Errors-and-Erasures Decoding for Block Codes with Feedback

Barış Nakiboğlu Lizhong Zheng
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA, 02139
Email: {nakib, lizhong} @mit.edu

***Abstract*— Fixed length block codes on discrete memoryless channels with feedback are considered for errors and erasures decoding. Upper and lower bounds are derived for the error exponent in terms of the rate and the erasure exponents. In addition the converse result of Burnashev for variable length block codes is extended to include list decoding.**

I. Introduction:

Early results about the use of feedback on discrete memoryless channels (DMCs) in terms conventional performance criteria were negative. Not only the capacity was not increasing with feedback, as Shannon showed in [9], but also the error exponent was not increasing at high rates, [4], [6], at least for symmetric channels. Relaxations like errors-and-erasures decoding or variable length coding was needed to increase the error exponent for those channels with the use of feedback.

Burnashev, [2], considered variable length block codes with feedback, instead of fixed-length ones and obtained the exact expression for the error exponent at all rates.[1] Later Yamamoto and Itoh, [13], suggested a coding scheme which achieves the best error exponent for variable length block codes by using a fixed length block code with errors-and-erasures decoding and feedback repetitively until a decoding without erasures occurs. However the average transmission time is only a first order measure, for analyzing the benefits of errors-and-erasures decoding in which the strain of retransmissions are simply ignored if they are rare enough. Although these two results can be reinterpreted together, to reveal the error exponent of fixed-length block codes with errors-and-erasures decoding and feedback at all rates below capacity, they can be reinterpreted so only when erasure probability is decaying sub-exponentially with block length.

A separate stream of research focused on benefits errors-and-erasures decoding for block codes on DMC's without feedback. First Forney, [5], considered errors-and-erasures decoding without feedback and obtained an achievable trade-off between the exponents of error and erasure probabilities. Then Csiszár and Körner, [3] achieved same performance using universal coding and decoding algorithms. Later Telatar and Gallager, [12], introduced a strict improvement to the previously mentioned results. Recently there has been a revived interest in the problem, for universally achievable performances and alternative methods of analysis, [8], [7].

Our main aim in this work is to complement both streams of work by characterizing the error exponents for fixed length block codes with feedback at positive erasure exponents, by finding upper and lower bounds to it. We will first introduce our model and notation. Then we will derive a lower bound using a two phase coding algorithm similar to the one described by Yamamoto and Ito in [13]. However our decoding rule and analyzing techniques, inspired by Telatar's in [11] for the non-feedback case, will give us a better characterization of the trade-off between error and erasure exponents. After that we will derive an upper bound by advancing an idea introduced by Shannon, Gallager and Berlekamp in [10]. Resulting bound will be strictly convex for any fixed rate. Finally in the last part of the paper we will briefly discuss the connections to the erasure exponent of zero error codes with feedback.

II. Model and Notation:

Input and output alphabets of the forward channel are $\{1,\dots,|\mathcal{X}|\}$ and $\{1,\dots,|\mathcal{Y}|\}$, respectively. Corresponding symbols at time k are denoted by X_k and Y_k. The feedback channel is perfect, i.e, a symbol Z_k is chosen by the receiver from an arbitrarily large alphabet, $\{1,\dots,|\mathcal{Z}|\}$ after observing Y_k and is received by the transmitter without any error before the transmission of X_{k+1}.

The forward channel is a stationary and memoryless one characterized by an $|\mathcal{X}|$ by $|\mathcal{Y}|$ transition matrix $\{W_{jl}\}$.

$$\mathbf{P}\left[Y_k\middle|\, X^k, Y^{k-1}, Z^{k-1}\right] = \mathbf{P}\left[Y_k\middle|\, X_k\right] = W_{X_kY_k} \qquad (1)$$

The coding algorithm for a fixed length block code with feedback is a sequence of functions, $X_k(\cdot)$, which assigns an input symbol $X_k \in \mathcal{X}$ to each message, i in message set, $\mathcal{M} = \{1,2,\dots,|\mathcal{M}|\}$, at each time k, depending on the previous feedback symbols Z^{k-1}, i.e. $X_k = X_k(i, Z^{k-1})$.

The message, θ, is drawn from $\mathcal{M}$ with a uniform distribution and is given to the transmitter at time zero. At any time $k \in [1,\mathbf{n}]$ the input symbol $X_k(i, Z^{k-1})$ is sent. After receiving $Y^{\mathbf{n}}$ the receiver will decode a $\hat{\theta}(Y^{\mathbf{n}}) \in \{\mathbf{x}, 1,\dots,|\mathcal{M}|\}$ where $\mathbf{x}$ is the erasure symbol. Then error and erasure probabilities of a message $i \in \mathcal{M}$ are,

$$P_{\mathbf{x}}(i) \triangleq \mathbf{P}\left[\hat{\theta} = \mathbf{x}\middle|\, \theta = i\right] \quad P_{\mathbf{e}}(i) \triangleq \mathbf{P}\left[\hat{\theta} \neq \theta\middle|\, \theta = i\right] - P_{\mathbf{x}}(i).$$

[1] It is evident that including erasures will not result in an increase in the exponent for variable length block codes with feedback. However it was not clear in the light of converse techniques that has been previously employed on the problem [2],[1] whether such an increase is possible or not with list decoding. We have shown that it is not; a brief discussion will be given in section IV.

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

We will use a somewhat abstract but rigorous approach in defining the rate and achievable exponent pairs. A reliable sequence with erasures, $\mathcal{Q}$, will be a sequence of codes indexed by their block lengths such that

$$\limsup_{\mathbf{n}\to\infty} {P_{\mathbf{e}}}^{(\mathbf{n})} = 0 \qquad \limsup_{\mathbf{n}\to\infty} {P_{\mathbf{x}}}^{(\mathbf{n})} = 0.$$

Definition 1: The rate, erasure exponent, and error exponent of a reliable sequence $\mathcal{Q}$ are given by

$$R_{\mathcal{Q}} \triangleq \liminf_{\mathbf{n}\to\infty} \frac{\ln|\mathcal{M}^{(\mathbf{n})}|}{\mathbf{n}}\ , \ E_{\mathbf{x}\,\mathcal{Q}} \triangleq \liminf_{\mathbf{n}\to\infty} \frac{-\ln {P_{\mathbf{x}}}^{(\mathbf{n})}}{\mathbf{n}}$$

$$E_{\mathbf{e}\,\mathcal{Q}} \triangleq \liminf_{\mathbf{n}\to\infty} \frac{-\ln {P_{\mathbf{e}}}^{(\mathbf{n})}}{\mathbf{n}}.$$

Haroutunian, [6], has already established a strong converse for erasure free block codes with feedback which in our setting imlies that $\lim_{\mathbf{n}\to\infty}({P_{\mathbf{e}}}^{(\mathbf{n})} + {P_{\mathbf{x}}}^{(\mathbf{n})}) = 1$. Thus we will consider only rates below capacity, $\mathbf{C}$.

Definition 2: $\forall R \le \mathbf{C}$ and $\forall E_{\mathbf{x}} \ge 0$ the error exponent is,

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \triangleq \sup_{\mathcal{Q}:\ \substack{R(\mathcal{Q})\ge R \\ E_{\mathbf{x}}(\mathcal{Q})\ge E_{\mathbf{x}}}} \mathcal{E}_{\mathbf{e}}(\mathcal{Q}). \tag{2}$$

It is worth noting at this point that

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \mathcal{E}(R) \qquad \forall E_{\mathbf{x}} > \mathcal{E}(R) \tag{3}$$

where $\mathcal{E}(R)$ is the (true) error exponent of erasure free block-codes on DMCs with feedback.[2] The benefit of the errors-and-erasures decoding is the, possible, increase in the error exponent as the erasure exponents goes below $\mathcal{E}(R)$.

Determining $\mathcal{E}(R)$ for all R's and for all channels is still an open problem; only upper and lower bounds to $\mathcal{E}(R)$ are known. Our investigation will focus on quantifying the gains of errors-and-erasures decoding instead of finding $\mathcal{E}(R)$, consequently we will restrict ourselves the regions where erasure exponents are lower than the error exponents.

III. Achievability: A Lower Bound to $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$

Consider a two phase coding scheme, in the first phase of which transmitter uses a fixed composition code of length $\alpha\mathbf{n}$ and rate $\frac{R}{\alpha}$. At the end of the first phase, the receiver makes a maximum mutual information decoding[3] and get a temporary decision, $\tilde{\theta}$. The transmitter knows what $\tilde{\theta}$ is because of feedback and confirms $\tilde{\theta}$ if it is correct by sending accept codeword, rejects otherwise by sending reject codeword.[4] At the end of the second phase the receiver decodes using a partial order. If the pair $(y^{\mathbf{n}}, \tilde{\theta})$ satisfies $(y^{\mathbf{n}}, \tilde{\theta}) \succ (y^{\mathbf{n}}, j)$, for all $j \ne \tilde{\theta}$ in $\mathcal{M}$, i.e. if the pair $(y^{\mathbf{n}}, \tilde{\theta})$ dominates all other pairs, then $\tilde{\theta}$ becomes the final decision else an erasure is declared. Our goal in rest of this section is analyzing the performance of the coding architecture described above and finding a proper partial order[5] for it. Before starting this analysis let us recall certain basic features of fixed composition codes.

A. Fixed Composition Codes and Packing Lemma

Codes, whose all codewords have the same empirical distribution are called fixed composition codes. In such a code, for any message i and any output sequence $y^{\mathbf{n}}$ corresponding empirical distribution of transitions, from input letters to the output letters, $\mathbb{V}(y^{\mathbf{n}}, i)$, is called the conditional type. Furthermore the set of $y^{\mathbf{n}}$'s with a conditional type V, will be denoted by $T_V(x^{\mathbf{n}}(i))$ i.e.,

$$T_V(x^{\mathbf{n}}(i)) = \{y^{\mathbf{n}} : \mathbb{V}(y^{\mathbf{n}}, i) = V\} \tag{4}$$

The following packing lemma is proved by Csiszár and Körner, [3, Chapter 2, Lemma 5.1]

Lemma 1: For every $R > 0$, $\delta > 0$ and every type P of the sequences $\mathcal{X}^n$ satisfying $H(P) > R$, there exist at least $\lfloor e^{\mathbf{n}(R-\delta)} \rfloor$ distinct sequences $x^{\mathbf{n}}(i) \in \mathcal{X}^n$ of type P such that for every pair of stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$, $\hat{V} : \mathcal{X} \to \mathcal{Y}$ and for every i

$$|T_V(x^{\mathbf{n}}(i)) \bigcap \cup_{j\ne i} T_{\hat{V}}(x^{\mathbf{n}}(j))| \le |T_V(x^{\mathbf{n}}(i))| e^{-\mathbf{n}|I(P,\hat{V})-R|^+}$$

B. Coding Algorithm:

We use a length $\alpha\mathbf{n}$ code satisfying the property described in Lemma 1, of rate $\frac{R}{\alpha}$ and type P. At the end of the first phase the receiver makes a temporary decoding by choosing the codeword that has the maximum empirical mutual information with the output sequence $y^{\alpha\mathbf{n}}$.

$$\tilde{\theta}(y^{\alpha\mathbf{n}}) = \{i : \mathcal{I}\left(P, \mathbb{V}(y^{\alpha\mathbf{n}}, i)\right) > \mathcal{I}\left(P, \mathbb{V}(y^{\alpha\mathbf{n}}, j)\right) \quad \forall j \ne i\}$$

If $\tilde{\theta}(y^{\alpha\mathbf{n}}) = \theta$ the transmitter will send the accept codeword $\mathbf{x}_a$ else the transmitter will send the reject codeword $\mathbf{x}_r$ instead. Codewords $\mathbf{x}_a$ and $\mathbf{x}_r$ will have joint type $\varphi(i, j)$, i.e. the ratio of the number of time instances in which $\mathbf{x}_a$ has an i and $\mathbf{x}_r$ has a j to the length of the codewords, $\ell = (1-\alpha)\mathbf{n}$, will be $\varphi(i, j)$. The joint conditional type $\upsilon(l|i, j)$ of $y_{\alpha\mathbf{n}+1}^{\mathbf{n}}$, the output sequence in the second phase, is the ratio of number of time instances in which $y_k = l$, $x_a = i$ and $x_r = j$ to the overall number of time instances in which $x_a = i$ and $x_r = j$ i.e. $(1-\alpha)\varphi(i, j)\mathbf{n}$.

C. Decoding Rule:

The receiver decodes correctly, when $\tilde{\theta}(y^{\alpha\mathbf{n}}) = \theta$ and $(y^{\mathbf{n}}, \theta) \succ (y^{\mathbf{n}}, j)$ for all $j \ne \theta$.Thus an error or an erasure will occur only when the correct message can not dominate all other messages, i.e. when $\exists j \ne \theta$ such that $(y^{\mathbf{n}}, \theta) \not\succ (y^{\mathbf{n}}, j)$ consequently,

$$P_{\mathbf{e}}(i) + P_{\mathbf{x}}(i) = \mathbf{P}\left[\{y^{\mathbf{n}} : \exists j \ne i \text{ s.t.} (y^{\mathbf{n}}, i) \not\succ (y^{\mathbf{n}}, j)\} \middle| \theta = i\right] \tag{5}$$

[2] In order to see this consider a reliable sequence with erasures, $\mathcal{Q}$, and replace its decoding algorithm by any erasure free one such that, $\hat{\theta}'(y^{\mathbf{n}}) = \hat{\theta}(y^{\mathbf{n}})$ if $\hat{\theta}(y^{\mathbf{n}}) \ne \mathbf{x}$. Then ${P_{\mathbf{e}}}_{\mathcal{Q}'}^{(\mathbf{n})} \le {P_{\mathbf{x}}}_{\mathcal{Q}}^{(\mathbf{n})} + {P_{\mathbf{e}}}_{\mathcal{Q}}^{(\mathbf{n})}$; thus $E_{\mathbf{e}\mathcal{Q}'} = \min(E_{\mathbf{x}\mathcal{Q}}, E_{\mathbf{e}\mathcal{Q}})$ and $R_{\mathcal{Q}'} = R_{\mathcal{Q}}$. This together with the definition of $\mathcal{E}(R)$ leads to the relation given in equation (3).

[3] One might, possibly, improve this result by doing a list decoding at the end of the first phase, if the control phase and decoding algorithms are also modified accordingly. We have avoided that discussion in this first attempt.

[4] In general the codewords used in the second phase can depend on the observation in the first phase $y^{\alpha\mathbf{n}}$. Furthermore the coding in the second phase can actively use the feedback. These approaches have not been analyzed in this first attempt either.

[5] Binary relation $\succ$ is a strict partial order, on message, output sequence pairs, i.e. for all such pairs a, b and c following three holds,

$$(i)\{a \not\succ a\} \qquad (ii)\{a \succ b \Rightarrow b \not\succ a\} \qquad (iii)\{(a \succ b), (b \succ c) \Rightarrow (a \succ c)\}$$

Similarly an error will occur only when an incorrect message dominate all other messages, i.e. when $\exists j \neq \theta$ such that $(y^{\mathbf{n}}, j) \succ (y^{\mathbf{n}}, k)$ for all $k \neq j$.

$$P_{\mathbf{e}}(i)=\mathbf{P}\left[\{y^{\mathbf{n}} : \exists j \neq i \text{ s.t.}(y^{\mathbf{n}}, j) \succ (y^{\mathbf{n}}, k) \quad \forall k \neq j\} | \, \theta = i\right]$$

Note that $\{y^{\mathbf{n}} : \exists j \neq i \text{ s.t.}(y^{\mathbf{n}}, j) \succ (y^{\mathbf{n}}, k) \quad \forall k \neq j\}$ is a subset of $\{y^{\mathbf{n}} : \exists j \neq i \text{ s.t.}(y^{\mathbf{n}}, j) \succ (y^{\mathbf{n}}, i)\}$. Thus

$$P_{\mathbf{e}}(i) \leq \mathbf{P}\left[\{y^{\mathbf{n}} : \exists j \neq i \text{ s.t.}(y^{\mathbf{n}}, j) \succ (y^{\mathbf{n}}, i)\} | \, \theta = i\right] \quad (6)$$

The optimal order for $(y^{\mathbf{n}}, i)$'s for a given coding algorithm, might depend on the conditional types of $y^{\mathbf{n}}$ corresponding to all of the messages in the first phase, i.e. all $\mathbb{V}(i)$'s, and the conditional type of the second phase, υ. However we will restrict ourselves to the orders that can be written pair wise, i.e. which only depends on the conditional types of the two messages that are compared and the conditional type of the second phase.

We assume that $(y^{\mathbf{n}}, i) \succ (y^{\mathbf{n}}, j)$ implies $\mathcal{I}\left(P, \mathbb{V}(y^{\alpha \mathbf{n}}, i)\right) \geq \mathcal{I}\left(P, \mathbb{V}(y^{\alpha \mathbf{n}}, j)\right)$ and use packing lemma to bound the number of $y^{\mathbf{n}}$ in $T_V(x^{\mathbf{n}}(\theta))$ that are also in $T_{\hat{V}}(x^{\mathbf{n}}(i))$ of some $i \neq \theta$, then equation (5) becomes[6]

$$P_{\mathbf{e}}(i) + P_{\mathbf{x}}(i) \leq \sum_{\substack{(V,\upsilon) \not\succ (\hat{V},\upsilon) \\ (VP) \doteq (\hat{V}P)}} e^{-\overbrace{\left[\alpha \mathbf{D}\left(V \| W | P\right) + \left|\alpha \mathcal{I}\left(P, \hat{V}\right) - R\right|^{+} + (1-\alpha)\mathbf{D}\left(\upsilon \| W_a | \varphi\right)\right]}^{\lambda(P,R,\alpha,\varphi,V,\hat{V},\upsilon)} \mathbf{n}}$$

where $\mathbf{D}\left(\upsilon \| W_a | \varphi\right) = \sum_{i,j,l} \varphi(i,j) \upsilon(l|i,j) \log \frac{\upsilon(l|i,j)}{W_{i,l}}$ and $(VP) \doteq (\hat{V}P)$ means that PV and $P\hat{V}$ has same marginal output distributions, i.e. $\forall j$, $\sum_i P_i V_{i,j} = \sum_i P_i \hat{V}_{i,j}$.

Similarly we can write equation (6) as,

$$P_{\mathbf{e}}(i) \leq \sum_{\substack{(\hat{V},\upsilon) \succ (V,\upsilon) \\ (VP) \doteq (\hat{V}P)}} e^{-[\alpha \mathbf{D}(V \| W | P) + \left|\alpha \mathcal{I}\left(P, \hat{V}\right) - R\right|^{+} + (1-\alpha)\mathbf{D}(\upsilon \| W_r | \varphi)]\mathbf{n}}$$

$$\leq \sum_{\substack{(V,\upsilon) \succ (\hat{V},\upsilon) \\ (VP) \doteq (\hat{V}P)}} e^{-\overbrace{[\alpha \mathbf{D}\left(\hat{V} \| W | P\right) + |\alpha \mathcal{I}\left(P, V\right) - R|^{+} + (1-\alpha)\mathbf{D}\left(\upsilon \| W_r | \varphi\right)]}^{\beta(P,R,\alpha,\varphi,V,\hat{V},\upsilon)} \mathbf{n}}$$

where $\mathbf{D}\left(\upsilon \| W_r | \varphi\right) = \sum_{i,j,l} \varphi(i,j) \upsilon(l|i,j) \log \frac{\upsilon(l|i,j)}{W_{j,l}}$.

Note that the number of the terms in both of the sums are polynomial in $\mathbf{n}$, thus the term with the minimum exponent in each sum determine the decay rate of the sum. On the other hand two sums are over two disjoint and collectively exhaustive subsets of the set of all possible $\{(V,\upsilon), (\hat{V},\upsilon)\}$ pairs. In order to have an erasure exponent higher then say $E_{\mathbf{x}}$, all of the $\{(V,\upsilon), (\hat{V},\upsilon)\}$ pairs satisfying

$$\lambda(P, R, \alpha, \varphi, V, \hat{V}, \upsilon) \leq E_{\mathbf{x}} \quad (7)$$

should satisfy $(V,\upsilon) \succ (\hat{V},\upsilon)$. Adding more pairs to these, however can only decrease the error exponent. Thus we will chose $\succ$, so that $(V,\upsilon) \succ (\hat{V},\upsilon)$ if and only if equation (7) is satisfied. One can prove that for all values of $E_{\mathbf{x}} < \alpha E_r\left(\frac{R}{\alpha}, P\right)$, $(y^{\mathbf{n}}, i) \succ (y^{\mathbf{n}}, j)$ implies $\mathcal{I}\left(P, \mathbb{V}(y^{\alpha \mathbf{n}}, i)\right) \geq \mathcal{I}\left(P, \mathbb{V}(y^{\alpha \mathbf{n}}, j)\right)$ as we have previously assumed. Thus following is achievable

$$E_{\mathbf{e}}(R, P, E_{\mathbf{x}}, \alpha, \varphi) = \min_{\substack{(PV) \doteq (P\hat{V}) \\ \lambda(P,R,\alpha,\varphi,V,\hat{V},\upsilon) \leq E_{\mathbf{x}}}} \beta(P, R, \alpha, \varphi, V, \hat{V}, \upsilon) \quad (8)$$

D. Optimizing The Relative Durations of Phases:

Note that in our scheme $\mathbf{P}\left[\tilde{\theta}(y^{\alpha \mathbf{n}}) \neq \theta\right]$ decays no faster than $e^{-\alpha E_r\left(\frac{R}{\alpha}, P\right)\mathbf{n}}$. Thus either error or erasure exponent of the code has to be smaller than or equal to $\alpha E_r\left(\frac{R}{\alpha}, P\right)$. We are characterizing the error exponent using the region where erasure exponent is lower than the error exponent, thus we are interested in the region where $\alpha E_r\left(\frac{R}{\alpha}, P\right) \geq E_{\mathbf{x}}$. This implies,

$$\alpha \geq \alpha^*(R, P, E_{\mathbf{x}}) = \frac{R}{\xi_P^{-1}\left(\frac{E_{\mathbf{x}}}{R}\right)} \quad (9)$$

where $\xi_P(R) = \frac{E_r(R,P)}{R}$.

$$E_{\mathbf{e}}(R, P, E_{\mathbf{x}}) = \max_{\varphi} \max_{\alpha \in [\alpha^*, 1]} E_{\mathbf{e}}(R, P, E_{\mathbf{x}}, \alpha, \varphi) \quad (10)$$

Thus only α's in the interval $[\alpha^*(R, P, E_{\mathbf{x}}), 1]$ are permissible. Furthermore it can be proved that, for any quadruple $(R, P, E_{\mathbf{x}}, \varphi)$, $E_{\mathbf{e}}(R, P, E_{\mathbf{x}}, \alpha, \varphi)$ is a convex function of α on the same interval. Consequently its maximum is on the boundaries. For some φ's optimum α is 1, the trivial φ's which correspond to identical accept and reject codewords for example. For others, optimal α is α^* and resulting value of $E_{\mathbf{e}}$ is strictly greater than $E_{\mathbf{e}}(R, P, E_{\mathbf{x}}, 1, \varphi)$. Noting that when $\alpha = 1$, the value of $E_{\mathbf{e}}$ is same for all φ's we get

$$E_{\mathbf{e}}(R, P, E_{\mathbf{x}}) = \max_{\varphi} E_{\mathbf{e}}(R, P, E_{\mathbf{x}}, \alpha^*, \varphi) \quad (11)$$

For $E_{\mathbf{x}} = 0$ equation (11) leads to $\left(1 - \frac{R}{\mathbf{C}}\right)\mathbf{D}$ where $\mathbf{D} = \max_{i,j} \sum_l W_{i,l} \log \frac{W_{i,l}}{W_{j,l}}$, which can be proved to be optimal using the converse part of [2] or the upper bounds established in the next section.

For channels which has non-zero zero-error capacity, one can prove using equation (11) that, for any $E_{\mathbf{x}} < E_r(R)$ will be $E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty$, where $E_r(R)$ is the random coding exponent. Using upper bounds we establish in the next section one can prove that this is the optimum, i.e. $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ will be finite for all $E_{\mathbf{x}}$'a strictly greater than $E_r(R)$, at least for symmetric channels at rates above the critical rate.

IV. Converse: An Upper Bound to $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$

Shannon, Gallager and Berlekamp showed in, [10, Theorem 1], that for fixed length block codes, with list decoding and without feedback

$$\tilde{\mathcal{P}}_{\mathbf{e}}(M, \mathbf{n}, L) \geq \tilde{\mathcal{P}}_{\mathbf{e}}(M, \mathbf{n}_1, L_1) \tilde{\mathcal{P}}_{\mathbf{e}}(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L) \quad (12)$$

where $\tilde{\mathcal{P}}_{\mathbf{e}}(M, \mathbf{n}, L)$ denotes the minimum error probability of erasure free codes of length $\mathbf{n}$ with M messages and with decoding list size of L. As they have mentioned in [10] theorem continues to hold in the case when there exist a feedback link from receiver to the transmitter; although $\tilde{\mathcal{P}}_{\mathbf{e}}$'s

[6] We are not ignoring the fact that whenever $\tilde{\theta} \neq i$, the codeword $\mathbf{x}_r$ will be sent. But we are merely using the fact that $\sum_{\upsilon} e^{-(1-\alpha)\mathbf{D}(\upsilon \| W_r | \varphi)\mathbf{n}} \geq 1$.

would be different when feedback is available, the relation given in equation (12) will still hold.

Unlike them we are interested in the error probabilities of the codes with non-zero erasure probabilities. We will denote the minimum error probability of length $\mathbf{n}$ block codes, with M messages, decoding list size L and erasure probability $P_{\mathbf{x}}$ by $\mathcal{P}_{\mathbf{e}}(M, \mathbf{n}, L, P_{\mathbf{x}})$. Instead of the relation given in (12) we have the following.

Theorem 1: For any $\mathbf{n}$, M, L, $P_{\mathbf{x}}$ and for any $\mathbf{n}_1 \leq \mathbf{n}$, L_1, minimum error probabilities of fixed length block codes with feedback is satisfy,

$$\mathcal{P}_{\mathbf{e}}(M, \mathbf{n}, L, P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M, \mathbf{n}_1, L_1, 0) \cdot \mathcal{P}_{\mathbf{e}}\left(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L, \tfrac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M, \mathbf{n}_1, L_1, 0)}\right) \quad (13)$$

Similar to [10, Theorem 1], theorem 1 lower bounds error probability of a longer code, using that of shorter ones. After giving a sketch of the proof, we will derive upper bounds to error exponent using Theorem 1. Let us first consider the following lemma about, achievable error and erasure probabilities of codes, which has a nonuniform a priori probability distribution on the messages.

Lemma 2: For any block length $\mathbf{n}$, message set $\mathcal{M}$, list decoding size L and a priori probability distribution $f(\cdot)$ on $\mathcal{M}$ and for any integer K such that $\Omega(f, K) > 0$, all of the achievable error and erasure probability pairs satisfy,

$$(P_{\mathbf{x}}, P_{\mathbf{e}}) = \Omega(f, K)\,\vartheta \quad \text{for some } \vartheta \in \Psi(K+1, \mathbf{n}, L) \quad (14)$$

where $\Omega(f, K) = \min_{\mathcal{S}:|\mathcal{S}|=|\mathcal{M}|-K} f(\mathcal{S})$ and $\Psi(K+1, \mathbf{n}, L)$ is the set of achievable error, erasure probability pairs for length $\mathbf{n}$ block codes, with decoding list size of L and message set of size $K+1$ with uniform a priori probability distribution.

Proof: For any size $(K+1)$ subset $\mathcal{M}'$ of $\mathcal{M}$, consider the original coding and decoding rule. If a message which is not in $\mathcal{M}'$ was in the decoding list we will ignore it, if the decoding list is solely composed of such messages we will declare an erasure. Clearly this is a length $\mathbf{n}$ code with $(K+1)$ messages and list decoding size of L. Thus

$$\frac{1}{K+1} \sum_{i \in \mathcal{M}'} (P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) \in \Psi(K+1, \mathbf{n}, L) \quad (15)$$

Let the smallest non-zero element of $\{f(1), f(2), \ldots f(|\mathcal{M}|)\}$ be $f(\xi_1)$, pick any size $(K+1)$ subset of $\mathcal{M}$ which includes ξ_1 and has all non-zero elements, say $\mathcal{M}_1$, then we have,

$$\begin{aligned}(P_{\mathbf{x}}, P_{\mathbf{e}}) &= \sum_{i \in \mathcal{M}} f(i)\,(P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) \\ &= \sum_{i \in \mathcal{M} \setminus \mathcal{M}_1} f(i)(P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) + f(\xi_1) \sum_{i \in \mathcal{M}_1} (P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) \\ &\quad + \sum_{i \in \mathcal{M}_1} (f(i) - f(\xi_1))(P_{\mathbf{x}}(i), P_{\mathbf{e}}(i))\end{aligned}$$

Using equation (15) we get,

$$(P_{\mathbf{x}}, P_{\mathbf{e}}) = \sum_{i \in \mathcal{M}} f^{(1)}(i)(P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) + f(\vartheta_1)\vartheta_1 \quad (16)$$

for some $\vartheta_1 \in \Psi(K+1, \mathbf{n}, L)$, where $f(\vartheta_1) = (K+1)f(\xi_1)$ and $f^{(1)}(i) = f(i) - f(\xi_1)\mathbb{I}\{i \in \mathcal{M}_1\}$. Consequently

$$\sum_{i \in \mathcal{M}} f^{(1)}(i) + f(\vartheta_1) = 1 \quad (17)$$

Furthermore the number of non-zero $f^{(1)}(i)$'s is at least one less than that of non-zero $f(i)$'s. The remaining probabilities, $f^{(1)}(i)$, have a minimum, $f^{(1)}(\xi_2)$ among its non-zero elements. We can repeat the same argument once more using that element and reduce the number of non-zero elements at least one more. After at most $|\mathcal{M}| - K$ iterations like this we will get,

$$(P_{\mathbf{x}}, P_{\mathbf{e}}) = \sum_{j=1}^{|\mathcal{M}|-K} f(\vartheta_j)\vartheta_j + \sum_{i \in \mathcal{M}} f^{(|\mathcal{M}|-K)}(i)(P_{\mathbf{x}}(i), P_{\mathbf{e}}(i)) \quad (18)$$

where at most K of $f^{(|\mathcal{M}|-K)}(i)$'s are non-zero and all of them are non-negative and less than or equal to corresponding $f(i)$.

Consequently the first sum is equal to a weighted sum of ϑ_j's multiplied by a scalar greater than $\Omega(f, K)$. Furthermore the second sum is equal to a pair with non-negative entries. Recalling that $\Psi(K+1, \mathbf{n}, L)$ is convex, and the fact that

$$\forall a \geq 1,\ \forall b_1, b_2 \geq 0 \quad \vartheta \in \Psi \Rightarrow (a \cdot \vartheta + (b_1, b_2)) \in \Psi \quad (19)$$

we get equation (14). ∎

For proving theorem 1, we will first write the error and erasure probabilities, as weighted sum of error and erasure probabilities of length $\mathbf{n} - \mathbf{n}_1$ codes with a priori probability distribution $f_{y^{\mathbf{n}_1}}(\cdot) = \mathbf{P}[\theta = \cdot\,|\, y^{\mathbf{n}_1}]$, over $y^{\mathbf{n}_1}$'s. Then we will apply similar convexity arguments and use the fact that,

$$\sum_{y^{\mathbf{n}_1}} \mathbf{P}[y^{\mathbf{n}_1}]\,\Omega(f_{y^{\mathbf{n}_1}}, L_1) \geq \mathcal{P}_{\mathbf{e}}(M, \mathbf{n}_1, L_1, 0) \quad (20)$$

to prove that $(P_{\mathbf{x}}, P_{\mathbf{e}}) = \mathcal{P}_{\mathbf{e}}(M, \mathbf{n}_1, L_1, 0)\vartheta$ for some ϑ in $\Psi(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L)$. Using the fact that $\mathbf{P}_{\mathbf{e}}(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L, \mathbf{P}_{\mathbf{x}})$ is the boundary of the $\Psi(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L)$ and $\mathbf{P}_{\mathbf{e}}(L_1 + 1, \mathbf{n} - \mathbf{n}_1, L, \mathbf{P}_{\mathbf{x}})$ is a decreasing function of $P_{\mathbf{x}}$, we get the desired result, equation (13).

Like the result of Shannon, Gallager and Berlekamp in [10, Theorem 1], theorem 1 is correct both with and without feedback. Although $\mathcal{P}_{\mathbf{e}}$'s will be different in each case, the relationship between them given in equation (13) holds in both cases.

As a digression we have employed these ideas in variable length block codes. The main difference in the analysis was, using decoding time $\mathbf{E}[\tau]$ instead of erasure probability $P_{\mathbf{x}}$. By incorporating the strong converse of Haroutunian for fixed length block codes with feedback to our analysis we have proved the tightness of Burnashev exponent even when list decoding is allowed. This fact was not known, prior to this work. Details of the calculation are omitted because of the limited space here.

A. Generalized Straight Line Bound for Error-Erasure Exponents

One can convert Theorem 1 into an upper bound on error exponent of errors-and-erasures codes in terms of error exponents of erasure-free codes and error-and-erasures codes, by taking the logarithm of equation (13) and dividing by $\mathbf{n}$. Resulting bounds will constitute a family of straight lines.

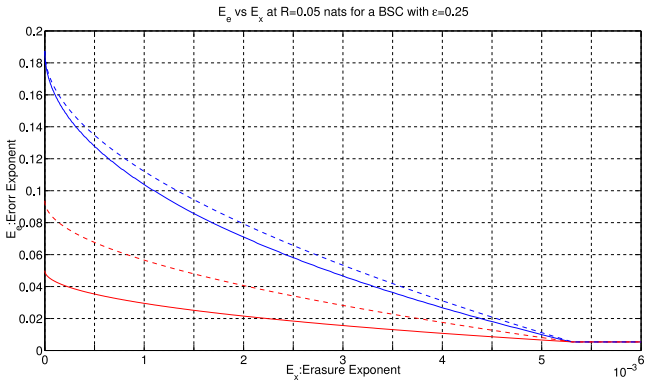

Fig. 1. Solid lines are lower bounds on the error exponent for block codes on DMC with feedback, which have established in this work, and without feedback, which was previously established [5], [3], [11]. Dashed lines are the upper bounds obtained using Theorem 2 for $\gamma = \gamma^*(R, E_{\mathbf{x}})$. While calculating the upper bound for the case without feedback we have used Telatar's result [11] on the error exponent at zero rate and zero erasure exponent.

One end of each line will be $(R, E_{\mathbf{x}}, \tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}))$ where $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ is an upper bound on the error exponent of errors-and-erasures codes. The other end will be a point of the forms $(R, \tilde{E}(R), \tilde{E}(R))$ where $\tilde{E}(R)$ is an upper bound on error exponent of erasure-free codes with feedback. In the following theorem we have used Haroutunian's upper bound, $E_h(R)$, on error exponents of erasure-free codes with feedback.

Theorem 2: For any rate $R \geq 0$, $E_{\mathbf{x}} \leq E_h(R)$, for any $\gamma \in \left(\frac{R}{\mathbf{C}}, \gamma^*(R, E_{\mathbf{x}})\right)$

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq \gamma E_h\left(\frac{R}{\gamma}\right) + (1-\gamma)\mathcal{E}_{\mathbf{e}}\left(0, \frac{E_{\mathbf{x}} - \gamma E_h\left(\frac{R}{\gamma}\right)}{1-\gamma}\right)$$

where $\gamma^*(R, E_{\mathbf{x}})$ is the unique solution of $\gamma E_h\left(\frac{R}{\gamma}\right) = E_{\mathbf{x}}$ if it exists, 1 else.

For the ease of discussion we have restricted our discussion to the case where list size is one, but one can easily derive corresponding results for larger but fixed list sizes and for list sizes increasing with block length.

In Fig. 1. upper and lower bounds to the error exponent are plotted as a function of erasure exponent at a fixed rate, $R = 0.05$ nats per channel use, for a binary symmetric channel with a cross over probability of $\epsilon = 0.25$ both with and without feedback.

V. Zero-Error Codes with Feedback

Zero error codes with erasure and feedback is part of the general error and erasure exponents discussion. In order to state the links between these problems clearly let us define 'zero error reliable sequences with erasure', $\mathcal{Q}_0$'s, as sequences of codes such that

$$\forall \mathbf{n} \quad P_{\mathbf{e}}{}^{(\mathbf{n})} = 0 \qquad \limsup_{\mathbf{n}\to\infty} P_{\mathbf{x}}{}^{(\mathbf{n})} = 0$$

The highest rate achievable with zero-error codes with erasures is the zero-error capacity with feedback and erasures and it will be denoted by $\mathbf{C}_{x,0}$. One can show that $\mathbf{C}_{x,0}$ is zero, if all of the transition probabilities are positive. Furthermore, if there is one or more zero probability transitions, $\mathbf{C}_{x,0}$ is equal to channel capacity $\mathbf{C}$.

Definition 3: $\forall R \leq \mathbf{C}_{x,0}$ the zero error erasure exponent with feedback is defined as

$$\mathcal{E}_{\mathbf{x},0}(R) \triangleq \sup_{\mathcal{Q}_0 : R(\mathcal{Q}_0) \geq R} E_{\mathbf{x}}(\mathcal{Q}_0)$$

Clearly any zero-error reliable sequence with erasures, $\mathcal{Q}_0$ is also a reliable sequence with erasures. Thus

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty \qquad \forall E_{\mathbf{x}} \leq \mathcal{E}_{\mathbf{x},0}(R) \tag{21}$$

Indeed, it can be shown that this reasoning works both ways i.e., $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ is finite for values of $E_{\mathbf{x}}$ greater than $\mathcal{E}_{\mathbf{x},0}(R)$.

VI. Acknowledgment

Authors are thankful to Emre Telatar for his encouragement on the problem, for various discussion on zero error case and for his counter example on the erasure exponent of zero-error case.

References

[1] P. Berlin, B. Nakiboğlu, B. Rimoldi, and İ. E. Telatar. A simple derivation of burnahsevs converse. arXiv:cs/0610145v2 [cs.IT].

[2] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problemy Perdachi Informatsii*, 12, No. 4:10–30, 1976.

[3] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982.

[4] 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.

[5] Jr. G. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *IEEE Transactions on Information Theory*, Vol.14, Iss.2:206–220, 1968.

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

[7] N. Merhav. Error exponents of erasure/list decoding revisited via moments of distance enumerators. arXiv:0711.2501v1 [cs.IT].

[8] P. Moulin. A neyman-pearson approach to universal erasure and list decoding. arXiv:0801.4544v1 [cs.IT].

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

[10] 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.

[11] İ. E. Telatar. *Multi-Access Communications with Decision Feedback Decoding*. Ph.D. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 1992.

[12] İ. E. Telatar and R. G. Gallager. New exponential upper bounds to error and erasure probabilities. In *ISIT 1994, Trondheim, Norway June 27-July 1, 1994*, 1994.

[13] H. Yamamoto and K. Itoh. Asymptotic performance of a modified schalkwijk-barron scheme for channels with noiseless feedback. *IEEE Transactions on Information Theory*, Vol.25, Iss.6:729–733, 1979.