A Lower Bound on the Error Probability of Block-codes with Feedback and Finite Memory

Giacomo Como
Massachusetts Institute of Technology
Laboratory for Information and Decision Systems

Barış Nakiboğlu
Massachusetts Institute of Technology
Research Laboratory of Electronics at MIT

***Abstract*—A lower bound bound is established on the error probability of fixed-length block-coding systems with finite memory feedback, which can be described in terms of a time dependent finite state machine. It is shown that the reliability function of such coding systems over discrete memoryless channels is upper-bounded by the sphere-packing exponent.**

I. Introduction

Although feedback is effective in reducing latency as well as complexity of coding systems, and in improving the capacity of certain channels with memory, there is a long history of negative results for transmission over discrete memoryless channels (DMCs) with feedback: After Shannon proved that feedback does not increase the capacity on DMCs [5], it was shown that *fixed-length block-coding with feedback* does not allow one to beat the *sphere-packing bound* on symmetric DMCs [2]. Whether such a result continues to hold for *non-symmetric DMCs* is a long-standing conjecture: An upper bound on the reliability function is given by the Haroutunian exponent [3], which is typically larger than the sphere-packing exponent on non-symmetric DMCs. In [7], the aforementioned conjecture was claimed to be proved, but the proposed proof appears to suffers from major gaps.

In the present paper, we shall be concerned with fixed-length block-coding over DMCs with *finite memory feedback*. In particular, we shall consider the case when the feedback information can only be stored by a time inhomogeneous finite-state machine whose state is updated each time with the channel output. Under some mild technical assumptions, we shall prove that the reliability function is upper-bounded by the sphere-packing exponent of the channel.

The proof we present in this paper partially follows the line of reasoning of [7], complementing the two major gaps therein with measure concentration, mixing, and fixed-composition arguments, made possible by the finite memory assumption. Our results may be thought as complementary to those in [4], where a lower bound on the error probability of block-coding schemes with delayed feedback has been derived.

The remainder of this paper is organized as follows. In Sect. II we describe the transmission model, with the finite memory restriction on the feedback encoders, and state our main result as Theorem 1. Then, in Sect. III we derive a lower bound to the error probability via a change of measure argument, using Holder's inequality, and give a brief discussion on how the sphere-packing bound can established when there is no feedback. In Sect. IV, we study the mixing properties of certain Markov chains in order to make a similar measure change argument for the encoding schemes satisfying our assumption. In Sect. V we combine the results of section III and IV using a method of types argument, and complete the proof of Theorem 1.

II. Model and Main Result

We consider a DMC with finite input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $W(y|x)$ such that

$$\delta_W := \min\{W(y|x) : x \in \mathcal{X}, y \in \mathcal{Y}\} > 0 \qquad (1)$$

The sphere-packing exponent of the DMC is given by

$$E_{sp}(R) := \max_{\rho\geq 0} E_0(\rho) - \rho R \qquad R \geq 0$$
$$E_0(\rho) := \max_{P(\cdot)} -\ln \sum\nolimits_y \Big(\sum\nolimits_x P(x) W^{\frac{1}{1+\rho}}(y|x)\Big)^{1+\rho}$$

where $P(\cdot)$ is probability distribution over $\mathcal{X}$.

A fixed-length block-coding system with feedback of rate R, and length $\mathbf{n}$, consists of: a message set $\mathcal{M}$ of cardinality $|\mathcal{M}| = \exp(\mathbf{n}R)$; a sequence of encoding functions

$$\phi_t : \mathcal{M} \times \mathcal{Y}^{t-1} \to \mathcal{X}\,, \qquad 1 \leq t \leq \mathbf{n}\,; \qquad (2)$$

and a decoder $\Psi : \mathcal{Y}^{\mathbf{n}} \to \mathcal{M}$. Any feedback encoder as in (2), induces a joint probability distribution on $\mathcal{M} \times \mathcal{Y}^{\mathbf{n}}$, given by[1]

$$w(m, y^{\mathbf{n}}) := |\mathcal{M}|^{-1} \prod\nolimits_{t=1}^{\mathbf{n}} W(y_t|\phi_t(m, y_1^{t-1}))\,. \qquad (3)$$

Then, the error probability can be written as $P_e := w(\mathcal{M}_e)$, where $\mathcal{M}_e := \{(m, y^n) : \Psi(y^{\mathbf{n}}) \neq m\}$.

In this paper we shall consider a particular class of feedback encoders with finite memory for which the encoding function at time t depends on the past channel outputs y_1^{t-1}, only through a state $s_t \in \mathcal{S}$ which is updated using a time-dependent finite-state machine, i.e.,

$$\phi_t(m, y_1^{t-1}) = \Phi_t(m, s_t)\,, \qquad s_{t+1} = \Gamma_t(s_t, y_t) \qquad (4)$$

for some $\Phi_t : \mathcal{M} \times \mathcal{S} \to \mathcal{X}$, $\Gamma_t : \mathcal{S} \times \mathcal{Y} \to \mathcal{S}$. To denote the state transitions over multiple time units we use the shorthand

$$\begin{aligned} s_{t+1} &= \Gamma_u^t(s_u, y_u^t) \\ &:= \Gamma_t(\Gamma_{t-1}(\ldots\Gamma_{u+1}(\Gamma_u(s_u, y_u), y_{u+1})\ldots, y_{t-1}), y_t)\,. \end{aligned}$$

Notice that, if one allows $\mathcal{S}$ to be infinite, then any feedback encoder as in (2) can be easily represented in the form (4). In contrast, assuming –as we shall– that (4) for some finite

[1]Throughout, for $u \leq t$, we shall use the notation $y_u^t := \{y_u, \ldots, y_t\}$.

$\mathcal{S}$ induces a real constraint. We shall also assume that there exists $k \geq 1$ such that

$$\forall t \geq 1,\, \forall i, j \in \mathcal{S}, \quad \exists y_{t+1}^{t+k} \in \mathcal{Y}^k : \quad \Gamma_{t+1}^{t+k}(i, y_{t+1}^{t+k}) = j. \quad (5)$$

The condition above ensures that effect of past channel outputs whiters away fast enough. Observe, that (4) and (5) are naturally satisfied when $\phi_t(m, y_1^{t-1}) = \phi_t(m, y_{t-k}^{t-1})$, i.e. when the transmitter uses only the latest k channel outputs. Indeed, it is sufficient to choose $\mathcal{S} = \mathcal{Y}^k$ and $\Gamma_t(y_{t-k}^{t-1}, y_t) = y_{t-k+1}^t$. We use this fact to establish the following bound on the error probability of finite memory feedback transmission systems.

Theorem 1: For any rate R, length $\mathbf{n}$ blockcode with feedback of the form (4), satisfying (5) on a DMC satisfying (1)

$$P_e \geq e^{-\mathbf{n}(E_{sp}(R-\epsilon(\ell))+\epsilon(\ell))} \qquad \forall \ell = 1, 2, \ldots, \mathbf{n} \quad (6)$$

where

$$\epsilon(\ell) = \frac{2k \ln \frac{e}{\delta_W}}{\ell \delta_W^k} + \frac{\ell(\ln 4 + |\mathcal{S}| \ln |\mathcal{X}|)}{\mathbf{n}} + \frac{e^{|\mathcal{S}|\ell(|\mathcal{Y}| \ln |\mathcal{S}| + \ln |\mathcal{X}|)} \ln(1+\frac{\mathbf{n}}{\ell})}{\mathbf{n}}.$$

III. A First Lower Bound on the Error Probability

Let $V(\cdot|\cdot)$ be the transition probabilities of a DMC with input alphabet $\mathcal{X}$, and output alphabet $\mathcal{Y}$, and let $Q(\cdot)$ be a probability distribution over $\mathcal{Y}$, such that $V(\cdot|\cdot)$ is absolutely continuous with respect to both W and Q.[2] For $\mathcal{A} \subseteq \mathcal{M}$, define the sets $\mathcal{A}_e$ and $\mathcal{A}_c$ as

$$\mathcal{A}_e := \{(m, y^{\mathbf{n}}) : m \in \mathcal{A} : \Psi(y^{\mathbf{n}}) \neq m\}$$
$$\mathcal{A}_c := \{(m, y^{\mathbf{n}}) : m \in \mathcal{A} : \Psi(y^{\mathbf{n}}) = m\}.$$

Let the probability distributions $q(\cdot)$ and $v(\cdot)$ over $\mathcal{M} \times \mathcal{Y}^{\mathbf{n}}$ be

$$q(m, y^{\mathbf{n}}) = \frac{\mathbb{1}(m \in \mathcal{A})}{|\mathcal{A}|} \prod\nolimits_{1 \leq t \leq \mathbf{n}} Q(y_t) \quad (7a)$$
$$v(m, y^{\mathbf{n}}) = \frac{\mathbb{1}(m \in \mathcal{A})}{|\mathcal{A}|} \prod\nolimits_{1 \leq t \leq \mathbf{n}} V(y_t|\phi_j(m, y^{t-1})). \quad (7b)$$

Let $\mathbf{E}_v[\cdot]$ be the expectation under $v(\cdot)$. The following result holds for feedback encoders not necessarily satisfying (4).

Lemma 1: For all $\beta > 0$, and $\mathcal{A} \subseteq \mathcal{M}$,

$$w(\mathcal{A}_e) \geq v(\mathcal{A}_e)^{\frac{1+\beta}{\beta}} \frac{|\mathcal{A}|}{|\mathcal{M}|} \mathbf{E}_v \left[\prod\nolimits_{t=1}^{\mathbf{n}} e^{\beta \ln \frac{v(y_t|m, y^{t-1})}{w(y_t|m, y^{t-1})}} \right]^{\frac{-1}{\beta}} \quad (8a)$$

$$v(\mathcal{A}_e) \geq 1 - \left[\frac{|\mathcal{M}|}{|\mathcal{A}|} \right]^{\frac{\beta}{1+\beta}} \mathbf{E}_v \left[\prod\nolimits_{t=1}^{\mathbf{n}} e^{\beta(\ln \frac{v(y_t|m, y^{t-1})}{q(y_t)} - R)} \right]^{\frac{1}{1+\beta}} \quad (8b)$$

Proof: Using $w(\cdot)$ and $v(\cdot)$ given in (3), (7) and the reverse Holder's inequality, one gets

$$\begin{aligned}
&\mathbf{E}_v \left[\mathbb{1}(\mathcal{A}) e^{\beta \ln \frac{v(y^{\mathbf{n}}|m)}{w(y^{\mathbf{n}}|m)}} \right] \\
&= \frac{|\mathcal{A}|^\beta}{|\mathcal{M}|^\beta} \sum \mathbb{1}(\mathcal{A}) v(m, y^{\mathbf{n}})^{1+\beta} w(m, y^{\mathbf{n}})^{-\beta} \\
&\geq \frac{|\mathcal{A}|^\beta}{|\mathcal{M}|^\beta} \sum \mathbb{1}(\mathcal{A}_e) v(m, y^{\mathbf{n}})^{1+\beta} w(m, y^{\mathbf{n}})^{-\beta} \\
&\geq \frac{|\mathcal{A}|^\beta}{|\mathcal{M}|^\beta} \left[\sum \mathbb{1}(\mathcal{A}_e) v(m, y^{\mathbf{n}}) \right]^{1+\beta} \left[\sum \mathbb{1}(\mathcal{A}_e) w(m, y^{\mathbf{n}}) \right]^{-\beta} \\
&= \frac{|\mathcal{A}|^\beta}{|\mathcal{M}|^\beta} v(\mathcal{A}_e)^{1+\beta} w(\mathcal{A}_e)^{-\beta} \quad (9)
\end{aligned}$$

[2] i.e. $V(y|x) = 0$ whenever $W(y|x) = 0$ or $Q(y) = 0$.

Following similar steps one can prove that

$$\mathbf{E}_v \left[\mathbb{1}(\mathcal{A}_c) e^{\beta \ln \frac{v(y^{\mathbf{n}}|m)}{q(y^{\mathbf{n}}|m)}} \right] \geq v(\mathcal{A}_c)^{1+\beta} q(\mathcal{A}_c)^{-\beta} \quad (10)$$

Then the lemma follows equations (9), (10) and the observations $q(\mathcal{A}_c) \leq |\mathcal{A}|^{-1}$ and $v(\mathcal{A}_e) = 1 - v(\mathcal{A}_c)$. ∎

Note that equations (8a) and (8b) bound the error probability from below. We shall show in the following sections how Lemma 1 leads to Theorem 1. In order to introduce some of the ideas of that proof let us show how Lemma 1 can be used to establish a lower bound on the error probability in the case without feedback.

Theorem 2: For any length $\mathbf{n}$ block code error probability is lower bounded as

$$P_e \geq e^{-\mathbf{n}(E_{sp}(R-\epsilon_1)+\epsilon_1)} \quad (11)$$

where $\epsilon_1 = \frac{(\ln \frac{e}{\delta_W})^2 \sqrt{\mathbf{n}} + |\mathcal{X}| \ln(\mathbf{n}+1)}{\mathbf{n}}$.

Proof: Let $Q(\cdot) = Q_\rho(\cdot)$ be in a parametric form to be specified later. Let V for $v(\cdot)$ be

$$V_\rho(y|x) = \frac{W^{\frac{1}{1+\rho}}(y|x) Q_\rho^{\frac{\rho}{1+\rho}}(y)}{r_\rho(x)} \quad (12)$$

where $r_\rho(x) = \sum_{\tilde{y}} W^{\frac{1}{1+\rho}}(\tilde{y}|x) Q_\rho^{\frac{\rho}{1+\rho}}(\tilde{y})$.

Given m, the y_t's are mutually independent. Thus,

$$\mathbf{E}_v \left[\prod\nolimits_{t=1}^{\mathbf{n}} e^{\beta \ln \frac{v(y_t|m, y^{t-1})}{w(y_t|m, y^{t-1})}} \middle| m \right] = \prod\nolimits_{t=1}^{\mathbf{n}} \mathbf{E}_v \left[e^{\beta \ln \frac{v(y_t|m)}{w(y_t|m)}} \middle| m \right] \quad (13a)$$

$$\mathbf{E}_v \left[\prod\nolimits_{t=1}^{\mathbf{n}} e^{\beta \ln \frac{v(y_t|m, y^{t-1})}{q(y_t)}} \middle| m \right] = \prod\nolimits_{t=1}^{\mathbf{n}} \mathbf{E}_v \left[e^{\beta \ln \frac{v(y_t|m)}{q(y_t)}} \middle| m \right] \quad (13b)$$

If we could upper bound $\mathbf{E}[e^z]$ by $e^{\mathbf{E}[z]}$ then we could use equations (8) and (13) to lower bound the error probability but Jensen's inequality works in the opposite direction, i.e. $\mathbf{E}[e^z] \geq e^{\mathbf{E}[z]}$. Thus we need to make an approximation. If $z - \mathbf{E}[z] \leq 1$ with probability one then[3]

$$\mathbf{E}[e^z] \leq e^{\mathbf{E}[z] + \mathbf{E}\left[(z - \mathbf{E}[z])^2\right]} \quad (14)$$

Thus if $\beta \leq -\ln \min\{\delta_W, \delta_{Q_\rho}\}$ for all m and t, the bound[4]

$$\mathbf{E}_v \left[\left(\ln \frac{v(y_t|m)}{w(y_t|m)} - \mathbf{E}_v \left[\ln \frac{v(y_t|m)}{w(y_t|m)} \middle| m \right] \right)^2 \middle| m \right] \leq (\ln \tfrac{1}{\delta_Q})^2 + \tfrac{4}{e^2}$$

implies that

$$\mathbf{E}_v \left[e^{\beta \ln \frac{v(y_t|m)}{w(y_t|m)}} \middle| m \right] \leq e^{\beta \mathcal{D}(V_\rho \| W | x_t(m)) + \beta^2 [[\ln \frac{1}{\delta_W}]^2 + \frac{4}{e^2}]} \quad (15a)$$

$$\mathbf{E}_v \left[e^{\beta \ln \frac{v(y_t|m)}{q(y_t)}} \middle| m \right] \leq e^{\beta \mathcal{D}(V_\rho \| Q_\rho | x_t(m)) + \beta^2 [[\ln \frac{1}{\delta_Q}]^2 + \frac{4}{e^2}]} \quad (15b)$$

where $x_t(m)$ is the input letter for message m at time t.

For any block-length $\mathbf{n}$ there are $\binom{\mathbf{n}+|\mathcal{X}|-1}{|\mathcal{X}|-1} \leq (\mathbf{n}+1)^{|\mathcal{X}|}$ empirical types. Thus if we choose messages of the most populous type, say $P(\cdot)$, to be $\mathcal{A}$ we get, $|\mathcal{A}| \geq \frac{|\mathcal{M}|}{(\mathbf{n}+1)^{|\mathcal{X}|}}$. Thus using Lemma 1 and equations (13) and (15) we get,

$$w(\mathcal{A}_e) \geq v(\mathcal{A}_e)^{\frac{1+\beta}{\beta}} e^{-\mathbf{n}(\mathcal{D}(V_\rho \| W | P) + \epsilon_2(\beta, \mathbf{n}))} \quad (16a)$$

$$v(\mathcal{A}_e) \geq 1 - e^{\frac{\beta \mathbf{n}}{1+\beta}(\mathcal{D}(V_\rho \| Q_\rho | P) - \epsilon_2(\beta, \mathbf{n}) - R)}. \quad (16b)$$

[3] Using techniques similar to those in [1], see Appendix A for details.

[4] The details of the bound on the variance are presented in Appendix B.

where $\epsilon_2(\beta,\mathbf{n}) = \beta[(\ln \frac{1}{\min\{\delta_W, \delta_{Q_\rho}\}})^2 + \frac{4}{e^2}] + \frac{|\mathcal{X}|\ln(1+\mathbf{n})}{\mathbf{n}}$.

There exists a parametric family of $Q_\rho(\cdot)$'s which is continuous in ρ, see [6] or appendix D for details, such that,

$$\ln r_\rho(x) \geq -\frac{E_0(\rho)}{1+\rho} \tag{17}$$

Furthermore when all entries of $W(\cdot|\cdot)$ is positive, $\delta_{Q_\rho} \geq \delta_W$ for all ρ. Using equations (12) and (17) we get,

$$\mathcal{D}\left(V_\rho \| W | P\right) \leq E_0(\rho) - \rho \mathcal{D}\left(V_\rho \| Q_\rho | P\right) \qquad \forall P, \rho. \tag{18}$$

Note that for any P,

$$\text{either} \quad \mathcal{D}\left(V_\rho \| Q_\rho | P\right)|_{\rho=0} \leq R - \epsilon_2(\beta,\mathbf{n}) - \tfrac{(1+\beta)\ln 2}{\beta \mathbf{n}} \quad \text{(a)}$$

$$\text{or} \quad \mathcal{D}\left(V_\rho \| Q_\rho | P\right)|_{\rho=0} > R - \epsilon_2(\beta,\mathbf{n}) - \tfrac{(1+\beta)\ln 2}{\beta \mathbf{n}} \quad \text{(b)}$$

If (a) is the case: using (12) and (16) at $\rho = 0$ we get

$$w(\mathcal{A}_e) \geq (\tfrac{1}{2})^{\frac{1+\beta}{\beta}} e^{-\mathbf{n}\epsilon_2(\beta,\mathbf{n})} \tag{19}$$

If (b) is the case: note that $\lim_{\rho\to\infty} \mathcal{D}\left(V_\rho \| Q_\rho | P\right) = 0 \ \forall P$. Thus, by the intermediate value theorem, there exist some ρ_P^* such that $\mathcal{D}\left(V_\rho \| Q_\rho | P\right)|_{\rho=\rho_P^*} = R - \epsilon_2(\beta,\mathbf{n}) - \frac{(1+\beta)\ln 2}{\beta\mathbf{n}}$. Using equations (16) and (18) at $\rho = \rho_P^*$ together with the fact that $E_{sp}(R) \geq E_0(\rho) - \rho R \ \forall \rho \geq 0$ we get

$$w(\mathcal{A}_e) \geq (\tfrac{1}{2})^{\frac{1+\beta}{\beta}} e^{-\mathbf{n}[E_{sp}\left(R - \epsilon_2(\beta,\mathbf{n}) - \frac{(1+\beta)\ln 2}{\beta\mathbf{n}}\right) + \epsilon_2(\beta,\mathbf{n})]} \tag{20}$$

Equation (11) follows equations (19) and (20) and the identity $(\ln \frac{1}{\delta_W})^2 + \frac{4}{e^2} + 2\ln 2 \leq (\ln \frac{e}{\delta_W})^2$ by setting $\beta = \mathbf{n}^{-1/2}$. ∎

Notice that, when there is feedback, the input letter at any time t for any message m depends on the previous channel outputs. Thus, we can not

- claim conditional independence of Y_t's or equation (13).
- make an expurgation over types

However for the particular encoding schemes satisfying the assumption 5 we can address both of the issues. For doing that we need to analyze the mixing properties of Markov chains resulting from Γ and Φ for each $m \in \mathcal{M}$ under $v(\cdot)$.

IV. Finite State Machine Encoders and Mixing

Let $v(\cdot)$ and $q(\cdot)$ be of the form given in (7) for some channel output probability distribution $Q(\cdot)$, and some transition probabilities $V(\cdot|\cdot)$. Then, for encoding schemes with feedback we can not write (13), because the channel outputs are not conditionally independent given the message. However, when (5) holds, the dependence between y_t and y_u vanishes as $t-u$ increases. Lemma 2 below uses this property to bound the terms $\mathbf{E}_v[\exp(\beta\mu_u^t)|\, m, s_u]$ and $\mathbf{E}_v[\exp(\beta\nu_u^t)|\, m, s_u]$, where

$$\mu_u^t := \ln \tfrac{v(y_u^t|m,s_u)}{w(y_u^t|m,s_u)} \qquad \nu_u^t := \ln \tfrac{v(y_u^t|m,s_u)}{q(y_u^t|m,s_u)} \tag{21}$$

The upper bound provided is in terms of the empirical type

$$P_{s^*}(x) := \tfrac{1}{t-u+1} \sum_{j=u}^{t} v(\Phi_j(m,s_j) = x | m, s_u = s^*) \tag{22}$$

for any $s^* \in \mathcal{S}$. Observe that, in contrast to the case without feedback, the type $P_{s^*}(\cdot)$ depends on the particular $V(\cdot|\cdot)$ that is used. However, this will not prevent us from applying of the intermediate value theorem because of the continuity of the term $\mathcal{D}\left(V \| W | P\right)$ in P.

Lemma 2: For any feedback encoder of the form (4), satisfying (5), $u \leq t$, $s^* \in \mathcal{S}$, $\beta \in (0, \frac{\ln \max\{\delta_W^{-1}, \delta_Q^{-1}\}}{t-u+1}]$,

$$\mathbf{E}_v\left[\exp(\beta\mu_u^t)\middle|\, m, s_u\right] \leq e^{(t-u+1)\beta(\mathcal{D}(V\|W|P_{s^*})+\epsilon_3)} \tag{23a}$$

$$\mathbf{E}_v\left[\exp(\beta\nu_u^t)\middle|\, m, s_u\right] \leq e^{(t-u+1)\beta(\mathcal{D}(V\|Q|P_{s^*})+\epsilon_3)} \tag{23b}$$

where $\epsilon_3 := \frac{k\ln\delta_W^{-1} + e^{-1}}{\delta_V^{2k}} \left[(t-u+1)^{-1} + 2\beta\ln(e\delta_W^{-1})\right]$.

Proof: If $z - \mathbf{E}[z] \leq 1$ with probability one, then[5]

$$\mathbf{E}[\exp(z)] \leq \exp(\mathbf{E}[z] + \mathbf{E}\left[(z - \mathbf{E}[z])^2\right]).$$

Hence, if $\beta \leq -(t-u+1)^{-1}\ln\delta_W$, then

$$\mathbf{E}_v\left[\exp(\beta\mu_u^t)\middle|\, m, s_u\right] \leq \exp\left[\beta\mathbf{E}_v\left[\mu_u^t\middle|\, m, s_u\right] + \epsilon_4\right], \tag{24}$$

where[6] $\epsilon_4 = 2(t-u+1)\beta^2\delta_V^{-2k}(k\ln\delta_W^{-1} + e^{-1})\ln(e\delta_W^{-1})$. Now, we consider $\mathbf{E}_v[\mu_u^t|\, m, s_u]$ and bound its dependence on s_u. Observe that, conditioned on m and s_u, the state sequence s_u^t forms a Markov chain on $\mathcal{S}$ whose time-dependent transition probabilities are given by

$$\Pi_j(s_{j+1}|s_j) := \sum_{y:\Gamma_j(s_j,y)=s_{j+1}} V(y|\Phi(m,s_j)) \qquad \forall u \leq j < t.$$

Let us consider a copy of this Markov chain, $\tilde{s}_u^t$, which starts at time u in $\tilde{s}_u = s^*$, evolves independently from s_u^t according to the same transition kernel $\Pi_j(\cdot|\cdot)$ until the first time they meet, and sticks to it thenceforth. Let $\tilde{\mu}_u^t$ be defined as in (21) with s^* replacing s_u, and notice that

$$\mathbf{E}_v\left[\tilde{\mu}_u^t\middle|\, m, s_u\right] = (t-u+1)D(V||W|P_{s^*}) \tag{25}$$

Moreover as a result of assumption (5) whenever $s_t \neq \tilde{s}_t$, $\forall y_t^{t+k}$, there exist at least one sequence of $\tilde{y}_t^{t+k}$'s such that $\Gamma_t^{t+k}(s_t, y_t^{t+k}) = \Gamma_t^{t+k}(\tilde{s}_t, \tilde{y}_t^{t+k})$. Thus whenever $s_t \neq \tilde{s}_t$ for any encoding scheme Φ and any t

$$\mathbf{P}_v\{\tilde{s}_{t+k} = s_{t+k}|\, m, \tilde{s}_t, s_t\} \geq \delta_V^k$$

As a consequence,

$$\mathbf{P}_v\{\tilde{s}_{u+ik} \neq s_{u+ik}|\, m, s_u\} \leq (1-\delta_V^k)^i, \quad i \geq 0. \tag{26}$$

Using the fact that $\mu_{j+1}^{j+k} \leq k\ln\delta_W^{-1}$, and the inequality $-x\ln x \leq e^{-1}$, one gets, for $\gamma = k\ln\delta_W^{-1} + e^{-1}$

$$\mathbf{E}_v\left[\mu_{j+1}^{j+k} - \tilde{\mu}_{j+1}^{j+k}\middle|\, m, s_u\right] \leq \gamma\mathbb{1}(s_{u+ik} \neq \tilde{s}_{u+ik}) \tag{27}$$

Using equations (26) and (27) , we get

$$\begin{aligned}\mathbf{E}_v\left[\mu_u^t - \tilde{\mu}_u^t\middle|\, m, s_u\right] &\leq \sum\nolimits_{i=0}^{\lfloor (t-u+1)/k\rfloor} \gamma(1-\delta_V^k)^i \\ &\leq \gamma\sum\nolimits_{i\geq 0}(1-\delta_V^k)^i \\ &\overset{(a)}{=} \gamma\delta_V^{-2k}\end{aligned} \tag{28}$$

Using equations (25) and (28)

$$\mathbf{E}_v\left[\mu_u^t\middle|\, m, s_u = i\right] \leq (t-u+1)D(V||W|P) + \gamma\delta_V^{-k} \tag{29}$$

Then, (23a) follows from (24), and (29). Equation (23b) can be derived from a similar discussion. ∎

[5] See Appendix A for details.

[6] Details of the bound on $\mathbf{E}_v\left[(\mu_u^t - \mathbf{E}_v\left[\mu_u^t\middle|\, m, s_u\right])^2\middle|\, m, s_u\right]$ are presented in Appendix C.

V. Super-letters and Fixed Composition Argument

Note that, as a result of Lemma 2 we know that if $(t-u)$ is large enough both $\mathbf{E}_v[\exp(\beta\mu_u^t)|\, m, s_u]$ and $\mathbf{E}_v[\exp(\beta\nu_u^t)|\, m, s_u]$ are bounded independently of s_u. On the other hand iff $(t-u)$ is not too large one can interpret $(t-u)$-long encoding functions together with $(t-u)$-long sequence of Γ_t's as an input letters and make a fixed composition argument to bound $\mathbf{E}_v[\exp(\beta\mu_1^{\mathbf{n}})|\, m, s_1]$ and $\mathbf{E}_v[\exp(\beta\nu_1^{\mathbf{n}})|\, m, s_1]$. The rest of this section is devoted to making this argument precise and establishing the bound given in Theorem 1, using Lemmas 1 and 2.

First note that for any $t \in [1, \mathbf{n}]$ and $m \in \mathcal{M}$, Γ_t is a mapping form $\mathcal{S}\times\mathcal{Y}$ to $\mathcal{S}$, i.e. an element of $\mathcal{S}^{\mathcal{S}\times\mathcal{Y}}$ and $\Phi_t(m)$ is a mapping form $\mathcal{S}$ to $\mathcal{X}$, i.e. an element of $\mathcal{X}^{\mathcal{S}}$. Thus for any $m \in \mathcal{M}$ any ℓ-long part of Γ and $\Phi(m)$, say $(\Gamma_{t+1}^{t+\ell}, \Phi_{t+1}^{t+\ell}(m))$ is an element of $\mathcal{Z} = \left(\mathcal{S}^{\mathcal{S}\times\mathcal{Y}} \times \mathcal{X}^{\mathcal{S}}\right)^{\ell}$.

If we interpret ℓ-long parts of the encoding function and finite state machine, $(\Gamma_{i\ell+1}^{(i+1)\ell}, \Phi_{i\ell+1}^{(i+1)\ell}(m))$, as super letters, $\mathbf{z}_i$'s for $i = 0, 1, \ldots (\lfloor \frac{\mathbf{n}}{\ell} \rfloor - 1)$ then the codeword for a message $m \in \mathcal{M}$ is composed of $\lfloor \frac{\mathbf{n}}{\ell} \rfloor$ super-letters and an $(\mathbf{n} - \lfloor \frac{\mathbf{n}}{\ell} \rfloor \ell)$ long extension. Including different extensions there are less than $(\lfloor \frac{\mathbf{n}}{\ell} \rfloor + 1)^{|\mathcal{Z}|} |\mathcal{X}|^{|\mathcal{S}|\ell}$ different types. Thus if we choose $\mathcal{A}$ to be the most populous type we will have

$$\frac{|\mathcal{A}|}{|\mathcal{M}|} \geq \exp(-(|\mathcal{S}|^{\ell|\mathcal{S}||\mathcal{Y}|}|\mathcal{X}|^{\ell|\mathcal{S}|}) \ln(1+\tfrac{\mathbf{n}}{\ell}) - |\mathcal{S}|\ell \ln|\mathcal{X}|) \quad (30)$$

Note that codewords of the message in set $\mathcal{A}$ differ only in first $\mathbf{n}_1 = \ell\lfloor \frac{\mathbf{n}}{\ell} \rfloor$ time instances. Although ordering of these super letters will effect the actual value of $\mathbf{E}_v\left[\exp(\beta \ln \frac{v(y^{\mathbf{n}_1}|m)}{w(y^{\mathbf{n}_1}|m)})\middle|\, m\right]$ and $\mathbf{E}_v\left[\exp(\beta \ln \frac{v(y^{\mathbf{n}_1}|m)}{q(y^{\mathbf{n}_1})})\middle|\, m\right]$ we can bound all of those expectations in way that is independent of the ordering using lemma 2. Consequently $\beta \in [0, \frac{-\ln\min\{\delta_W, \delta_Q\}}{\ell}]$,

$$\mathbf{E}_v\left[\exp(\beta \ln \tfrac{v(y^{\mathbf{n}_1}|m)}{w(y^{\mathbf{n}_1}|m)})\middle|\, m\right] = e^{\beta\mathbf{n}_1(\mathcal{D}(V\|W|P_m)-\epsilon_5)} \quad (31a)$$

$$\mathbf{E}_v\left[\exp(\beta \ln \tfrac{v(y^{\mathbf{n}_1}|m)}{q(y^{\mathbf{n}_1})})\middle|\, m\right] = e^{\beta\mathbf{n}_1(\mathcal{D}(V\|Q|P_m)-\epsilon_5)} \quad (31b)$$

where $\epsilon_5 = \frac{(k\ln\frac{1}{\delta_W}+e^{-1})}{\delta_V^{2k}}[\frac{1}{\ell} + 2\beta\ln\frac{e}{\delta_W}]$ and

$$P_m(x) = \frac{\sum_{i=0}^{\frac{\mathbf{n}_1}{\ell}-1}\sum_{t=1}^{\ell}\mathbf{P}_v\{\Phi_{i\ell+t}(m,s_{i\ell+t})=x|m,s_{i\ell+1}=j\}}{\mathbf{n}_1} \quad (32)$$

Note that unlike P_m in the non-feedback case P_m in equation (32) depends on V. However that dependence is continuous in V. Furthermore P_m is identical for all messages in a given type, no matter what V is.

As we did in the non-feedback case we will use V_ρ and Q_ρ and use equation (18). Using equations (30) and (31) together with corollary 1 we get,

$$w(\mathcal{A}_e) \geq v(\mathcal{A}_e)^{\frac{1+\beta}{\beta}} e^{-\mathbf{n}_1(\mathcal{D}(V_\rho\|W|P_m)+\epsilon_6)} \quad (33a)$$

$$v(\mathcal{A}_e) \geq 1 - e^{\frac{\beta\mathbf{n}_1}{1+\beta}(\mathcal{D}(V_\rho\|Q_\rho|P_m)+\epsilon_6-\frac{\mathbf{n}}{\mathbf{n}_1}R)} \quad (33b)$$

where $\epsilon_6 = \epsilon_5 + \frac{(|\mathcal{S}|^{\ell|\mathcal{S}||\mathcal{Y}|}|\mathcal{X}|^{\ell|\mathcal{S}|})\ln(1+\frac{\mathbf{n}_1}{\ell})+|\mathcal{S}|\ell\ln|\mathcal{X}|}{\mathbf{n}_1}$.

Note that both V_ρ and Q_ρ are continuous functions of ρ, then for any type P_m is continuous function of ρ. Consequently $\mathcal{D}(V_\rho\|\, Q_\rho|P_m)$ is continuous in ρ. Thus either $\mathcal{D}(V_\rho\|\, Q_\rho|P_m)|_{\rho=0}$ either less than or strictly greater than $\frac{\mathbf{n}}{\mathbf{n}_1}R - \epsilon_6 + \frac{(1+\beta)\ln 2}{\beta\mathbf{n}_1}$. Using exact same reasoning with the non-feedback case we get, $\beta \in [0, \frac{-\ln\min\{\delta_W,\delta_Q\}}{\ell}]$,

$$w(\mathcal{A}_e) \geq (\tfrac{1}{2})^{\frac{1+\beta}{\beta}} e^{-\mathbf{n}_1[E_{sp}(\frac{\mathbf{n}}{\mathbf{n}_1}R-\epsilon_6-\frac{(1+\beta)\ln 2}{\beta\mathbf{n}_1})+\epsilon_6]} \quad (34)$$

Note that when all entries of W are positive $\delta_{Q_\rho} \geq \delta_W$ and $\delta_{V_\rho} \geq \delta_W$. Using equation (34) and setting $\beta = \frac{1}{2\ell(1-\ln\delta_W)}$ and using the fact that $E_{sp}(\cdot)$ is decreasing function of its argument we recover equation (6).

VI. Discussion

In this paper, we have proved a lower bound on the error probability of fixed-length block-codes with finite state machine encoders over discrete memoryless channels with feedback. We have shown that, when the transmitter is only allowed use the state of finite state machine in its encoding, where the state of the finite state machine is updated as channel outputs are fed back to it, the sphere-packing bound continues to hold even on non-symmetric DMCs. Ongoing work includes relaxing some of the technical assumptions, and extending our results to channels with memory.

Acknowledgments

The work leading to this manuscript has grown out of several conversations between the authors and Harikrishna R. Palaiyanur and Prof. Anant Sahai of Berkley, Prof. Sekhar Tatikonda of Yale, and Prof. Sanjoy Mitter of MIT.

Appendix

A. $\mathbf{E}[\exp(z)]$ *vs* $\exp(\mathbf{E}[z])$:

Let $g(z) = 2\frac{\exp(z)-1-z}{z^2}$ than one can show that

$$g(z) \geq 0 \qquad g'(z) \geq 0 \qquad g''(z) \geq 0$$

Let z be r.v. such that $z - \mathbf{E}[z] \leq 1$ then,

$$\begin{aligned}\mathbf{E}[\exp(z-\mathbf{E}[z])] &= \mathbf{E}\left[1 + z - \mathbf{E}[z] + \tfrac{(z-\mathbf{E}[z])^2 g(z-\mathbf{E}[z])}{2}\right] \\ &\leq 1 + \mathbf{E}\left[\tfrac{(z-\mathbf{E}[z])^2 g(1)}{2}\right] \\ &\leq \exp(\tfrac{\mathbf{E}[(z-\mathbf{E}[z])^2 g(1)]}{2}).\end{aligned}$$

Using $g(1) \leq 1$ we get,

$$\mathbf{E}[\exp(z)] \leq \exp(\mathbf{E}[z] + \mathbf{E}[z^2] - \mathbf{E}[z]^2). \quad (35)$$

B. Bounding $\sum_y f_y(\ln\frac{f_y}{g_y})^2 - (\sum_y f_y \ln\frac{f_y}{g_y})^2$:

Let f_y and g_y be two probability distributions on $\mathcal{Y}$ then

$$\begin{aligned}\sum_y f_y(\ln\tfrac{f_y}{g_y})^2 &= \sum_y g_y \tfrac{f_y}{g_y}(\ln\tfrac{f_y}{g_y})^2 \\ &\overset{(a)}{\leq} \sum_y g_y\left[\tfrac{f_y}{g_y}(\ln\tfrac{1}{\delta_g})^2 \mathbb{1}(\tfrac{f_y}{g_y} \geq 1) + \mathbb{1}(\tfrac{f_y}{g_y} \leq 1)\tfrac{4}{e^2}\right] \\ &\leq (\ln\tfrac{1}{\delta_g})^2 + \tfrac{4}{e^2} \quad (36)\end{aligned}$$

where in step (a) we used the facts that $f_y \leq 1$, $g_y \geq \delta_g$ and $x(\ln x)^2 \leq \frac{4}{e^2}$ for $x \in [0,1]$. Thus

$$\sum_y f_y(\ln\tfrac{f_y}{g_y})^2 - (\sum_y f_y \ln\tfrac{f_y}{g_y})^2 \leq (\ln\tfrac{1}{\delta_g})^2 + \tfrac{4}{e^2} \quad (37)$$

C. Bounding $\mathbf{E}_v\left[(\mu_u^t - \mathbf{E}_v[\mu_u^t|\, m, s_u])^2 \middle|\, m, s_u\right]$:

Let us denote $\mathbf{E}_v[\mu_t|\, m, s_u]$ by $\overline{\mu_t}$ for brevity.

$$\begin{aligned}\mathbf{E}_v&\left[(\mu_u^t - \overline{\mu_u^t})^2 \middle|\, m, s_u\right] \\ &= \sum_{j=u}^{t} \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})^2 \middle|\, m, s_u\right] \\ &\quad + \sum_{j=u}^{t-1} \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})(\mu_{j+1}^t - \overline{\mu_{j+1}^t}) \middle|\, m, s_u\right] \quad (38)\end{aligned}$$

Note that

$$\begin{aligned}\mathbf{E}_v\left[(\mu_j - \overline{\mu_j})^2 \middle|\, m, s_u\right] &= \mathbf{E}_v\left[{\mu_j}^2 \middle|\, m, s_u\right] \\ &= \mathbf{E}_v\left[\mathbf{E}_v\left[{\mu_j}^2 \middle|\, m, s_t\right] \middle|\, m, s_u\right] \\ &\le (\ln \delta_W^{-1})^2 + 4e^{-2} \quad (39)\end{aligned}$$

the inequality following from (36). For $u \le j < t$, it holds

$$\begin{aligned}&\mathbf{E}_v\left[(\mu_j - \overline{\mu_j})(\mu_{j+1}^t - \overline{\mu_{j+1}^t}) \middle|\, m, s_u\right] \\ &= \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})(\mu_{j+1}^t - \mathbf{E}_v\left[\mu_{j+1}^t \middle|\, m, s_{j+1}\right])\middle|\, m, s_u\right] \\ &\quad + \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})(\mathbf{E}_v\left[\mu_{j+1}^t \middle|\, m, s_{j+1}\right] - \overline{\mu_{j+1}^t}) \middle|\, m, s_u\right] \\ &\overset{(a)}{=} \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})(\mathbf{E}_v\left[\mu_{j+1}^t \middle|\, m, s_{j+1}\right] - \overline{\mu_{j+1}^t}) \middle|\, m, s_u\right] \\ &\overset{(b)}{\le} \mathbf{E}_v\left[(\mu_j - \overline{\mu_j})^2 \middle|\, m, s_u\right]^{\frac{1}{2}} \\ &\quad \cdot \mathbf{E}_v\left[(\mathbf{E}_v\left[\mu_{j+1}^t \middle|\, m, s_{j+1}\right] - \overline{\mu_{j+1}^t})^2 \middle|\, m, s_u\right]^{\frac{1}{2}} \quad (40)\end{aligned}$$

where (a) follows from the Markovian property of the encoding and (b) follows from Schwarz's inequality. In addition, as a result of the Markovian property, we have,

$$\overline{\mu_j^t} = \mathbf{E}_v\left[\mathbf{E}_v\left[\mu_j^t \middle|\, m, s_j\right] \middle|\, m, s_u\right]. \quad (41)$$

Using equation (29) together with equation (41) we get,

$$|\mathbf{E}_v\left[\mu_{j+1}^t \middle|\, m, s_{j+1}\right] - \overline{\mu_j^t}| \le \delta_V^{-2k}(-k \ln \delta_W + e^{-1}). \quad (42)$$

Using equation (38), (39), (40) and (42) and the fact that $\ln^2 \delta_W^{-1} + 4e^{-2} \le \ln^2(e/\delta_W)$ we get,

$$\begin{aligned}\mathbf{E}_v\left[(\mu_u^t - \overline{\mu_u^t})^2 \middle|\, m, s_u\right] &\le (t-u+1)\left(\tfrac{k \ln \delta_W^{-1} + e^{-1}}{\delta_V^{2k}} + \ln \tfrac{e}{\delta_W}\right) \ln \tfrac{e}{\delta_W} \\ &\le (t-u+1)\left(\tfrac{k \ln \delta_W^{-1} + e^{-1}}{\delta_V^{2k}}\right) 2 \ln \tfrac{e}{\delta_W} \quad (43)\end{aligned}$$

D. Uniqueness and Continuity of Q_ρ

Recall that

$$\begin{aligned}e^{-E_0(\rho,Q)} &= \min_x \Big(\sum_y W(y|x)^{\frac{1}{1+\rho}} Q(y)^{\frac{\rho}{1+\rho}}\Big)^{1+\rho} \\ e^{-E_0(\rho)} &= \max_Q e^{-E_0(\rho,Q)}\end{aligned}$$

Note that maximizing Q_ρ satisfies

$$\sum_y W(y|x)^{\frac{1}{1+\rho}} Q_\rho(y)^{\frac{\rho}{1+\rho}} \ge e^{\frac{-E_0(\rho)}{1+\rho}} \qquad \forall x \quad (44)$$

and with equality for some x.

Note that if there are two distinct optimal distributions $Q_{\rho,a}$ and $Q_{\rho,b}$ then

$$\begin{aligned}\sum_y &W(y|x)^{\frac{1}{1+\rho}} (\alpha Q_{\rho,a}(y)^{\frac{\rho}{1+\rho}} + (1-\alpha) Q_{\rho,b}(y)^{\frac{\rho}{1+\rho}}) \\ &< \sum_y W(y|x)^{\frac{1}{1+\rho}} (\alpha Q_{\rho,a}(y) + (1-\alpha) Q_{\rho,b}(y))^{\frac{\rho}{1+\rho}}\end{aligned}$$

All of their linear combinations will lead to a strictly larger $E_0(\rho)$ so they can not be the optimal Q_ρ simultaneously. Thus there exist a unique Q_ρ.

Note that $e^{-E_0(\rho,Q)}$ is a decreasing function for all decreasing in ρ. Because

$$\mathbf{E}\left[x^{\frac{1}{1+\rho}}\right]^{1+\rho} = \mathbf{E}\left[x^{\frac{1}{1+\rho'}\frac{1+\rho'}{1+\rho}}\right]^{\frac{1+\rho}{1+\rho'}(1+\rho')} \ge \mathbf{E}\left[x^{\frac{1}{1+\rho'}}\right]^{1+\rho'}$$

for $\rho' \ge \rho$. Then $e^{-E_0(\rho)}$ is also a decreasing function of ρ. Thus

$$\begin{aligned}0 &\le e^{-E_0(\rho)} - e^{-E_0(\rho+\epsilon)} \le e^{-E_0(\rho,Q_\rho)} - e^{-E_0((\rho+\epsilon),Q_\rho)} \\ 0 &\le e^{-E_0(\rho)} - e^{-E_0(\rho+\epsilon)} \le \delta_1(Q_\rho, \epsilon) \quad (45)\end{aligned}$$

where $\lim_{\epsilon\to 0} \delta_1(Q_\rho, \epsilon) = 0$ and last step follows from the continuity of $e^{-E_0(\rho,Q)}$ in ρ for any Q.

Furthermore

$$\begin{aligned}e^{-E_0(\rho)} - e^{-E_0(\rho+\epsilon)} &\ge e^{-E_0(\rho,Q_\rho)} - e^{-E_0(\rho,Q_{\rho+\epsilon})} \\ &\ge \zeta(||Q_\rho - Q_{\rho+\epsilon}) \quad (46)\end{aligned}$$

where the $\zeta(\rho)$ is strictly increasing function such that $\zeta(0) = 0$. Last step follows from the strict convexity of $-e^{-E_0(\rho,Q)}$ in Q.

Thus as result of equations (45) and (46) we get,

$$||Q_\rho - Q_{\rho+\epsilon}|| \le \zeta^{-1}(\delta_1(Q_\rho, \epsilon)) \quad (47)$$

$\lim_{\epsilon\to 0} \delta_1(Q_\rho, \epsilon) = 0$ and $\zeta^{-1}(\cdot)$ is also a strictly increasing function such that $\zeta^{-1}(0) = 0$. Thus Q_ρ is continuous in ρ.

REFERENCES

[1] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math*, 3(1):79–127, 2006.

[2] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, vol 8:161–168, 1962.

[3] E. A. Haroutunian. A lower bound of the probability of error for channels with feedback. *Problemy Peredachi Informatsii*, vol 13:36–44, 1977.

[4] H. Palaiyanur and A. Sahai. A bound for block codes with delayed feedback related to sphere-packing bound. *preprint*, 2010.

[5] C. Shannon. The zero error capacity of a noisy channel. *IEEE Transactions on Information Theory*, Vol. 2, Iss 3:8–19, 1956.

[6] C.E. Shannon, R.G. Gallager, and E.R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. *Information and Control*, 10, No. 1:65–103, 1967.

[7] A.Yu. Sheverdyaev. Lower bound for error probability in a discrete memoryless channel with feedback. *Problemy Perdachi Informatsii*, 18, No. 4:5–15, 1982.