To Feed or Not to Feed Back

Himanshu Asnani
Information Systems Lab
Stanford University
Stanford, CA, USA
Email: asnani@stanford.edu

Haim Permuter
Elec. & Comp. Engg. Dept.
Ben-Gurion University
Beer-Sheva, Israel
Email: haimp@bgu.ac.il

Tsachy Weissman
Information Systems Lab
Stanford University
Stanford, CA, USA
Email: tsachy@stanford.edu

***Abstract*—We prove capacity results for a communication system with Finite State Channels (FSCs), where the encoder and the decoder can control the availability or the quality of the noise-free feedback. The instantaneous feedback is a function of a cost constrained action taken by the encoder, a cost constrained action taken by the decoder, and the channel output. Achievability is through construction of a sequence of convergent achievable rates, using a simple scheme based on '*code tree*' generation, that generates channel input symbols along with encoder and decoder actions. For a given block length N and probability of error, ϵ_N, we give an upper bound on the maximum achievable rate. For stationary *indecomposable* channels without intersymbol interference (ISI), the capacity is given as the limit of normalized directed information between the input and output sequence, maximized over an appropriate set of causally conditioned distributions. As important special cases, we characterize (a) the framework of '*to feed or not to feed back*' where either the encoder or the decoder takes binary actions to determine whether current channel output will be fed back to the encoder, with a constraint on the fraction of channel outputs that are fed back, (b) the capacity of '*coding on the backward link*' in FSCs, i.e. when the decoder sends limited-rate instantaneous coded noise-free feedback on the backward link.**

I. Introduction

In his book [1], Gallager introduced finite state channels (FSCs) as an apt model for a very broad family of channels with memory. When no feedback is present and the channel is stationary and indecomposable (defined in later sections) without intersymbol interference (ISI), the capacity was shown by Gallager in [1] and by Blackwell, Breiman and Thomasian in [2] to be

$$C_{NF} = \lim_{N\to\infty} \frac{1}{N} \max_{P(x^N)} I(X^N;Y^N). \quad (1)$$

For the case of no ISI, stationary and indecomposable finite state channels with time invariant deterministic feedback, the capacity was shown in [3] to be,

$$C_{FB} = \lim_{N\to\infty} \frac{1}{N} \max_{Q(x^N \| z^{N-1})} I(X^N \to Y^N), \quad (2)$$

where $Q(x^N \parallel z^{N-1})$ is causal conditioning introduced by Kramer in [4], defined as,

$$Q(x^N \parallel z^{N-1}) \quad \triangleq \quad \prod_{i=1}^{N} Q(x_i|x^{i-1}, z^{i-1}). \quad (3)$$

and $I(X^N \to Y^N)$ is the directed information introduced by Massey in [6], where he credits it to Marko [7] which also appears in the work of Tatikonda, [8]. Here Z_i is a time-invariant deterministic function of the output Y_i. When the channels have memory, feedback can increase the capacity even for single user channels, such as the recent result on capacity of trapdoor channel with feedback in [5]. In [10], the notion of *actions* in source coding context was introduced, where now the decoder can take actions based on the index obtained from the encoder to affect the formation or availability of side information. In [11], the channel coding dual is studied where the transmitter takes actions that affect the formation of channel states. Recently, in [12], the channel coding setting in [11] was generalized, to accommodate the case where both the encoder and the decoder take channel probing actions, with associated costs, to maximize the rate of reliable communication. This was referred to as the '*Probing Capacity*'.

In this paper, we introduce the notion of actions in acquisition of noise-free feedback or its deterministic function for FSCs. The main contribution of this paper is in characterizing the cost-capacity trade-off when the feedback observed by the encoder is a deterministic function of an action taken by the encoder, an action taken by the decoder, and the channel output, when actions are required to satisfy an average cost constraint. More precisely, the encoder observes '*sampled*' feedback $Z_i = f(A_{e,i}, A_{d,i}, Y_i)$, where $f(\cdot)$ is a deterministic function, Y_i is the channel output, $A_{e,i} = A_{e,i}(M, Z^{i-1})$ is the action taken by the encoder as a function of the message and the past sampled feedback, and $A_{d,i}$ is the action taken by the decoder, where we study two scenarios: one where that action is strictly causal in the channel output, i.e., $A_{d,i} = A_{d,i}(Y^{i-1})$, and one where it can depend also on the present channel output, i.e., $A_{d,i} = A_{d,i}(Y^i)$. The problem is motivated by practical applications where acquisition of the feedback may be costly, and either or both the encoder and decoder influence whether and what from the channel output is to be fed back. We motivate and compute certain special cases as that of *to feed or not to feed back*, i.e., where actions are binary corresponding to observing the channel output or not observing it, the cost constraint corresponding to the fraction of channel output observations allowed, and the channel states evolve as a markov chain independent of the channel input process. Another special case is of the coding on the backward link, where the decoder sends a symbol from the action alphabet based on the channel outputs observed so far, thus operating at an instantaneous rate which is log the cardinality of said alphabet.

The rest of paper is organized as follows. Section II describes the channel model and formulates the problem studied in this paper. The main results of this paper are outlined in Section III. Section IV is dedicated to capacity-achieving coding schemes along with the characterization of the capacity for stationary, indecomposable, finite state channels without intersymbol interference (ISI). Section V presents single letter lower bounds for a specific example of *to feed or not to feed back* (i.e. when actions are binary) for Markovian channels when only one of the two, encoder or decoder, takes the actions. Section VI computes for an example of a Markovian channel the case of coding on the backward link for FSCs. The paper is summarized and concluded in Section VII.

II. Channel Model and Problem Formulation

Let upper case, lower case, and calligraphic letters denote, respectively, random variables, specific or deterministic values they may assume, and their alphabets. For two jointly distributed random variables, X and Y, let P_X, P_{XY} and $P_{X|Y}$ respectively denote the marginal of X, joint distribution of (X,Y) and conditional distribution of X given Y. X_m^n is a shorthand for $n-m+1$ tuple $\{X_m, X_{m+1}, \cdots, X_{n-1}, X_n\}$. X^n will also denote X_1^n. When $i \leq 0$, X^i denotes null string as it is also for X_i^j, when $i \geq j$. $X^{n \setminus i}$ denotes $\{X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n\}$. The cardinality of an alphabet $\mathcal{X}$ is denoted by $|\mathcal{X}|$. We impose the assumption of finiteness of cardinality on all alphabets, unless otherwise indicated.
We use the *Causal Conditioning* notation $(\cdot \parallel \cdot)$ as in Eq. (3) and the following notation as introduced in [3] :

$$P(y^N \parallel x^{N-1}) \triangleq \prod_{i=1}^{N} P(y_i|x^{i-1}, y^{i-1}). \tag{4}$$

The directed information conditioned on a random object S, $I(X^N \to Y^N|S)$, is defined as,

$$I(X^N \to Y^N|S) \triangleq \sum_{i=1}^{N} I(X^i; Y_i|Y^{i-1}, S). \tag{5}$$

We model discrete time channels with memory as Finite State Channels (FSCs) where channel input symbols take values in the finite alphabet $\mathcal{X}$ and output denoted by Y takes values in finite alphabet $\mathcal{Y}$. The state takes values in a finite alphabet $\mathcal{S}$. The stationary channel is characterized by the conditional probability law $P(y_i, s_i|x_i, s_{i-1})$ satisfying,

$$P(y_i, s_i|x^i, s^{i-1}, y^{i-1}, a_e^i, a_d^i, m) = P(y_i, s_i|x_i, s_{i-1}), \tag{6}$$

where $a_{e,i} \in \mathcal{A}_e$ and $a_{d,i} \in \mathcal{A}_d$ are the encoder and decoder actions respectively as will be explained later. Messages $M \in \mathcal{M}$ are assumed to be independent of initial state, s_0. The FSC is without intersymbol interference (ISI) if

$$P(s_i|s_{i-1}, x_i) = P(s_i|s_{i-1}). \tag{7}$$

The basic framework in this paper is the setting depicted in Fig. 1. The communication system has the following building blocks :

- *Encoder Feedback Logic* : Generates encoder actions, $\{A_{e,i}\}_{i=1}^N$, using the function $f_{A_{e,i}} : \mathcal{M} \times \mathcal{Z}^{i-1} \to \mathcal{A}_e$ i.e., $A_{e,i} = f_{A_{e,i}}(M, Z^{i-1})$, where $Z_i \in \mathcal{Z}$ is the sampled feedback component.
- *Decoder Feedback Logic* : Generates decoder actions, $\{A_{d,i}\}_{i=1}^N$, using the function $f_{A_{d,i}} : \mathcal{Y}^{i-1} \to \mathcal{A}_d$ i.e., $A_{d,i} = f_{A_{d,i}}(Y^{i-1})$. where $Y_i \in \mathcal{Y}$ is the channel output.
- *Feedback Sampler* : Generates *sampled* feedback, $Z_i = f(A_{e,i}, A_{d,i}, Y_i)$, where f is a deterministic function.
- *Channel Encoder* : Constructs channel input symbol, $X_i(M, Z^{i-1})$, using the encoding function, $f_{e,i} : \mathcal{M} \times \mathcal{Z}^{i-1} \to \mathcal{X}$.
- *Channel Decoder* : Generates the best estimate of the message given the channel output, $\hat{M}(Y^N)$, using the decoding function, $f_d : \mathcal{Y}^N \to \mathcal{M}$.

We are interested in characterizing the maximal rate of reliable communication under the average cost constraint,

$$\mathsf{E}\left[\Lambda(A_e^N, A_d^N)\right] = \mathsf{E}\left[\frac{1}{N}\sum_{i=1}^{N} \Lambda(A_{e,i}, A_{d,i})\right] \leq \Gamma, \tag{8}$$

where $\Lambda(\cdot,\cdot)$ is a given cost function satisfying $\max_{a_e \in \mathcal{A}_e, a_d \in \mathcal{A}_d} \Lambda(a_e, a_d) = \Lambda_{\max} < \infty$.

Definition 1: A rate R is said to be *achievable* if there exists a sequence of block codes $(N, \lceil 2^{NR} \rceil)$ satisfying (8) such that the maximal probability of error,

$$\max_{m \in \{1,\cdots,\lceil 2^{NR} \rceil\}} \Pr(\hat{m} \neq m|\text{message } m \text{ was sent}),$$

vanishes as $N \to \infty$. The capacity of such a system is denoted by C which is the supremum of all achievable rates.

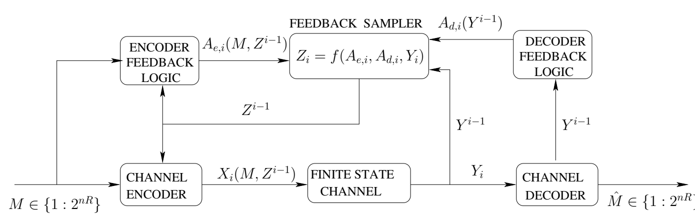

Fig. 1. Modeling **Feedback Sampling** for the acquisition of feedback in Finite State Channels (FSCs).

III. Main Results

Let s_0 denote the initial state. We define $\underline{C}_N(\Gamma)$ and $\overline{C}_N(\Gamma)$ as,

$$\underline{C}_N(\Gamma) \triangleq \frac{1}{N} \max \min_{s_0} I(X^N \to Y^N|s_0) \tag{9}$$

$$\overline{C}_N(\Gamma) \triangleq \frac{1}{N} \max \max_{s_0} I(X^N \to Y^N|s_0). \tag{10}$$

Here **max** denotes maximization over the joint probability distribution,

$$P(s_0, x^N, a_e^N, a_d^N, y^N, z^N) = P(s_0)Q(x^N, a_e^N \parallel z^{N-1}) \times Q(a_d^N \parallel y^{N-1})P(y^N \parallel x^N, s_0)\prod_{i=1}^{N}\mathbf{1}_{\{z_i=f(a_{e,i},a_{d,i},y_i)\}},$$

such that $\mathsf{E}[\Lambda(A_e^N, A_d^N)] \leq \Gamma$, where z_i will stand for $f(a_{e,i}, a_{d,i}, y_i)$. Note that effectively maximization in definition of $\underline{C}_N(\Gamma)$ and $\overline{C}_N(\Gamma)$ is over $Q(x^N, a_e^N \parallel z^{N-1})Q(a_d^N \parallel y^{N-1})$ as $P(s_0)$ is fixed and $P(y^N \parallel x^N, s_0)$ (and likewise $P(y^N \parallel x^N)$) is a characteristic of the channel given by (Lemma 6 of [3]). Our main results are as follows,

- **Achievable Rate :** For a communication abstraction as in Fig. 1, any rate R is achievable such that,
$$R < \lim_{N\to\infty} \underline{C}_N(\Gamma) = \sup_N \left[\underline{C}_N(\Gamma) - \frac{\log|S|}{N}\right]. \quad (11)$$
- **Converse :** Consider a coding scheme with rate R which achieves reliable communication over the FSC with feedback sampling as in Fig. 1. This implies the existence of $(N, \lceil 2^{NR}\rceil)$ codes such that the probability of error P_e^N goes to zero as $N \to \infty$. For such a scheme given $\epsilon_N > 0$, $\exists$ block length N_0 such that for all block lengths $N > N_0$ we have
$$R \leq \overline{C}_N(\Gamma) + \epsilon_N. \quad (12)$$
Note that unlike in [3], limit of $\overline{C}_N(\Gamma)$ may not exist because sub-additivity (like the one in Theorem 16 in [3]) breaks due to the presence of cost constraints. Hence for a general FSC, we have the above converse result for a give blocklength N and probability of error, ϵ_N.
- **Capacity :** In the following cases we characterize the capacity exactly,
  1) For an FSC where the probability of the initial state is positive for all $s_0 \in \mathcal{S}$, the capacity is evaluated exactly,
  $$C(\Gamma) = \lim_{N\to\infty} \underline{C}_N(\Gamma). \quad (13)$$
  2) For stationary '*indecomposable*' channels without ISI with feedback sampling as in Fig. 1, the capacity is,
  $$C(\Gamma) = \lim_{N\to\infty} \frac{1}{N} \max I(X^N \to Y^N), \quad (14)$$
  where **max** denotes maximization over the joint probability distribution,
  $$P(x^N, a_e^N, a_d^N, y^N) = Q(x^N, a_e^N \parallel z^{N-1}) \times Q(a_d^N \parallel y^{N-1})P(y^N \parallel x^N), \quad (15)$$
  such that $\mathsf{E}[\Lambda(A_e^N, A_d^N)] \leq \Gamma$.

IV. Achievability

We present here only the encoding and decoding schemes due to space constraints. The proof of existence of limit of $\underline{C}_N(\Gamma)$, the analysis of probability of error and the converse result is deferred to [9].

A. Encoding Scheme

Encoding is based on generating separate code trees which is described below. These are then revealed to the encoder and the decoder.

- *Encoder Code-Tree* : 2^{NR} **code-trees** are generated as follows, the i^{th} encoder action and channel input symbol is generated using a probability mass function which depends on previous encoder action and channel input symbols and on the past sampled feedback sequence, i.e. $Q(x^i, a_e^i | x^{i-1}, a_e^{i-1}, z^{i-1})$.
- *Decoder Action Code-Tree* : We generate a single code tree at random, where the vertex represents decoder action symbol, $a_{d,i}$ generated with distribution $Q(a_{d,i}|a_d^{i-1}, y^{i-1})$. Thus the present decoder action depend on the past actions as well as the past channel output.

Note that $\{Q(x^i, a_e^i|x^{i-1}, a_e^{i-1}, z^{i-1})\}_{i=1}^N$ and $\{Q(a_{d,i}|a_d^{i-1}, y^{i-1})\}_{i=1}^N$ correspond to the joint distribution on $(X^N, A_e^N, A_d^N, S^N, Y^N)$ such that constraint $\mathsf{E}\left[\Lambda(A_e^N, A_d^N)\right] \leq \Gamma$ is satisfied.

Using the decoder action symbol $a_{d,i}$, along with encoder actions, $a_{e,i}$ and channel output y_i, feedback sampler produces sampled feedback as $z_i = f(a_{e,i}, a_{d,i}, y_i)$. In this way, given a message m, and the complete sampled feedback sequence z^{N-1} thus obtained, there is a particular (x^N, a_e^N) which can be found from the collection of *encoder code trees*. The encoder thus sends the corresponding x^N though the channel.

B. Decoding

The decoder performs ML decoding, i.e. it chooses the message m for which $P(y^N|m)$ is maximized.

$$\begin{aligned}
& P(y^N|m) \\
= & \prod_{i=1}^{N} P(y_i|y^{i-1}, m) \\
\overset{(a)}{=} & \prod_{i=1}^{N} P(y_i|y^{i-1}, a_d^i(y^{i-1}), m, x^i(m, z^{i-1}), a_e^i(m.z^{i-1})) \\
\overset{(b)}{=} & \prod_{i=1}^{N} P(y_i|y^{i-1}, a_d^i(y^{i-1}), x^i(m, z^{i-1}), a_e^i(m.z^{i-1})) \\
= & P(y^N \parallel x^N, a_e^N, a_d^N) \\
\overset{(c)}{=} & P(y^N \parallel x^N),
\end{aligned}$$

where (a) follows from the fact that knowing m and y^{i-1}, we know (x^i, a_e^i, a_d^i). This can be iteratively shown. Given m we know $(x_1(m), a_1(m))$. We also know $a_{d,1}$. Given y_1, $z_1 = f(a_{e,1}, a_{d,1}, y_1)$. Hence now we know, $(x_2(m, z_1), a_{e,2}(m, z_1), a_{d,2}(y_1))$. Iteratively we can conclude that for a given message m and true feedback sequence, y^{i-1}, we can construct (x^i, a_e^i, a_d^i) knowing the codebooks. (b) follows from the assumption on channel model in Eq. (6) and (c) follows from our encoding scheme (refer to Lemma 1 in [9]). Hence ML decoding to construct message estimate,

$\hat{m}$ can also be done my maximizing causal conditioning, i.e.,

$$\hat{m} = \underset{m}{\text{argmax}}\, P(y^N|m) = \underset{x^N}{\text{argmax}}\, P(y^N \parallel x^N). \quad (16)$$

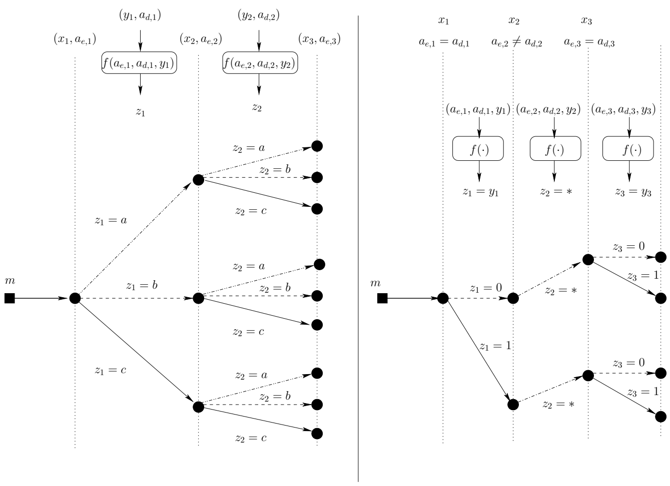

Fig. 2. This figure illustrates *Encoder Code-Trees* in our coding scheme. The left hand side figure depicts a general setting where $\mathcal{Z} = \{a, b, c\}$, and $z_i = f(a_{e,i}, a_{d,i}, y_i)$. The tree is shown for $N = 3$. The right hand side shows a specific example where $a_{d,i} = 0\ \forall i$ and output is binary. Actions of encoder, $a_{e,i} \in \{0, 1\}$ and $z_i = f(a_{e,i}, a_{d,i}, y_i) = y_i$ if $a_{e,i} = a_{d,i}$ or $a_{e,i} = 0$, else it is erasure($= *$). Hence some portion of the tree collapses as by knowing $a_{e,i}$ we know the possible values of z_i, for e.g. $a_{e,i} = 1$ implies $z_i = *$ and $a_{e,i} = 0$ implies, $z_i = 0$ or 1.

C. *Capacity for Stationary Indecomposable FSC without ISI*

We assume now that state transition is a separate markov chain and does not depend on input, i.e., $P(y_i, s_i|s_{i-1}, x_i) = P(s_i|s_{i-1})P(y_i|s_i, s_{i-1}, x_i)$. Such a channel is said to have no ISI. We further assume this channel is *indecomposable* as the definition given below,

Definition 2: An FSC without ISI is said to be indecomposable if, for every $\epsilon > 0$, $\exists N_0$ such that $\forall N > N_0$

$$|P(s_N|s_0) - P(s_N|s_0')| \leq \epsilon\ \forall\ s_N, s_0, s_0'. \quad (17)$$

The channel is stationary if $P(s_0) = \pi(s_0)$.

Theorem 1: For a stationary and indecomposable FSC without ISI and with communication abstraction as in Fig. 1, the capacity of the channel is given by,

$$C(\Gamma) = \lim_{N\to\infty} C^N(\Gamma) = \lim_{N\to\infty} \frac{1}{N} \max I(X^N \to Y^N), \quad (18)$$

where **max** denotes maximization over $Q(x^N, a_e^N \parallel z^{N-1})Q(a_d^N \parallel y^{N-1})$ such that $\mathsf{E}[\Lambda(A_e^N, A_d^N)] \leq \Gamma$.

Proof: The proof is similar to proof of Theorem 18 in [3] with $Q(x^N \parallel z^{N-1})$ replaced by $Q(x^N, a_e^N \parallel z^{N-1})Q(a_d^N \parallel y^{N-1})$. ■

V. Numerical Example 1 : To Feed or Not to Feed back

The actions now are binary, i.e., $\mathcal{A} = \{0, 1\}$. In this setting, action sequence determine *to feed or not to feed back* (by either encoder or decoder) a deterministic function of the past channel output, i.e.,

$$\begin{aligned} Z_i &= f(A_i, Y_i) = g(Y_i), \text{ if } A_i = 1 \\ Z_i &= f(A_i, Y_i) = *, \text{ if } A_i = 0, \end{aligned} \quad (19)$$

where $*$ stands for erasure or no information about feedback. As a specific example for such a setting, consider the communication system involving Markovian channel with encoder feedback logic as in Fig. 3, which is essentially a no ISI, stationary, indecomposable FSC. The cost function, $\Lambda(a) = a$, $a \in \mathcal{A}$ and the cost constraint is $\Gamma \in [0, 1]$.

Theorem 2: The capacity of the system in Fig. 3 with encoder feedback logic is lower bounded as,

$$C_{enc}(\Gamma) \geq C_{enc,lower}(\Gamma) = \max I(X; Y|S), \quad (20)$$

where maximization is over joint probability distribution,

$$P_{S,A,Z,X,Y}(s, a, z, x, y) = \pi_S(s)P_A(a)\mathbf{1}_{\{z=f(a,s)\}} \times P_{X|Z,A}(x|z, a)P_{Y|X,S}(y|x, s),$$

and $\mathsf{E}[\Lambda(A)] \leq \Gamma$. Now if instead of encoder we have decoder taking actions, but causally dependent on the channel output and the state (See Note 1), the capacity of such a system is lower bounded as,

$$C_{dec}(\Gamma) \geq C_{dec,lower}(\Gamma) = \max I(X; Y|S), \quad (21)$$

where maximization is over joint probability distribution,

$$P_{S,A,Z,X,Y}(s, a, z, x, y) = \pi_S(s)P_{A|S}(a|s)\mathbf{1}_{\{z=f(a,s)\}} \times P_{X|Z,A}(x|z, a)P_{Y|X,S}(y|x, s),$$

such that $\mathsf{E}[\Lambda(A)] \leq \Gamma$.

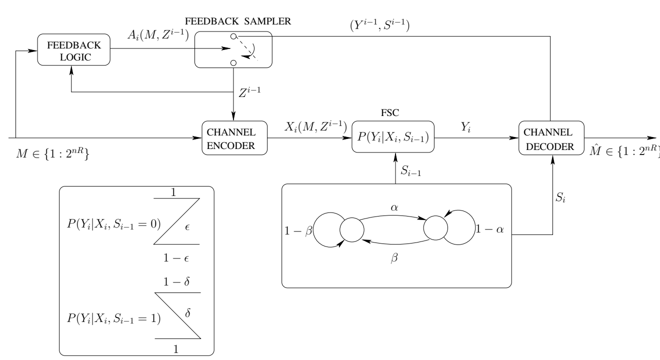

Fig. 3. **To feed or not to feed back** when encoder takes actions and decoder knows the state. States are stationary and evolve as a markov process.

Proof: Omitted (refer to Theorem 12 and 13 in [9]). ■

Note 1: Note this is slightly different from the canonical setting in Fig. 1, as the decoder takes action causally dependent on the channel outputs and states. However as shown in Section VII of [9] that the results for the setting in Fig. 1 can be applied to get fundamental limits for this setting.

We evaluate the lower bounds for the example in Fig. 3, when $\alpha = \beta = \epsilon = \delta = 0.5$. The region is shown in Fig. 4.

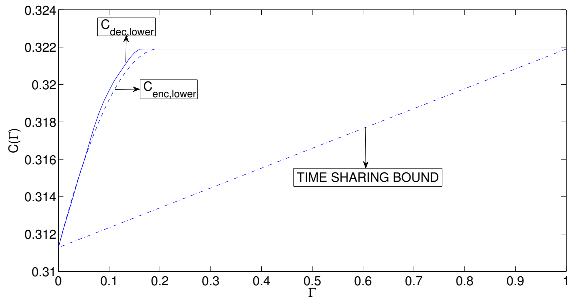

Fig. 4. Cost-capacity trade off for example in Fig. 3. $C_{enc,lower}$ is the lower bound on capacity with encoder feedback logic. If instead of encoder decoder decides (causally dependent on channel output and state) when encoder will sample feedback, $C_{dec,lower}$ is a lower bound on the capacity. The straight line represents time sharing scheme which is strictly sub-optimal.

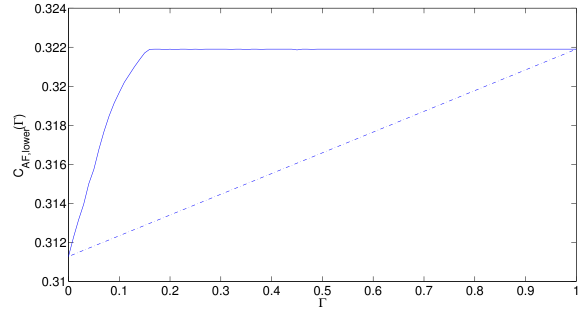

Fig. 6. Cost-capacity trade off for example in Fig. 5. $C_{AF,lower}$ is the lower bound on capacity. The straight line represents naive time sharing scheme.

VI. Numerical Example 2 : Coding on the Backward Link in FSC

Consider the setting depicted in Fig. 5. We allow *coding on the backward link*, i.e., decoder encodes the channel outputs causally $(A_i(Y^i, S^i) \in \mathcal{A})$[again refer to Note 1] and sends it to the encoder. The encoder uses the acquired *active feedback* symbols to generate channel input symbols, i.e., $X_i(M, A^{i-1})$. For stationary indecomposable FSCs with active feedback we denote the capacity by C_{AF}.

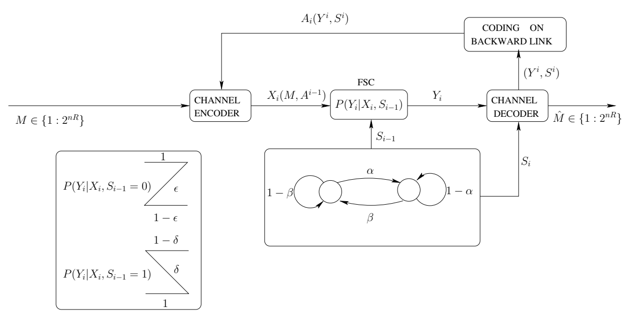

Fig. 5. Modeling **coding on the backward link** for markovian channel with binary states.

Theorem 3: For the system in Fig. 5, the capacity is lower bounded as,

$$C_{AF}(\Gamma) \geq C_{AF,lower}(\Gamma) = \max I(X; Y|S), \tag{22}$$

where maximization is over joint probability distribution,

$$P_{S,A,X,Y}(s, a, x, y) = \pi_S(s) P_{A|S}(a|s) \times P_{X|A}(x|a) P_{Y|X,S}(y|x, s),$$

where $\mathsf{E}[\Lambda(A)] \leq \Gamma$.

Proof: Omitted. (refer to Theorem 15 in [9]) ■

As an example under this setting with $\Lambda(a) = a$, $a \in \{0, 1\}$, Fig. 5. models the scenario of cost constrained *one-bit active feedback* in the given FSC. The plot for $\alpha = \beta = \delta = 0.5$ is shown in Fig. 6.

VII. Conclusion

In this paper, we studied systems with finite state channels (FSCs), where the encoder and decoder adaptively decide what to feed back from the decoder to encoder to optimize for the rate of reliable communication, under an average cost constraint. For no ISI, stationary indecomposable FSCs, we have the exact characterization of the capacity. We also discuss and compute capacities for the special cases of *to feed or not to feed back* and of *coding on the backward link* for FSCs.

Acknowledgment

This work is supported by Stanford Graduate Fellowship and NSF Grants CCF-1049413 and 4101-38047.

References

[1] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.
[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "Proof of shannon's transmission theorem for finite-state indecomposable channels," *The Annals of Mathematical Statistics*, vol. 29, no. 4, pp. 1209–1220, 1958.
[3] H. H. Permuter, T. Weissman, and A. J. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 644–662, 2009.
[4] G. Kramer, "Directed information for channels with feedback," PhD Dissertation, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1998.
[5] H. Permuter, P. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3150 –3165, Jul. 2008.
[6] J. Massey, "Causality, feedback and directed information," in *Proc. 1990 Symp. Information Theory and Its Applications (ISITA-90), Waikiki, HI*, Nov. 1990, pp. 303–305.
[7] H. Marko, "The bidirectional communication theory–a generalization of information theory," *IEEE Transactions on Communications*, vol. COM-21, no. 12, pp. 1335–1351, Dec. 1973.
[8] S. C. Tatikonda, "Control under communication constraints," PhD Dissertation, MIT, Cambridge, MA, 2000.
[9] H. Asnani, H. H. Permuter, and T. Weissman, "To Feed or Not to Feed Back," *submitted to IEEE Transactions on Information Theory*, November 2010.
[10] H. H. Permuter and T. Weissman, "Source coding with a side information 'vending machine' at the decoder," in *ISIT'09: Proceedings of the 2009 IEEE international conference on Symposium on Information Theory*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1030–1034.
[11] T. Weissman, "Capacity of channels with action-dependent states," in *ISIT'09*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1794–1798.
[12] H. Asnani, H. H. Permuter, and T. Weissman, "Probing capacity," *submitted to IEEE Transactions on Information Theory*, October 2010.