{n}\log_2|\mathcal{Y}| + \frac{k}{n}\log_2|\mathcal{X}| + 2\frac{\log_2|\mathcal{S}|}{n}\\
&\overset{(a)}{\leq} |\mathcal{S}|\epsilon\frac{n-k}{n}\log_2|\mathcal{Y}| + \frac{k}{n}\log_2|\mathcal{X}| + 2\frac{\log_2|\mathcal{S}|}{n}\\
&\overset{n\to\infty}{\longrightarrow} \epsilon|\mathcal{S}|\log_2|\mathcal{Y}|
\end{aligned}$$

where (a) follows from $\left|\tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n})\right| \leq \sum_{\mathcal{X}^k} \tilde{q}(x^k)\left|\tilde{p}(s_k|x^k,\tilde{s}_{0,n}) - \tilde{p}(s_k|x^k,\underset{\sim}{s}_{0,n})\right| \leq \epsilon$. ■

4) Combining Lemmas A.1 and A.3: From Lemma A.3, we conclude that $\lim_{n\to\infty} \underset{\sim}{\mathcal{R}}_n(k(n)) = \lim_{n\to\infty} \widetilde{\mathcal{R}}_n(k(n))$. Since from (A.3) $\lim_{n\to\infty} \underset{\sim}{\mathcal{R}}_n(k(n)) = \lim_{n\to\infty} \underline{\mathcal{R}}_n$, we conclude that C_{FB} can be evaluated as in (11).

REFERENCES

[1] B. McMillan, "The basic theorems of information theory," *Ann. Math. Statist.*, vol. 24, no. 2, pp. 196–219, Jun. 1953.

[2] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[3] G. G. Raleigh and M. M. Cioffi, "Spatio-temporal coding for wireless channels," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[4] P. Golden, H. Dedieu, and K. S. Jacobsen, Eds., *Fundamentals of DSL Technology*. Auerbach Publications, 2004.

[5] K. A. S. Immink, "Coding techniques for the noisy magnetic recording channel: A state-of-the-art report," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 413–419, May 1989.

[6] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Control*, vol. 1, pp. 6–25, Sep. 1957.

[7] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Statist.*, vol. 29, no. 4, pp. 1209–1220, 1958.

[8] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.

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

[10] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 57–85, Jan. 2010.

[11] S. Yang, A. Kavcic, and S. Tatikonda, "On the feedback capacity of power-constrained Gaussian noise channels with memory," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 929–954, Mar. 2007.

[12] 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.

[13] H. H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455–2477, Jun. 2009.

[14] R. Dabora and A. J. Goldsmith, "The capacity region of the degraded finite-state broadcast channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1828–1851, Apr. 2010.

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

[16] 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.

[17] J. L. Massey, "Causality, feedback and directed information," in *Proc. Int. Symp. Inf. Theory Appl.*, Waikiki, HI, Nov. 1990, pp. 303–305.

[18] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[19] A. J. Goldsmith and P. 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.

[20] E. Seneta, *Non-Negative Matrices and Markov Chains*. New York: Springer-Verlag, 1981.

[21] R. A. Rankin, *An Introduction to Mathematical Analysis*. New York: Dover, 2007.

[22] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," in *Proc. Amer. Math. Soc.*, vol. 14, no. 5, pp. 733–737, Oct. 1963.

Ron Dabora (M'03) received his B.Sc. and M.Sc. degrees in 1994 and 2000, respectively, from Tel-Aviv University, and his Ph.D. degree in 2007 from Cornell University, all in Electrical Engineering. From 1994 to 2000 he worked as an engineer at the Ministry of Defense of Israel, and from 2000 to 2003, he was with the Algorithms Group at Millimetrix Broadband Networks, Israel. From 2007 to 2009 he was a postdoctoral researcher at the Department of Electrical Engineering, Stanford University. Since 2009 he is an assistant professor at the Department of Electrical and Computer Engineering, Ben-Gurion University, Israel. His research interests include network information theory, wireless communications and power line communications. He currently serves as an associate editor for the IEEE SIGNAL PROCESSING LETTERS.

Andrea J. Goldsmith (S'90–M'93–SM'99–F'05) is a professor of Electrical Engineering at Stanford University, and was previously an assistant professor of Electrical Engineering at Caltech. She co-founded Accelera Mobile Broadband, Inc. and Quantenna Communications Inc., and has previously held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. Dr. Goldsmith is a Fellow of the IEEE and of Stanford, and she has received several awards for her work, including the IEEE Communications Society and Information Theory Society joint paper award, the National Academy of Engineering Gilbreth Lecture Award, the IEEE Wireless Communications Technical Committee Recognition Award, the Alfred P. Sloan Fellowship, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. She is author of the book "Wireless Communications" and co-author of the books "MIMO Wireless Communications" and "Principles of Cognitive Radio," all published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from U.C. Berkeley.

Dr. Goldsmith has served as associate editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and as editor for the *Journal on Foundations and Trends in Communications and Information Theory and in Networks*. She previously served as an editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and for the IEEE WIRELESS COMMUNICATIONS MAGAZINE. Dr. Goldsmith participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She is a Distinguished Lecturer for both societies, served as the President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory Society, and currently chairs the Emerging Technology Committee and is a member of the Strategic Planning Committee in the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award, served as Chair of its Faculty Senate, and currently serves on its Faculty Senate and on its Budget Group.