On the Capacity of Indecomposable Finite-State Channels with Feedback

Ron Dabora and Andrea Goldsmith

Wireless Systems Lab
Department of Electrical Engineering
Stanford University

***Abstract*—We study the capacity of indecomposable finite-state channels (FSCs) with feedback. In this class of channels, the effect of the initial state on the state transition probabilities for every *given* input sequence becomes negligible as time evolves. It is known that for indecomposable FSCs without feedback the capacity is independent of the initial state. Similar results were obtained for indecomposable finite-state multiple access channels and indecomposable degraded finite-state broadcast channels. However, when feedback is present, such a result does not exist except for FSCs without intersymbol interference (ISI). In this paper we show that the capacity-achieving distribution of indecomposable FSCs with feedback can be computed without minimizing over all initial channel states.**

I. Introduction

Consider a digital communication system operating over a finite-memory 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 and n_i denote bandlimited, additive, white Gaussian noise at time i. Letting $\{h_j\}_{j=0}^{J}$ be the channel coefficients, the relationship between the channel input and its output (after A/D conversion at the receiver) is given by

$$y_i = Q_K\left[h_0 x_i + \sum_{j=1}^{J} h_j x_{i-j} + n_i\right], \qquad (1)$$

where $Q_K[\cdot]$ is a quantizer with K levels, K even,

$$Q_K[r] = \begin{cases} -\frac{K}{2} & , r \le -\frac{K}{2}+1 \\ -\delta & , -\delta < r \le -\delta + 1 \\ \delta & , \delta - 1 < r \le \delta \\ \frac{K}{2} & , r > \frac{K}{2} - 1 \end{cases}, \quad \delta = 1, 2, ..., \frac{K}{2} - 1.$$

This channel is depicted in Figure 1. As is evident from Equation (1), before the A/D (point A in Figure 1) this channel has a memory that consists of the last J channel input symbols $(x_{i-J}, ..., x_{i-1})$. Since quantization is a memoryless operation, the overall memory can be represented by a finite state 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, $s_{i-1} = (x_{i-J}, ..., x_{i-1})$. This channel belongs to a special class of finite-state channels called indecomposable FSCs. Before reviewing the known results on indecomposable FSCs we present the formal definition of this class of channels:

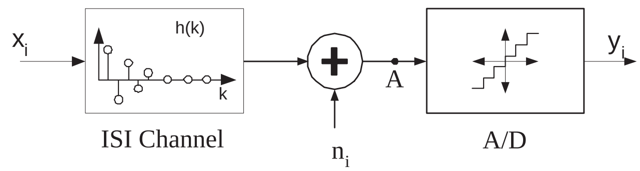

Fig. 1. A schematic description of a multipath digital communication channel. x_i is the channel input symbol and y_i is the sampled output at the receiver, at time i.

Definition 1. ([2, Equation 4.6.26]) *A finite-state channel is called indecomposable if for any $\epsilon > 0$ there exists a time index $K(\epsilon)$ such that for all $k > K(\epsilon)$ and any channel states s_k, s_0, s_0' and input sequence x^k it holds that*

$$|p(s_k|x^k, s_0) - p(s_k|x^k, s_0')| < \epsilon. \qquad (2)$$

As indecomposable FSCs are frequently encountered in communication systems, understanding the fundamental limits for this class of channels is of particular importance. In his work on point-to-point (PtP) FSCs without feedback (NFB) [2, Chapter 4.6], Gallager showed that for indecomposable FSCs the initial state does not affect capacity. To place the current work in the right context, we briefly discuss this result. Gallager first showed that the capacity of PtP-FSCs (non-indecomposable and indecomposable channels) is given by [2, Section 5.9][1]

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

If the channel is indecomposable (ID), Gallager further showed that [2, Section 4.6]

$$\begin{aligned} C_{NFB-ID} &= \lim_{n\to\infty} \max_{p(x^n)} \min_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n; Y^n|s_0) \\ &= \lim_{n\to\infty} \max_{p(x^n)} \max_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n; Y^n|s_0). \qquad (4) \end{aligned}$$

Thus, when the channel is indecomposable, lack of knowledge of the initial state at the receiver does not affect the maximum rate that can be achieved. Equation (4) also implies that

The authors are with the Wireless Systems Lab, Department of Electrical Engineering, Stanford University, Stanford, CA 94305. Email: {ron,andrea}@wsl.stanford.edu. This work was supported in part by the DARPA ITMANET program under grant 1105741-1-TFIND and the ARO under MURI award W911NF-05-1-0246.

[1] Subject to the definition of the probability of error in Definition 4.

evaluating the capacity of ID-FSC without feedback can be done by fixing the initial state to some arbitrary channel state and optimize only the input distribution (i.e. not perform the '$\min_{s_0 \in \mathcal{S}}$' part of the rate expression (3)). In Gallager's analysis of indecomposable channels in [2, Theorem 4.6.4], the actual condition that needed to be verified in order to show that the capacity of indecomposable FSCs is the same for all initial states is

$$|p(s_k, x^k|s_0) - p(s_k, x^k|s_0')| < \epsilon, \qquad \forall s_0, s_0', s_k, x^k. \quad (5)$$

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 that condition (2) is enough to satisfy (5). However, when feedback is present, $p(x^k|s_0) \neq p(x^k)$. This follows from the fact that feedback introduces memory into the channel by letting 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, Theorem 4.6.4] and conclude that the initial channel state does not affect the asymptotic performance of general indecomposable FSCs as defined in Definition 1.

For this reason, recent work on the capacity of point-to-point and multiple-access FSCs with feedback [3], [4] restricted the treatment of indecomposable channels to the class of Markov channels. These channels satisfy $p(s_i|x_i, s_{i-1}) = p(s_i|s_{i-1})$. Thus, the channel input does not affect the state transition when the previous state is given. Note that the finite-ISI channel (1) is an indecomposable channel [1] but it is not a Markov channel since the transition from state s_{i-1} to state s_i depends on the new channel input symbol x_i:

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

However, general indecomposable channels are of significant importance due to their prevalence in wireless and wired digital communications. In this work we first discuss the dependence of the definition of indecomposable FSCs on the existence of feedback and show that a channel may be indecomposable without feedback but non-indecomposable with feedback. We then show that when using feedback in indecomposable channels, the capacity achieving distribution can be found without searching over all initial channel states. This holds even if the channel satisfy Definition 1 only without feedback. The question whether for indecomposable FSCs with feedback the 'lim-max-min' expression equals the 'lim-max-max' in as in (4) is still open.

The rest of this paper is organized as follows: Section II recalls the relevant definitions and notations and presents an example of a channel that is indecomposable without feedback but non-indecomposable with feedback, Section III includes the main theorem and Section IV presents concluding remarks. The proof of the main theorem is relegated to Appendix A.

II. Definitions and Notations

First, a word about notation. In the following we denote random variables with upper case letters, e.g. X, Y, and their realizations with lower case letters x, y. A random variable (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 (p.m.f.) of a discrete RV X on $\mathcal{X}$. For brevity we may omit the subscript X when it is obvious from the context. We use $p_{X|Y}(x|y)$ to denote the conditional p.m.f. of X given Y. We denote vectors with boldface letters, e.g. $\mathbf{x}$, $\mathbf{y}$; the i'th element of a vector $\mathbf{x}$ is denoted with x_i and we use x_i^j where $i < j$ to denote the vector $(x_i, x_{i+1}, ..., x_{j-1}, x_j)$; x^j is short form notation for x_1^j, and $\mathbf{x} \equiv x^n$. A vector of random variables is denoted by $\mathbf{X} \equiv X^n$, and similarly we define $X_i^j \triangleq (X_i, X_{i+1}, ..., X_{j-1}, X_j)$ for $i < j$. We use $H(\cdot)$ to denote the entropy of a discrete random variable and $I(\cdot;\cdot)$ to denote the mutual information between two random variables, as defined in [5, Chapter 2]. $I(\cdot;\cdot)_q$ denotes the mutual information evaluated with a p.m.f. q on the channel inputs. We also define the following quantities, see also [6, Section II.C], [4], [7]:

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

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

$$Q_{k+1}(x^{n_2}||y^{n_2-1}) \triangleq \prod_{i=k+1}^{k+n_2} p(x_i|x_{k+1}^{i-1}, y_{k+1}^{i-1}).$$

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

Definition 2. *The* 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 end of the previous symbol transmission and* S *is the channel state at the end of the current symbol transmission.* $\mathcal{S}$, $\mathcal{X}$ *and* $\mathcal{Y}$ *are discrete alphabets of finite cardinalities.*

The p.m.f of a block of n transmissions 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 s_0 is the initial channel state. Here, (a) is because the transmitter is oblivious of the channel states and (b) captures the fact that given S_{i-1}, the symbols at time i are independent of the past.

Definition 3. *An* (R, n) deterministic code *for the FSC with feedback consists of a message set* $\mathcal{M} = \{1, 2, ..., 2^{nR}\}$*, a collection of mappings* $(\{f_i\}_{i=1}^n, g)$ *such that*

$$f_i : \mathcal{M} \times \mathcal{Y}^{i-1} \mapsto \mathcal{X} \quad (6)$$

is the encoding function at time i*,* $i = 1, 2, ..., n$*, and*

$$g : \mathcal{Y}^n \mapsto \mathcal{M},$$

is the decoder. Note that we assume no knowledge of the states at the transmitter and receiver.

Definition 4. *The* average probability of error *of a code of blocklength n is given by* $\max_{s_0 \in \mathcal{S}} P_e^{(n)}(s_0)$*, where*

$$P_e^{(n)}(s_0) = \Pr\left(g(Y^n) \neq M | s_0\right)$$

and the message M is selected independently and uniformly from $\mathcal{M}$.

We now revisit the Gallager's definition of indecomposable channels. Consider the following model for the received signal at time k:

$$y_k = Q_2\Big[x_k + a y_{k-1} + n_k\Big], \tag{7}$$

where $x_k, y_k \in \{-1, 1\}, a = 0.1, n_k \sim \mathcal{N}(0,1)$. This models a receiver with a one-tap equalizer operating in steady-state at low SNR[2]. The channel (7) is clearly a FSC with $S_i = Y_i$.

We note that the state transition matrix for this channel depends on the input sequence x^k, thus the state transitions are represented by a non-homogeneous Markov chain. Denoting with P_1 and $P_{(-1)}$ the transition matrices when $x = 1$ and $x = -1$ respectively we write

$$P_1 = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \quad P_{(-1)} = \begin{bmatrix} 1-q & q \\ p & 1-p \end{bmatrix}.$$

Here P_{ij} denotes the transition probability from state i to state j. When $i = 1$ the state is $y' = -1$ and when $i = 2$ the state is $y' = 1$. The values of p and q can be obtained as follows:

$$\begin{aligned} p &= \Pr(y_k = 1 | x_k = 1, y_{k-1} = -1) \\ &= \Pr\left(x_k + a y_{k-1} + n_k > 0 \middle| x_k = 1, y_{k-1} = -1\right) \\ &= \Pr\left(1 - a + n_k > 0\right) \\ &= \Pr\left(n_k > -1 + a\right) \\ &= \Pr\left(n_k < 1 - a\right) \\ &= \mathbb{Q}(1-a) \\ &= 0.8159, \end{aligned}$$

$$\begin{aligned} q &= \Pr(y_k = -1 | x_k = 1, y_{k-1} = 1) \\ &= \Pr\left(x_k + a y_{k-1} + n_k < 0 \middle| x_k = 1, y_{k-1} = 1\right) \\ &= \Pr\left(1 + a + n_k < 0\right) \\ &= \Pr\left(n_k < -1 - a\right) \\ &= \mathbb{Q}(-1-a) \\ &= 0.1357. \end{aligned}$$

It is easy to see that each of the matrices P_1, P_{-1} has a unique (but different) limit distribution:

$$\pi_1 = \left[\frac{q}{p+q}, \frac{p}{p+q}\right], \qquad \pi_{(-1)} = \left[\frac{p}{p+q}, \frac{q}{p+q}\right]$$

Without feedback, it can be shown that this channel is indecomposable: for k large enough

$$\left|p_{NFB}\big(y_k = 1 \big| x^k, y_0'\big) - p_{NFB}\big(y_k = 1 \big| x^k, y_0''\big)\right| \leq \epsilon \tag{8}$$

[2]Note that for simplicity of the exposition we assume no ISI. Thus, strictly speaking there is no need for an equalizer. In general the channel will include ISI as in Equation (1) and an equalizer as in Equation (7).

where $\epsilon > 0$ is an arbitrary constant, x^k is an arbitrary input sequence, and y_0', y_0'' are the initial channel states. Therefore, the effect of the initial state becomes negligible as time evolves.

We now show that when feedback is present the channel (7) is non-indecomposable. Assume that the feedback scheme sets $X_i = Y_{i-1}$. Clearly, if $x_k = 1$ then $p_{FB}\big(y_k = 1, y_{k-1} = -1 \big| x_k = 1, x^{k-1}, s_0\big) = 0$. Let $\{-1\}^{k-1}$ denote a vector of length $k-1$ with all elements equal to -1 and consider the same input sequence $x^k = \big(\{-1\}^{k-1}, x_k = 1\big)$ for the two cases, with and without feedback. With feedback we obtain

$$\begin{aligned} &p_{FB}\big(y_k = 1 \big| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\big) \\ &= \sum_{y_{k-1} \in \{-1,1\}} \Pr\left(y_k = 1, y_{k-1} \middle| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\right) \\ &= \Pr\left(y_k = 1, y_{k-1} = 1 \middle| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\right) \\ &= \Pr\left(y_{k-1} = 1 \middle| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\right) \\ &\quad \times \Pr\left(y_k = 1 \middle| y_{k-1} = 1, x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\right) \\ &= \Pr\left(y_k = 1 \middle| y_{k-1} = 1, x_k = 1\right) \\ &= \Pr\left(x_k + a y_{k-1} + n_k > 0 \middle| y_{k-1} = 1, x_k = 1\right) \\ &= \Pr\left(n_k > -1 - a\right) \\ &= \Pr\left(n_k < 1 + a\right) \\ &= \mathbb{Q}\big(1 + a\big) = 1 - q \end{aligned}$$

Consider the case without feedback. From the structure of the channel we have that for k large enough we can write the distribution of y_k as

$$\begin{aligned} p_{NFB}\big(y_k \big| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\big) &= [0,1] \cdot P_{(-1)}^{k-1} \cdot P_1 \\ &\approx \pi_{(-1)} \cdot P_1 \\ &= \left[\frac{p}{p+q}, \frac{q}{p+q}\right] \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \end{aligned}$$

thus

$$\begin{aligned} \Pr_{NFB}\big(y_k = 1 \big| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\big) \\ = \frac{p^2 + q - q^2}{p+q}. \end{aligned}$$

Note that unless $\frac{p^2+q-q^2}{p+q} = 1 - q$, which does not hold for this example[3], then taking larger blocklengths k we cannot make the difference

$$\begin{aligned} \Big| p_{NFB}\big(y_k = 1 \big| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\big) \\ - p_{FB}\big(y_k = 1 \big| x_k = 1, x^{k-1} = \{-1\}^{k-1}, y_0 = 1\big) \Big| \end{aligned}$$

arbitrarily small since it is bounded from below by $\left(\frac{1}{p+q} - 1\right) p > 0$. We conclude that when feedback is present then the definition of indecomposable channels (Definition 1) becomes more restrictive than it is for channels without

[3]This equality requires

$$\begin{aligned} \frac{p^2 + q - q^2}{p+q} &= 1 - q \\ \Rightarrow \quad p + q &= 1, \end{aligned}$$

but in our example $p + q = 0.9516$.

feedback. Therefore, when discussing indecomposable channels it should be explicitly stated if the property hold with feedback or only without feedback. In the following we refer to channels which are indecomposable only without feedback as *weakly indecomposable* channels and channels which satisfy the indecomposability property also with feedback as *strongly indecomposable* channels. The finite-state Markov channel is an example of a strongly indecomposable channel. Since no-feedback is a special case of feedback, then strongly indecomposable channels are also weakly indecomposable. In the next section we provide a simplification of the capacity expression for FSCs with feedback, for channels which are weakly indecomposable.

III. An Alternative Form of the Capacity Expression with Feedback

When feedback is present the capacity of the general FSC is given by [3]

$$C_{FB} = \lim_{n\to\infty} \max_{Q(x^n||y^{n-1})} \min_{s_0\in\mathcal{S}} \frac{1}{n} I(X^n \to Y^n|s_0). \quad (9)$$

In this work we show that for indecomposable FSCs with feedback the capacity can be found without searching over all initial channel states as the general expression (9) requires. This is stated in the following theorem:

Theorem 1. *Let $k(n)$ be a monotone non-decreasing function of n and denote*

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

For point-to-point weakly indecomposable FSCs with feedback the capacity is given by

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

where $p_{k(n)}Q_{k(n)+1}$ is a short notation for $p(x^{k(n)})Q_{k(n)+1}(x^{n-k(n)}||y^{n-k(n)-1})$ and $k(n)$ satisfies

$$\lim_{n\to\infty} k(n) = \infty, \qquad \text{and} \qquad \lim_{n\to\infty} \frac{k(n)}{n} = 0. \quad (12)$$

Proof: see Appendix A. ■

An example for $k(n)$ that satisfies (12) is $k(n) = [\sqrt{n}]$. This expression also provides insight into the channel, as explained in the following discussion.

Discussion

One way to interpret the expression in (10) is to note that without feedback, for both indecomposable and non-indecomposable channels, capacity (subject to Definition 4) is dominated by the state transitions in the "long-term"[4]. Thus, the contribution of the first $k(n)$ inputs to the capacity is negligible and it is enough to evaluate the rate achieved with the remaining $n - k(n)$ symbols. If the channel is non-indecomposable then there may be several "long-term" rates depending on the structure of the channel. If the channel is indecomposable, then there is only one "long-term" behavior; thus, the initial state does not matter what was the initial state. When feedback is applied, if the initial state is the worst-case state, then we will not reduce the rate by sending into the channel $k(n)$ symbols without feedback, since this will bring us to a steady-state achieved due to the indecomposability of the channel. Unless this steady-state is atomic in the worst-case state, the situation will improve as there is a positive probability that the channel state when feedback begins will not be the worst-case initial state. The feedback scheme now can be designed assuming it begins to operate when the initial states are distributed according to their steady-state distribution instead of assuming the worst-case state.

Consider the PtP-FSC without feedback. When the channel is non-indecomposable, then knowledge of the initial state at the receiver may significantly improve performance compared with a situation where the initial state is unknown: as long as the worst-case initial state does not occur all the time, then considering a block of B messages, we can achieve a higher average rate, as for each block, the rate $\lim_{n\to\infty} \max_{p(x^n)} \frac{1}{n} I(X^n \to Y^n|s_0)$ is achievable. Now, if for weakly indecomposable FSCs with feedback the capacity depends on the initial state, then this implies that feedback causes an indecomposable channel to behave like a "non-indecomposable" one. Thus, knowledge of the initial state can help to improve the average rate of the system. This will be in contrast to the indecomposable FSC without feedback. The question as to whether or not knowledge of the initial state increases capacity for indecomposable FSCs with feedback remains open

IV. Conclusions

In this work we focused on the capacity of weakly indecomposable FSCs with feedback. We first showed with an example that when feedback is present Gallager's definition of indecomposable channels becomes very restrictive. We then showed that the capacity-achieving distribution for weakly indecomposable FSCs with feedback, subject to the worst-case definition of the average probability of error, can be found without searching over all initial channel states.

Appendix A
Proof of Theorem 1

A. Codebook Generation and Achievable Rate

Let $m = (m_1, m_2)$, $m_q \in \mathcal{M}_q \triangleq \{1, 2, ..., 2^{n_q R}\}$, $q = 1, 2$, $n_1 = k$. Fix the p.m.fs $p(x^k)$ 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 p.m.f. $p(x^k)$. For each message $m_2 \in \mathcal{M}_2$ and feedback sequence $(y_{k+1}, y_{k+2}, ..., y_{k+n_2-1})$ 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 $\left\{x_i\left(m_2; y_{k+1}^{i-1}\right)\right\}_{i=k+1}^{n}$.

[4]By "long-term" we mean the behavior after the first $k(n)$ symbols.

As this is a special case of the scheme used in [3] to derive the capacity expression (9), we can write the achievable rate of this scheme for a given overall blocklength n and a length-k initial sequence as

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

where in (a) we used the fact that without feedback $\sum_{i=1}^k H(Y_i|Y^{i-1},X^i,s_0) = H(Y^k|X^k,s_0)$, [7].

B. 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 [3] it was established that $\lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}$ exists and is finite. Let $k(n)$ be a monotone non-decreasing function of n such that $\frac{k(n)}{n}$ is monotone non-increasing and

$$\lim_{n\to\infty} k(n) = \infty, \qquad \lim_{n\to\infty}\frac{k(n)}{n} = 0. \quad \text{(A.2)}$$

This can be satisfied, for example, by setting $k(n) = [\sqrt{n}]$. We note the obvious fact that

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

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

We now have the following lemma:

Lemma 1. *For every* $0<k<n$

$$\frac{n-k}{n}\underline{\mathcal{R}}_{n-k} - \frac{\log_2|\mathcal{S}|}{n} \le \mathcal{R}_n(k) \le \underline{\mathcal{R}}_n.$$

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.3)

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

Combined with Lemma 1, this implies that

$$\lim_{n\to\infty}\mathcal{R}_n(k(n)) = \lim_{n\to\infty}\underline{\mathcal{R}}_n = C_{FB}. \quad \text{(A.4)}$$

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 $\frac{1}{n}I(X^l;Y^n|s_0)_p + \frac{1}{n}\tilde{I}_{k(n)}(X^n\to Y^n|s_0)_p$. we now have for the same input distribution p

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

We also note that if $\forall x,\ a(x) > b(x)$ then $\max_x a(x) \ge \max_x b(x)$[5]. Therefore,

$$\begin{aligned}
&\lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}||y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\bigg(I(X^{k(n)};Y^{k(n)}|s_0) + \sum_{i=k(n)+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)\bigg) - \frac{\log_2|\mathcal{S}|}{n} \\
&\ge \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}||y^{n_2-1})} \min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k(n)+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)
\end{aligned}$$

and also

$$\begin{aligned}
&\lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}||y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\bigg(I(X^{k(n)};Y^{k(n)}|s_0) + \sum_{i=k(n)+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)\bigg) - \frac{\log_2|\mathcal{S}|}{n} \\
&\le \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}||y^{n_2-1})}\min_{s_0\in\mathcal{S}}\frac{1}{n}\bigg(k(n)\log_2|\mathcal{X}| + \sum_{i=k(n)+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)\bigg) - \frac{\log_2|\mathcal{S}|}{n} \\
&= \lim_{n\to\infty}\max_{p(x^{k(n)})Q_{k(n)+1}(x^{n_2}||y^{n_2-1})} \min_{s_0\in\mathcal{S}}\frac{1}{n}\sum_{i=k(n)+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)
\end{aligned}$$

giving the expression used in (10) and (11).

We now return to the proof of Lemma 1: in the proof of Lemma 1 we use [4, Lemma 2][6], restated here for convenience:

Lemma 2. [4, Lemma 2] *Let* (Z^n, U^n, S) *be a joint ensemble of random variables such that* $|\mathcal{S}|$ *is finite. For* $0<i_0<n$ *it holds that*

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

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

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

Proof of Lemma 1: Let $(\underline{p}_n, \underline{s}_{0,n})$ be the pair that achieves the max-min solution for $\overline{\mathcal{R}}_n(k)$ and let $(p^*_{n-k}, s^*_{0,n-k})$ achieve the max-min solution for $\underline{\mathcal{R}}_{n-k}$. We can also write $\underline{p}_n = \underline{p}_k \underline{Q}_{k+1}$ and $p^*_{n-k} = Q^*(n-k)$. Then

$$\begin{aligned}
\overline{\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} \\
&\qquad + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{p_n}\Bigg] - \frac{\log_2|\mathcal{S}|}{n} \\
&= \frac{1}{n}\Bigg[I(X^k;Y^k|\underline{s}_{0,n})_{\underline{p}_n} \\
&\qquad + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},\underline{s}_{0,n})_{\underline{p}_n}\Bigg] - \frac{\log_2|\mathcal{S}|}{n} \\
&= \frac{1}{n}\Bigg[I(X^k;Y^k|\underline{s}_{0,n})_{\underline{p}_k\underline{Q}_{k+1}} \\
&\qquad + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},\underline{s}_{0,n})_{\underline{p}_k\underline{Q}_{k+1}}\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)_{\underline{p}_k\underline{Q}^*_{k+1}} \\
&\qquad + \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{\underline{p}_k\underline{Q}^*_{k+1}}\Bigg] - \frac{\log_2|\mathcal{S}|}{n} \\
&\geq \min_{s_0\in\mathcal{S}} \frac{1}{n}\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} - \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)_{\underline{p}_k\underline{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)_{\underline{p}_k\underline{Q}^*_{k+1}} - 2\frac{\log_2|\mathcal{S}|}{n} \\
&= \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\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_0)_{\underline{p}_k\underline{Q}^*_{k+1}} \\
&\qquad - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] - \frac{\log_2|\mathcal{S}|}{n} \\
&= \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\Bigg[\frac{1}{n-k}\sum_{i=k+1}^{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0) \\
&\qquad \times I(X^i_{k+1};Y_i|Y^{i-1}_{k+1},s_k,s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] - \frac{\log_2|\mathcal{S}|}{n} \\
&= \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0) \\
&\qquad \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,s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] \\
&\qquad - \frac{\log_2|\mathcal{S}|}{n} \\
&\overset{(d)}{=} \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0) \\
&\qquad \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)_{\underline{p}_k\underline{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] \\
&\qquad - \frac{\log_2|\mathcal{S}|}{n} \\
&\overset{(e)}{=} \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0) \\
&\qquad \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)_{Q^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] \\
&\qquad - \frac{\log_2|\mathcal{S}|}{n} \\
&\geq \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0) \\
&\qquad \times \min_{s_l\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_l)_{Q^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n-k}\Bigg] \\
&\qquad - \frac{\log_2|\mathcal{S}|}{n} \\
&\overset{(f)}{=} \min_{s_0\in\mathcal{S}} \frac{n-k}{n}\sum_{\mathcal{S}} \tilde{p}(s_k|s_0)\underline{\mathcal{R}}_{n-k} - \frac{\log_2|\mathcal{S}|}{n} \\
&= \frac{n-k}{n}\underline{\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 $\underline{\mathcal{R}}_{n-k}$, i.e., $\underline{Q}^*_{k+1} = Q^*(n-k)$ with the appropriate index shift, (b) follows from Lemma 2 and the relationships

$$\begin{aligned}
&\sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} \\
&\quad \geq \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n} \\
&\quad \geq \min_{s_0\in\mathcal{S}} \sum_{i=k+1}^{n} I(X^i;Y_i|Y^{i-1},S_k,s_0)_{\underline{p}_k\underline{Q}^*_{k+1}} - \frac{\log_2|\mathcal{S}|}{n},
\end{aligned}$$

(c) follows from:

$$\begin{aligned}
&H(Y_i|X^i,Y^{i-1},S_k,s_0) \\
&= \sum_{\mathcal{X}^i\times\mathcal{Y}^{i-1}\times\mathcal{S}} p(x^i,y^{i-1},s_k|s_0) H(Y_i|x^i,y^{i-1},s_k,s_0) \\
&= \sum_{\mathcal{X}^i\times\mathcal{Y}^{i-1}\times\mathcal{S}} p(x^i,y^{i-1},s_k|s_0) \\
&\quad \times \sum_{\mathcal{Y}} p(y_i|x^i,y^{i-1},s_k,s_0)\log_2\frac{1}{p(y_i|x^i,y^{i-1},s_k,s_0)} \\
&\overset{(a')}{=} \sum_{\mathcal{X}^i\times\mathcal{Y}^{i-1}\times\mathcal{S}} p(x^i,y^{i-1},s_k|s_0) \\
&\quad \times \sum_{\mathcal{Y}} p(y_i|x^i_{k+1},y^{i-1}_{k+1},s_k)\log_2\frac{1}{p(y_i|x^i_{k+1},y^{i-1}_{k+1},s_k)} \\
&= \sum_{\mathcal{X}^i\times\mathcal{Y}^{i-1}\times\mathcal{S}} p(x^i,y^{i-1},s_k|s_0) H(Y_i|x^i_{k+1},y^{i-1}_{k+1},s_k) \\
&= \sum_{\mathcal{X}^{i-k}\times\mathcal{Y}^{i-k-1}\times\mathcal{S}} p(x^i_{k+1},y^{i-1}_{k+1},s_k|s_0) H(Y_i|x^i_{k+1},y^{i-1}_{k+1},s_k) \\
&= H(Y_i|X^i_{k+1},Y^{i-1}_{k+1},S_k,s_0),
\end{aligned}$$

To see (a') write

$$\begin{aligned}
p(y_i|x^i, y^{i-1}, s_k, s_0) &= \frac{p(y^i, x^i|s_k, s_0)}{p(y^{i-1}, x^i|s_k, s_0)} \\
&= \frac{\sum_{\mathcal{S}_{k+1}^i} \prod_{j=k+1}^i p(y_j, s_j|x_j, s_{j-1})}{\sum_{\mathcal{S}_{k+1}^{i-1}} \prod_{j=k+1}^{i-1} p(y_j, s_j|x_j, s_{j-1})},
\end{aligned}$$

which is independent of s_0, x^k and y^k. Next, (d) and (e) follow from the structure of the finite-state channel and the fact that only feedback from Y_{k+1}^n is used[7]. Finally, (f) is because in $\underline{\mathcal{R}}_{n-k}$ we take the minimizing initial state.

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

■

C. *The Asymptotic Expression for* (A.1)

We prove the following lemma:

Lemma 3. *Let* $p_n = p_k Q_{k+1}$ *and*

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

For every $\epsilon > 0$*, there exists* k *large enough such that*

$$\lim_{n\to\infty} \left| \widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k) \right| \le \epsilon |\mathcal{S}| \log_2 |\mathcal{Y}|.$$

Proof: Define first

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

Clearly $\underset{\sim}{\mathcal{R}}_n(k) \ge \underset{\sim}{\mathcal{R}}_n(k; p_n)$. Let $\big(\tilde{p}_n, \tilde{s}_{0,n}\big)$ 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

$$\begin{aligned}
&\left| \widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k) \right| \\
&\le \left| \widetilde{\mathcal{R}}_n(k) - \underset{\sim}{\mathcal{R}}_n(k; \tilde{p}_n) \right| \\
&= \Bigg| \frac{1}{n} \Bigg[I(X^k; Y^k|\tilde{s}_{0,n})_{\tilde{p}_n} + \sum_{i=k+1}^n I(X^i; Y_i|Y^{i-1}, \tilde{s}_{0,n})_{\tilde{p}_n} \Bigg] \\
&\quad - \frac{1}{n} \Bigg[I(X^k; Y^k|\underset{\sim}{s}_{0,n})_{\tilde{p}_n} + \sum_{i=k+1}^n I(X^i; Y_i|Y^{i-1}, \underset{\sim}{s}_{0,n})_{\tilde{p}_n} \Bigg] \Bigg| \\
&\le \Bigg| \frac{1}{n} \sum_{i=k+1}^n I(X^i; Y_i|Y^{i-1}, \tilde{s}_{0,n})_{\tilde{p}_n} \\
&\qquad - \frac{1}{n} \sum_{i=k+1}^n I(X^i; Y_i|Y^{i-1}, \underset{\sim}{s}_{0,n})_{\tilde{p}_n} \Bigg| + \frac{k}{n} \log_2 |\mathcal{X}| \\
&\le 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} \\
&\qquad - \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}| \\
&= 2\frac{\log_2 |\mathcal{S}|}{n} + \Bigg| \frac{1}{n} \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, S_k, \tilde{s}_{0,n})_{\tilde{p}_n} \\
&\quad - \frac{1}{n} \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, S_k, \underset{\sim}{s}_{0,n})_{\tilde{p}_n} \Bigg| + \frac{k}{n} \log_2 |\mathcal{X}| \\
&= \Bigg| \frac{1}{n} \sum_{\mathcal{S}} \tilde{p}(s_k|\tilde{s}_{0,n}) \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k, \tilde{s}_{0,n})_{\tilde{p}_n} \\
&\quad - \frac{1}{n} \sum_{\mathcal{S}} \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k, \underset{\sim}{s}_{0,n})_{\tilde{p}_n} \Bigg| \\
&\qquad + \frac{k}{n} \log_2 |\mathcal{X}| + 2\frac{\log_2 |\mathcal{S}|}{n} \\
&= \Bigg| \frac{1}{n} \sum_{\mathcal{S}} \tilde{p}(s_k|\tilde{s}_{0,n}) \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k)_{\tilde{Q}_{k+1}} \\
&\qquad - \frac{1}{n} \sum_{\mathcal{S}} \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k)_{\tilde{Q}_{k+1}} \Bigg| \\
&\qquad + \frac{k}{n} \log_2 |\mathcal{X}| + 2\frac{\log_2 |\mathcal{S}|}{n} \\
&= \frac{k}{n} \log_2 |\mathcal{X}| + 2\frac{\log_2 |\mathcal{S}|}{n} \\
&\qquad + \Bigg| \frac{1}{n} \sum_{\mathcal{S}} \Big(\tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \Big) \\
&\qquad\quad \times \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k)_{\tilde{Q}_{k+1}} \Bigg| \\
&\le \frac{k}{n} \log_2 |\mathcal{X}| + 2\frac{\log_2 |\mathcal{S}|}{n} \\
&\qquad + \Bigg| \frac{1}{n} \sum_{\mathcal{S}} \Big| \tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \Big| \\
&\qquad\quad \times \sum_{i=k+1}^n I(X_{k+1}^i; Y_i|Y_{k+1}^{i-1}, s_k)_{\tilde{Q}_{k+1}} \Bigg|
\end{aligned}$$

[7] For example we have that

$$\begin{aligned}
&p(y_i|y_{k+1}^{i-1}, s_k) \\
&= \frac{p(y_{k+1}^i|s_k)}{p(y_{k+1}^{i-1}|s_k)} \\
&= \frac{\sum_{\mathcal{X}^{i-k}} p(y_{k+1}^i, x_{k+1}^i|s_k)}{\sum_{\mathcal{X}^{i-k-1}} p(y_{k+1}^{i-1}, x_{k+1}^{i-1}|s_k)} \\
&= \frac{\sum_{\mathcal{X}^{i-k}} \prod_{j=k+1}^i p(y_j, x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}, s_k)}{\sum_{\mathcal{X}^{i-k-1}} \prod_{j=k+1}^{i-1} p(y_j, x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}, s_k)} \\
&= \frac{\sum_{\mathcal{X}^{i-k}} \prod_{j=k+1}^i p(x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}, s_k) p(y_j|y_{k+1}^{j-1}, x_{k+1}^{j}, s_k)}{\sum_{\mathcal{X}^{i-k-1}} \prod_{j=k+1}^{i-1} p(x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}, s_k) p(y_j|y_{k+1}^{j-1}, x_{k+1}^{j}, s_k)} \\
&= \frac{\sum_{\mathcal{X}^{i-k}} \prod_{j=k+1}^i p(x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}) p(y_j|y_{k+1}^{j-1}, x_{k+1}^{j}, s_k)}{\sum_{\mathcal{X}^{i-k-1}} \prod_{j=k+1}^{i-1} p(x_j|y_{k+1}^{j-1}, x_{k+1}^{j-1}) p(y_j|y_{k+1}^{j-1}, x_{k+1}^{j}, s_k)},
\end{aligned}$$

which can be completely evaluated using only Q_{k+1}^*. Thus $\underset{\sim}{p}_k$ is not needed.

$$
\begin{aligned}
&\leq \sum_{\mathcal{S}} \Big| \tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \Big| \frac{n-k}{n} \log_2|\mathcal{Y}| \\
&\qquad\qquad + \frac{k}{n}\log_2|\mathcal{X}| + 2\frac{\log_2|\mathcal{S}|}{n} \\
&\overset{(a)}{\leq} |\mathcal{S}|\epsilon\frac{n-k}{n}\log_2|\mathcal{Y}| + \frac{k}{n}\log_2|\mathcal{X}| + 2\frac{\log_2|\mathcal{S}|}{n} \\
&\overset{n\to\infty}{\longrightarrow} \epsilon|\mathcal{S}|\log_2|\mathcal{Y}|,
\end{aligned}
$$

(a) follows from

$$
\begin{aligned}
&\Big| \tilde{p}(s_k|\tilde{s}_{0,n}) - \tilde{p}(s_k|\underset{\sim}{s}_{0,n}) \Big| \\
&= \Big| \sum_{\mathcal{X}^k} \tilde{p}(s_k, x^k|\tilde{s}_{0,n}) - \sum_{\mathcal{X}^k} \tilde{p}(s_k, x^k|\underset{\sim}{s}_{0,n}) \Big| \\
&= \Big| \sum_{\mathcal{X}^k} p(x^k|\tilde{s}_{0,n})\tilde{p}(s_k|x^k, \tilde{s}_{0,n}) \\
&\qquad\qquad - \sum_{\mathcal{X}^k} p(x^k|\underset{\sim}{s}_{0,n})\tilde{p}(s_k|x^k, \underset{\sim}{s}_{0,n}) \Big| \\
&= \Big| \sum_{\mathcal{X}^k} p(x^k)\tilde{p}(s_k|x^k, \tilde{s}_{0,n}) - \sum_{\mathcal{X}^k} p(x^k)\tilde{p}(s_k|x^k, \underset{\sim}{s}_{0,n}) \Big| \\
&\leq \sum_{\mathcal{X}^k} p(x^k) \Big| \tilde{p}(s_k|x^k, \tilde{s}_{0,n}) - \tilde{p}(s_k|x^k, \underset{\sim}{s}_{0,n}) \Big| \\
&\leq \sum_{\mathcal{X}^k} p(x^k)\epsilon \\
&= \epsilon.
\end{aligned}
$$

■

D. Combining Lemma 1 and Lemma 3

From Lemma 3 we conclude that $\lim_{n\to\infty} \underset{\sim}{\mathcal{R}}_n(k(n))$ is independent of the initial state. Since from (A.4) $\lim_{n\to\infty} \underset{\sim}{\mathcal{R}}_n(k(n)) = \lim_{n\to\infty} \underline{\mathcal{R}}_n$ then we conclude that C_{FB} can be evaluated without searching over all initial states.

REFERENCES

[1] B. McMillan. "The Basic Theorems of Information Theory". *The Annals of Mathematical Statistics*, Vol-24(2):196–219, 1953.

[2] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley and Sons Inc., 1968.

[3] H. Permuter, T. Weissman and A. Goldsmith. "Finite-State Channels with Time-Invariant Deterministic Feedback". Submitted to *IEEE Trans. Inform. Theory 2007, revised 2008.*

[4] H. Permuter and T. Weissman. "Capacity Region of the Finite-State Multiple Access Channel with and without Feedback". Submitted to the *IEEE Trans. Inform. Theory*, 2007.

[5] T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., 1991.

[6] G. Kramer. "Capacity Results for Discrete Memoryless Networks". *IEEE Trans. Inform. Theory*, IT-49(1):4–21, 2003.

[7] J. L. Massey. "Causality, Feedback and Directed Information". *Proc. 1990 Intl. Symp. on Info. Th. and its Applications*, Waikiki, Hawaii, November, 1990.