Upper Bounds to Error Probability with Feedback

Barış Nakiboğlu Lizhong Zheng
Research Laboratory of Electronics at MIT
Cambridge, MA, 02139
Email: {nakib, lizhong }@mit.edu

***Abstract*—A new technique is proposed for upper bounding the error probability of fixed length block codes with feedback. Error analysis is inspired by Gallager's error analysis for block codes without feedback. Zigangirov-D'yachkov encoding scheme is analyzed with the technique on binary input channels and K-ary symmetric channels.**

I. Introduction

Shannon showed, [6], that capacity of the discrete memoryless channels (DMCs) does not increase with feedback. Later it was shown that the exponential decay rate of the error probability of fixed length bloc codes can not exceed sphere packing exponent even when there is feedback, first by Dobrushin, [3], for symmetric channels and later by Sheverdyaev, [7], for general DMCs. In other words for the rates above the critical rate even the error exponent does not increase with feedback, when we restrict ourselves to the fixed length block codes. Characterizing the improvement in the error exponent for the rates below the critical rate is the most pressing open question in this stream of research.[1]

The first work on the error analysis block codes with feedback was by Berlekamp [1]. He has obtained a closed form expression of the error exponent at zero rate for binary symmetric channels (BSCs). Later Zigangirov [8] proposed an encoding scheme, for BSCs which reaches sphere packing exponent for all rate over a critical rate R_{Zcrit}.[2] Furthermore at zero-rate Zigangirov's encoding scheme reaches optimal error exponent at zero rate, which is derived by Berlekamp in [1]. Later D'yachkov, [4], proposed a generalization of the encoding scheme of Zigangirov, and obtained a coding theorem for general DMCs. However the optimization problem in his coding theorem, is quite involved and does not allow for simplifications that will lead us to conclusions about the error exponents of general DMCs. In [4] after pointing out this fact, D'yachkov focuses on binary input channels and K-ary symmetric channels and derives the error exponent expressions for these channels.

Our approach will be similar to the D'yachkov's in the sense that we will first prove a coding theorem for general DMCs and then focus on particular cases and demonstrate its gains over non-feedback encoding schemes. We will start with introducing the channel model we have at hand and the notation we use. After that we will consider the encoding schemes that uses feedback and make an error analysis which is very similar to the that of Gallager in [5]. Then we will use our results in different cases, to recover the results of Zigangirov [8] and D'yachkov [4] for binary input channels and to improve D'yachkov's results, [4], for K-ary symmetric channels.[3]

II. Channel Model and Notation

We have a discrete memoryless channel with input alphabet $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$, output alphabet $\mathcal{Y} = \{1, 2, \ldots, |\mathcal{Y}|\}$. Channel transition probabilities are given by a $|\mathcal{X}|$-by-$|\mathcal{Y}|$ matrix $W(y|x)$. In addition we assume that a noiseless, delay free feedback link exists from transmitter to the receiver. Thus transmitter learns the channel output at time k before the transmission of the symbol at time $k+1$.

A length $\mathbf{n}$ block code with feedback for a messages set $\mathcal{M} = \{1, 2, \ldots, \lceil e^{\mathbf{n}R} \rceil\}$ is a feedback encoding scheme together with a decoding rule. The feedback encoding scheme, Ψ, is a mapping from the set of possible output sequences, $y^{j-1} \in \mathcal{Y}^{j-1}$ for $j \in \{1, 2, \ldots, \mathbf{n}\}$, to the set of possible input symbol assignment to the messages in the set $\mathcal{M}$

$$\Psi : \bigcup_{j=1}^{\mathbf{n}} \mathcal{Y}^{j-1} \rightarrow \mathcal{X}^{|\mathcal{M}|} \tag{1}$$

The input letter for the message $m \in \mathcal{M}$ at time j given $y^{j-1} \in \mathcal{Y}^{j-1}$ is the m^{th} element of $\Psi(y^{j-1})$, i.e., $\Psi_m(y^{j-1})$. Note that when there is no feedback $\Psi(y^{j-1}) = \Psi_j$, $\forall y^{j-1} \in \mathcal{Y}^{j-1}$. The probability of observing a $y^i \in \mathcal{Y}^i$ conditioned on message $m \in \mathcal{M}$ is,

$$\mathbf{P}\left\{y^i \middle| \theta = m\right\} = \prod_{j=1}^{i} W(y_j|\Psi_m(y^{j-1}))$$

A decoding rule is simply a mapping from the set of all possible length $\mathbf{n}$ output sequences $\mathcal{Y}^{\mathbf{n}}$ to the message set

[1] There are a number of closely related models in which error exponent analysis has been successfully applied, like variable-length block codes, fixed length block codes with errors-and-erasure decoding, block codes on additive white Gaussian noise channels, fixed/variable delay code on DMCs. We are refraining from discussing these variants mainly because understanding those variants will not help the reader much in understanding the work at hand.

[2] Evidently $R_{Zcrit} < R_{crit}$ where R_{crit} is the critical rate in the non-feedback case, i.e. the rate above which random coding exponent is equal to the sphere packing exponent.

[3] Indeed same improvement can be obtained within framework of the analysis introduced by Zigangirov and D'yachkov with some fairly minor modifications.

$\mathcal{M}$.

$$\Phi : \mathcal{Y}^{\mathbf{n}} \to \mathcal{M}$$

We use a maximum a-posteriori probability (MAP) decoder. When there is a tie the message with the smaller index is decoded.

III. Error Analysis

The error probability of the message $m \in \mathcal{M}$ is,

$$P_{\mathbf{e},m} = \sum_{y^{\mathbf{n}} \in \mathcal{Y}^{\mathbf{n}}} \mathbf{P}\left\{y^{\mathbf{n}} \middle| \theta = m\right\} \mathbb{I}\{\Phi(y^{\mathbf{n}}) \neq m\}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. Note that for a MAP decoder, for any $\rho > 0$, and $\eta > 0$ we have

$$\mathbb{I}\{\Phi(y^{\mathbf{n}}) \neq m\} \leq \left(\frac{\sum_{k \neq m} \mathbf{P}\left\{y^{\mathbf{n}} \middle| \theta = k\right\}^{\eta}}{\mathbf{P}\left\{y^{\mathbf{n}} \middle| \theta = k\right\}^{\eta}} \right)^{\rho}$$

Consequently

$$P_{\mathbf{e},m} \leq \sum_{y^{\mathbf{n}} \in \mathcal{Y}^{\mathbf{n}}} \mathbf{P}\left\{y^{\mathbf{n}} \middle| \theta = m\right\}^{1-\rho\eta} \left(\sum_{k \neq m} \mathbf{P}\left\{y^{\mathbf{n}} \middle| \theta = k\right\}^{\eta} \right)^{\rho}$$

If we introduce the short hand

$$\Gamma_{\rho,\eta,j}(y^j) = \sum_{m \in \mathcal{M}} \mathbf{P}\left\{y^j \middle| \theta = m\right\}^{1-\rho\eta} \left[\sum_{k \neq m} \mathbf{P}\left\{y^j \middle| \theta = k\right\}^{\eta} \right]^{\rho}$$

and recall that $P_{\mathbf{e}} = \frac{\sum_m P_{\mathbf{e},m}}{|\mathcal{M}|}$, we get,

$$P_{\mathbf{e}} \leq \sum_{y^{\mathbf{n}} \in \mathcal{Y}^{\mathbf{n}}} \frac{\Gamma_{\rho,\eta,\mathbf{n}}(y^{\mathbf{n}})}{|\mathcal{M}|} \tag{2}$$

If $\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1}) \neq 0$ we can divide and multiply by $\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})$, i.e,

$$\begin{aligned} P_{\mathbf{e}} &\leq \sum_{y^{\mathbf{n}} \in \mathcal{Y}^{\mathbf{n}}} \frac{\Gamma_{\rho,\eta,\mathbf{n}}(y^{\mathbf{n}})}{|\mathcal{M}|} \\ &= \sum_{y^{\mathbf{n}} \in \mathcal{Y}^{\mathbf{n}}} \frac{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})}{|\mathcal{M}|} \frac{\Gamma_{\rho,\eta,\mathbf{n}}(y^{\mathbf{n}})}{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})} \\ &= \sum_{y^{\mathbf{n}-1} \in \mathcal{Y}^{\mathbf{n}-1}} \frac{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})}{|\mathcal{M}|} \sum_{y_{\mathbf{n}} \in \mathcal{Y}} \frac{\Gamma_{\rho,\eta,\mathbf{n}}(y^{\mathbf{n}})}{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})} \end{aligned}$$

Thus

$$P_{\mathbf{e}} \leq \sum_{y^{\mathbf{n}-1} \in \mathcal{Y}^{\mathbf{n}-1}} \frac{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})}{|\mathcal{M}|} \xi(\rho, \eta, y^{j-1}, \Psi) \tag{3}$$

where $\xi(\rho, \eta, y^{\mathbf{n}-1}, \Psi)$ is given by,

$$\xi = \begin{cases} 0 & \text{if } \Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1}) = 0 \\ \sum_{y_{\mathbf{n}} \in \mathcal{Y}} \frac{\Gamma_{\rho,\eta,\mathbf{n}}(y^{\mathbf{n}})}{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})} & \text{if } \Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1}) \neq 0 \end{cases}$$

Let us define $\alpha(\rho, \eta, \Psi)$ as

$$\alpha(\rho, \eta, \Psi) = \max_{j \in \{1, \dots \mathbf{n}\}} \max_{y^{j-1} \in \mathcal{Y}^{j-1}} \xi(\rho, \eta, y^{j-1}, \Psi) \tag{4}$$

Then

$$\begin{aligned} P_{\mathbf{e}} &\leq \sum_{y^{\mathbf{n}-1} \in \mathcal{Y}^{\mathbf{n}-1}} \frac{\Gamma_{\rho,\eta,\mathbf{n}-1}(y^{\mathbf{n}-1})}{|\mathcal{M}|} \alpha(\rho, \eta, \Psi) \\ &\leq \frac{\Gamma_{\rho,\eta,0}(y^0)}{|\mathcal{M}|} \alpha(\rho, \eta, \Psi)^{\mathbf{n}} \\ &\leq |\mathcal{M} - 1|^{\rho} \alpha(\rho, \eta, \Psi)^{\mathbf{n}} \\ &\leq e^{\mathbf{n}(\rho R + \ln \alpha(\rho, \eta, \Psi))} \end{aligned} \tag{5}$$

Now what we find an encoding scheme, Ψ, with small $\alpha(\rho, \eta, \Psi)$. For that we will focus on the encoding schemes that are repetitions of the same 'one step encoding function'.

IV. One Step Encoding Functions

Let $\mathcal{Q}^{|\mathcal{M}|}$ be the set $|\mathcal{M}|$-dimensional vectors whose all entries are non-negative. Let Υ be a parametric function,

$$\Upsilon : \mathcal{Q}^{|\mathcal{M}|} \times \mathcal{X}^{|\mathcal{M}|} \to R$$

If q has at least two non-zero entries then Υ is defined as

$$\Upsilon_{\rho,\eta}(q, \chi) = \sum_{y,m} \frac{W(y|\chi_m)^{(1-\rho\eta)} q_m^{\frac{1-\rho\eta}{\eta}} \left(\sum_{k \neq m} W(y|\chi_k)^{\eta} q_k\right)^{\rho}}{\sum_m q_m^{\frac{1-\rho\eta}{\eta}} \left(\sum_{k \neq m} q_k\right)^{\rho}} \tag{6}$$

else $\Upsilon_{\rho,\eta}(q, \chi) = 0$.

Let us introduce the short hand,

$$\varphi(m|y^j) = \mathbf{P}\left\{y^j \middle| \theta = m\right\}^{\eta}$$

then we have

$$\varphi(m|y^j) = W(y_j|\Psi_m(y^{j-1}))^{\eta} \varphi(m|y^{j-1})$$

We can write ξ in terms of Υ as follows,

$$\xi(\rho, \eta, y^{j-1}, \Psi) = \Upsilon_{\rho,\eta}(\varphi(\cdot|y^{j-1}), \Psi(y^{j-1})) \tag{7}$$

Note that $\xi(\rho, \eta, y^{j-1}, \Psi)$ depends on the encoding in first $j-1$ time units only through the $|\mathcal{M}|$ dimensional vector $\varphi(\cdot|y^{j-1})$. Thus if we can find a χ_q for each $q \in \mathcal{Q}^{|\mathcal{M}|}$ such that $\Upsilon_{\rho,\eta}(q, \chi_q)$ is small, we can use it to obtain a block code with small error probability. These mappings are what we call one step encoding function.

An $|\mathcal{M}|$ dimensional one step encoding function, $\mathbb{X}$, is a mapping from $\mathcal{Q}^{|\mathcal{M}|}$ to the $\mathcal{X}^{|\mathcal{M}|}$, i.e.,

$$\mathbb{X} : \mathcal{Q}^{|\mathcal{M}|} \to \mathcal{X}^{|\mathcal{M}|}$$

We define $\alpha(\rho, \eta, \mathbb{X})$ as

$$\alpha(\rho, \eta, \mathbb{X}) = \max_{q \in \mathcal{Q}^{|\mathcal{M}|}} \Upsilon_{\rho,\eta}(q, \mathbb{X}(q)) \tag{8}$$

Lemma 1: For any $\rho > 0$, $\eta > 0$, $|\mathcal{M}|$ dimensional one step encoding function $\mathbb{X}$, and for any $\mathbf{n} \geq 1$ there exists and block code such that

$$P_{\mathbf{e}} \leq e^{\mathbf{n}(\ln \alpha(\rho, \eta, \mathbb{X}) + \rho R)} \tag{9}$$

where $R = \frac{|\mathcal{M}|}{\mathbf{n}}$

Proof: Consider the encoding scheme, Ψ such that,

$$\Psi_m(y^{j-1}) = \mathbb{X}_m(\varphi(\cdot|y^{j-1})) \tag{10}$$

As a result of equations (4), (7) and (8) we get,

$$\alpha(\rho,\eta,\Psi) \le \alpha(\rho,\eta,\mathbb{X}) \tag{11}$$

Using equation (5) we get the claim of the lemma. ■

When we are calculating the achievable error exponents the role of the minimum of $-\ln\alpha(\rho,\eta,\mathbb{X})$ will be very similar to that of $E_0(\rho)$ in the case without feedback, [5].

V. Examples:

A. Achievability of Random Coding Exponent:

In this subsection we will, as a sanity check, re-derive the achievability of random coding exponent for all DMC using Lemma 1. Let $\eta = \frac{1}{1+\rho}$. For any $|\mathcal{M}| > 1$, at each $q \in \mathcal{Q}$ consider the set of all possible mappings of messages to the input letters, $\mathcal{X}^{|\mathcal{M}|}$, and calculate the expected value of $\Upsilon_{\rho,\eta}(q,\chi)$, where the probability of each $\chi \in \mathcal{X}^{|\mathcal{M}|}$ is simply given by $\prod_{x\in\mathcal{X}} P(x)^{r(\chi,x)}$ where $r(\chi,x)$ is the number of messages assigned to input letter $x \in \mathcal{X}$. Then one can show that

$$\mathbf{E}\left[\Upsilon_{\rho,\frac{1}{1+\rho}}(q,\chi)\right] = e^{-E_0(\rho,P)} \quad \forall q \in \mathcal{Q}^{|\mathcal{M}|} \tag{12}$$

where $E_0(\rho,P) = \sum_y(\sum_x W(y|x)^{\frac{1}{1+\rho}}P(x))^{1+\rho}$ Thus for each $q \in \mathcal{Q}^{|\mathcal{M}|}$ there exist at least one $\chi_q \in \mathcal{X}^{|\mathcal{M}|}$ such that

$$\Upsilon_{\rho,\frac{1}{1+\rho}}(q,\chi_q) \le e^{-E_0(\rho,P)} \tag{13}$$

Thus for the one step encoding function $\mathbb{X}(q)$, such that $\mathbb{X}(q) = \chi_q$ we have

$$-\ln\alpha(\rho,\tfrac{1}{1+\rho},\mathbb{X}) \ge E_0(\rho,P)$$

Using this together with the lemma 1 we can conclude that:

Corollary 1: For any input distribution $P(\cdot)$ on $\mathcal{X}$, and $\rho \in (0,1]$, $R \ge 0$ any $\mathbf{n} > 1$ there exists a length $\mathbf{n}$ block code of the form given in equation (10) such that.

$$P_{\mathbf{e}} \le e^{\mathbf{n}(-E_0(\rho,P)+\rho R)} \tag{14}$$

Note that above description is not constructive in the sense that it proves the existence of a one-step-encoding scheme, $\mathbb{X}$ with the desired properties but it does not tell us how to find it. Encoding scheme we will investigate below however does specify a $\mathbb{X}$ with the desired properties.

B. Zigangirov-D'yachkov Encoding Scheme:

In this subsection we will describe the Zigangirov-D'yachkov encoding scheme and apply Lemma 1 to this encoding scheme on binary input channels and k-ary symmetric channels. This encoding scheme was first described by Zigangirov, [8], for binary symmetric channels than generalized by D'yachkov to general DMCs. Consider a probability distribution $P(\cdot)$ on input alphabet $\mathcal{X}$ and a $q \in \mathcal{Q}^{|\mathcal{M}|}$. Without loss of generality we can assume that[4] $\forall i,j \in \mathcal{M}$, if $i \le j$ then $q_i \ge q_j$. Now we can define mapping χ for a given q and P iteratively as follows:

$$\gamma_0(x) = 0$$
$$\chi_j = \arg\min_{x\in\mathrm{supp}(P)} \frac{\gamma_{j-1}(x)}{P(x)}$$
$$\gamma_j(x) = \sum_{1\le i\le j:\chi_i = x} q_i$$

For assigning $j \in \mathcal{M}$ we first calculate for each input letter, $x \in \mathcal{X}$, the total mass of all of the messages that has already been assigned to x, $\gamma_{j-1}(x)$. Then we divide $\gamma_{j-1}(x)$'s by the corresponding $P(x)$ values and assign the message $j \in \mathcal{M}$ to the $x \in \mathcal{X}$, for which $P(x) > 0$ and $\frac{\gamma_{j-1}(x)}{P(x)}$ is the minimum. If there is a tie we choose the input letter, x, with larger $P(x)$. If there is still a tie, we choose the input letter with smaller index.

1) Properties of Z-D Encoding Scheme: Now let us point out one property of this encoding scheme which will become handy later on. A Zigangirov-D'yachkov encoding scheme with $P(\cdot)$, will satisfy,

$$\zeta_m = \tfrac{q_{\chi_m} - q_m}{P(\chi_m)} \le \tfrac{q_x}{P(x)} \quad \forall x \in \mathcal{X} \quad \forall m \in \mathcal{M} \tag{15}$$

where $q_x = \gamma_{|\mathcal{M}|}(x,P,\chi^{|\mathcal{M}|})$. In order to see this, simply consider the last message assigned to each input letter $x \in \mathcal{X}$. They will satisfy this property by construction. Since the messages that are assigned to the same letter prior to the last message should have at least the same mass as the last one, they will satisfy the property given in (15) too. Thus for any $q \in \mathcal{Q}^{|\mathcal{M}|}$ and any input distribution $P(x)$, the mapping created by a Zigangirov-D'yachkov encoding scheme, satisfies

$$q_x - P(x)\zeta_m \ge 0 \quad \forall x \in \mathcal{X} \quad \forall m \in \mathcal{M} \tag{16}$$

In other words, with Zigangirov-D'yachkov encoding scheme, the mass of the q is distributed over the input letters in such a way that; when we consider all the mass distribution except an $m \in \mathcal{M}$, it is a linear combination of $P(x)$ and $\delta_{x,x'}$'s for $x' \ne \chi_m$. Thus

$$\tfrac{\sum_{k\ne m}\mathbb{I}\{\chi_k = x\}q_k}{\sum_{k\ne m} q_k} = \tfrac{\zeta_m P(x)}{\sum_{k\ne m} q_k} + \tfrac{\sum_{x\ne\chi_m}\mathbb{I}\{\chi_k = x\}(q_x - \zeta_m P(x))}{\sum_{k\ne m} q_k}$$

For $\rho \ge 1$, z^ρ is a convex function of z thus $\mathbf{E}[z]^\rho \le \mathbf{E}[z^\rho]$, thus

$$\left(\tfrac{\sum_{k\ne m} W(y|\chi_k)^\eta q_k}{\sum_{k\ne m} q_k}\right)^\rho \le \tfrac{\zeta_m}{\sum_{k\ne m} q_k}\left[\sum_x W(y|x)^\eta P(x)\right]^\rho + \sum_{x\ne\chi_m}\tfrac{(q_x - \zeta_m P(x))}{\sum_{k\ne m} q_k} W(y|x)^{\eta\rho} \tag{17}$$

Let us define f_m as

$$f_m = \sum_y W(y|\chi_m)^{(1-\rho\eta)}\left(\tfrac{\sum_{k\ne m} W(y|\chi_k)^\eta q_k}{\sum_{k\ne m} q_k}\right)^\rho$$

[4] If this is not the case for a q, we can rearrange the messages $m \in \mathcal{M}$, according to their q_m in decreasing order. If two or more messages have same mass, q, we order them with respect to their index numbers.

Using equation (17) we get,

$$\begin{aligned} f_m &\leq \frac{\zeta_m}{\sum_{k\neq m} q_k} \sum_y W(y|\chi_m)^{(1-\rho\eta)} \left[\sum_x W(y|x)^\eta P(x)\right]^\rho \\ &+ \sum_{x\neq\chi_m} \frac{(q_x - P(x)\zeta_m)}{\sum_{k\neq m} q_k} \sum_y W(y|\chi_m)^{(1-\rho\eta)} W(y|x)^{\eta\rho} \\ &\leq \frac{\zeta_m}{\sum_{k\neq m} q_k} e^{\beta_{\chi_m}(P,\rho,\eta)} (1 - \frac{\zeta_m}{\sum_{k\neq m} q_k}) e^{\mu_{\chi_m}(\rho\eta)} \\ &\leq e^{\max\{\beta_{\chi_m}(P,\rho,\eta),\mu_{\chi_m}(\rho\eta)\}} \end{aligned}$$

where $\forall i \in \mathcal{X}$, $\beta_i(P,\rho,\eta)$ and $\mu_i(\rho\eta)$ are defined as,

$$\beta_i(P,\rho,\eta) = \ln \sum_y W(y|i)^{(1-\rho\eta)} \left(\sum_x P(x) W(y|x)^\eta\right)^\rho$$

$$\mu_i(\rho\eta) = \max_{x\neq i, x\in\mathcal{X}} \ln \sum_y W(y|i)^{(1-\rho\eta)} W(y|x)^{\rho\eta}$$

Consequently for $\rho \geq 1$, $\eta \geq 0$,

$$\begin{aligned} \Upsilon_{\rho,\eta}(q,\chi_P) &= \frac{\sum_{y,m} W(y|\chi_m)^{(1-\rho\eta)} {q_m}^{\frac{1-\rho\eta}{\eta}} (\sum_{k\neq m} W(y|\chi_k)^\eta q_k)^\rho}{\sum_m {q_m}^{\frac{1-\rho\eta}{\eta}} (\sum_{k\neq m} q_k)^\rho} \\ &= \frac{\sum_m {q_m}^{\frac{1-\rho\eta}{\eta}} (\sum_{k\neq m} q_k)^\rho f_m}{\sum_m {q_m}^{\frac{1-\rho\eta}{\eta}} (\sum_{k\neq m} q_k)^\rho} \\ &\leq e^{\max_{x\in suppP(\cdot)} \max\{\beta_x(P,\rho,\eta),\mu_x(\rho\eta)\}} \end{aligned}$$

Thus for $\rho \geq 1$ for all input distributions P and for all $\eta > 0$,

$$\ln \alpha(\rho,\eta,\mathbb{X}_P) \leq \max_{x\in suppP(\cdot)} \max\{\beta_x(P,\rho,\eta),\mu_x(\rho\eta)\} \quad (18)$$

For certain channels property given in equation (18) together with Lemma 1 implies that sphere packing exponent is achievable on an interval of the form $[R_{Dcrit}, R_{crit}]$.

Corollary 2: If for a DMC,

$$\max_{x\in\mathcal{X}} \mu_x(\tfrac{\rho}{1+\rho}) \leq -E_0(\rho) \quad (19)$$

on an interval of the form $\rho \in [1, \rho_{Dc}]$. Then

$$\ln \alpha(\rho, \tfrac{1}{1+\rho}, \mathbb{X}) = -E_0(\rho) \quad \forall \rho \in [1, \rho_{Dc}]$$

Proof: In order to see this, first recall that for the $P(\cdot)$ that maximizes $E_0(\rho, P)$, i.e. P^* satisfying

$$\max_P E_0(\rho, P) = E_0(\rho, P^*)$$

we have,

$$E_0(\rho, P^*) = \beta_x(P^*, \tfrac{1}{1+\rho}, \rho) \qquad \forall x \in supp(P*)$$

Now the statement simply follows the equation (18) ■

Recall that sphere packing exponent is given

$$E_{sp}(R) = \max_{\rho\geq 0} E_0(\rho) - \rho R$$

Thus for each R there is a corresponding ρ_R, which is the maximizer of the the expression above. As a result of Lemma1 and Corollary2 for the rates with $\rho_R \in [1, \rho_{Dc}]$ sphere packing exponent will be achievable as error exponent for the channels satisfying condition given in equation (19). Two family of channels that satisfy the condition (19) are Binary input channels and K-ary symmetric channels.

2) Binary Input Channel: The binary input channel case have also been addressed by D'yachkov. We simply rederive his results here. A binary input channel is DMC for which $\mathcal{X} = \{0, 1\}$

For $\rho \geq 1$, $\eta \geq 0$, using equation (18) we get,

$$\alpha(\rho,\eta,\mathbb{X}) \leq e^{\max\{\beta_0(P,\eta,\rho),\beta_1(P,\eta,\rho),\mu_0(\rho\eta),\mu_1(\rho\eta)\}} \quad (20)$$

For the rates for which $\rho_R \in [1, \rho_{Dc}]$ this will lead to sphere packing exponent. For the rates for which $\rho_R > \rho_{Dc}$ we will simply minimize the expression in equation (20) over P,η and ρ to find the best possible error exponent achievable with this scheme.

For the rates such that $\rho_R \in (0, 1)$ one can also show that sphere packing exponent is achievable. For that one needs to consider,

$$\begin{aligned} f_m &= \sum_y W(y|\chi_m)^{(1-\rho\eta)} (\frac{\sum_{k\neq m} W(y|\chi_k)^\eta q_k}{\sum_{k\neq m} q_k})^\rho \\ &= \sum_y W(y|\chi_m)^{(1-\rho\eta)} (z_m W(y|\chi_k)^\eta + (1 - z_m) W(y|\bar{\chi}_k)^\eta)^\rho \end{aligned}$$

It can be shown that, for $\rho < 1$ and $\eta < 1$, $\frac{\partial}{\partial z_m} f_m \geq 0$. Furthermore as a result of equation (16) we know that $z_m \leq P(\chi_m)$. Thus for $\rho < 1$, $\eta < 1$,

$$f_m \leq e^{\beta_{\chi_m}(P,\rho,\eta)}$$

$$\alpha(\rho,\eta,\mathbb{X}) \leq e^{\max\{\beta_0(P,\eta,\rho),\beta_1(P,\eta,\rho)\}}$$

If we consider P^* that is the maximizer of $E_0(\rho, P)$, we get

$$\ln \alpha(\rho, \tfrac{1}{1+\rho}, \mathbb{X}) \leq -E_0(\rho)$$

Recalling the definition of ρ_R and $E_{sp}(R)$ we conclude that sphere packing exponent is achievable for R such that $\rho_R in (0, 1)$.

3) K-ary Symmetric Channel: Let us consider K-ary symmetric channel with $\epsilon < \frac{K-1}{K}$, i.e.

$$w(i|j) = \begin{cases} 1-\epsilon & i = j \\ \frac{\epsilon}{K-1} & i \neq j \end{cases}$$

Note that for any $\rho > 0, \eta > 0$ and x we have

$$\begin{aligned} \mu(\rho\eta) &= \mu_x(\rho\eta) \\ &= (1-\epsilon)\left(\frac{\epsilon}{(K-1)(1-\epsilon)}\right)^{\rho\eta} + \frac{\epsilon}{K-1}\left(\frac{(K-1)(1-\epsilon)}{\epsilon}\right)^{\rho\eta} + \frac{\epsilon(K-2)}{K-1} \end{aligned}$$

Furthermore note that for any $\rho > 0, \eta > 0$, $x \in \mathcal{X}$ and for $P(x) = 1/K$

$$\begin{aligned} \beta(\rho,\eta) &= \beta_x(P,\rho,\eta) \\ &= [(1-\epsilon)^{1-\rho\eta} + (K-1)(\tfrac{\epsilon}{K-1})^{1-\rho\eta}](\tfrac{(1-\epsilon)^\eta}{K} + \tfrac{K-1}{K}(\tfrac{\epsilon}{K-1})^\eta)^\rho \end{aligned}$$

Thus as a result of equation (18), for $\rho \geq 1$

$$\alpha(\rho,\eta,\mathbb{X}) \leq \max\{\beta(\rho,\eta), \mu(\rho\eta)\} \quad (21)$$

For $K = 2$ case these expressions are equivalent to those of Zigangirov in [8], which were specifically derived for BSCs. For $K \geq 3$, these expressions lead to $\alpha(\rho,\eta)$'s that are strictly better than the corresponding quantities in [4] for all $\rho > 1$. Figure **??** shows the resulting error exponents for a particular channel.

In [4], equivalent of equation (21), has $\beta_D(\rho,\eta)$ instead of $\beta(\rho,\eta)$ where,

$$\beta_D(\rho,\eta) = (1-\epsilon)^{1-\rho\eta}(\tfrac{(1-\epsilon)^\eta}{K} + \tfrac{K-1}{K}(\tfrac{\epsilon}{K-1})^\eta)^\rho \\ + (K-1)(\tfrac{\epsilon}{K-1})^{1-\rho\eta}(\tfrac{(1-\epsilon)^{\eta\rho}}{K} + \tfrac{K-1}{K}(\tfrac{\epsilon}{K-1})^{\eta\rho})$$

Since the function z^ρ is strictly convex in z for $\rho > 1$, $\beta(\rho,\eta) > \beta_D(\rho,\eta)$ for all $\rho > 1$ and $\eta > 0$.

4) Remarks on Z-D Encoding Scheme: The two example channels we have considered does not reveal the main weakness of the Zigangirov-D'yachkov encoding scheme. Consider the following four-by-four symmetric channel $w(y|x)$,

$$W(y|x) = \begin{bmatrix} 1-\epsilon & \epsilon & 0 & 0 \\ 0 & 1-\epsilon & \epsilon & 0 \\ 0 & 0 & 1-\epsilon & \epsilon \\ \epsilon & 0 & 0 & 1-\epsilon \end{bmatrix}$$

The error exponent of this channel is equal to sphere packing exponent at all rates. In order to see why, consider any rate $R > 0$. We already know that $\forall R > 0$ there is a fixed list size L_R such that: for any $\mathbf{n} \geq 1$ there is a length $\mathbf{n}$ code with decoding list size L_R whose error probability satisfies $P_\mathbf{e} \leq e^{-E_{sp}(R)\mathbf{n}}$. Using $\log_2 L_R$ extra channel uses transmitter can specify the correct message among the decoded list, if the correct message is in the list. Since this extra time will become negligible as $\mathbf{n}$ increases we can conclude that sphere packing exponent is achievable for this channel, when there is feedback.

However when we consider the Zigangirov-D'yachkov encoding scheme we can not reach the same conclusion. Clearly we should choose the $P(x)$ to be the uniform distribution. Consider for example the q, such that $q = [0.5\ 0.5\ 0\ 0\ \ldots]$ any smart encoding scheme would assign the first two messages into input letters that are not consecutive. But Zigangirov-D'yachkov encoding scheme will not necessarily do so. Considering similar q's for low enough ρ's one can show that Z-D encoding scheme leads to an $\alpha(\rho, \frac{1}{1+\rho})$ such that $\ln \alpha(\rho, \frac{1}{1+\rho}) > -E_0(\rho)$. As a result this will imply that Z-D encoding scheme will have an error exponent strictly less than the sphere pacing exponent.

On the other hand for all such anomalies we observed, we were also able to find a modification of Z-D encoding scheme, which employs a smarter encoding scheme for the first $|\mathcal{X}|$ messages, and which performs optimally. However we have yet to find a general modification on Z-D encoding scheme that works for all $q \in \mathcal{Q}$.

VI. Conclusions and Future Work

A part from the minor improvement for k-ary symmetric channels our results are essentially a rederivations of the results of Zigangirov, [8] and D'yachkov, [4]. However simplicity of the analysis does help us to see how to generalize the Zigangirov-D'yachkov encoding scheme to the general DMCs.

On a seperate note Burnashev, [2], considered the binary symmetric channels and showed that for all the rates between 0 and R_{Zcrit} one can reach and an error exponent strictly higher than the ones given in [8]. Indeed a similar modification are possible within our frame work too. We will present those in our journal paper on this subject.

References

[1] E. R. Berlekamp. *Block Coding with Noiseless Feedback*. Ph.D. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1964.

[2] M. V. Burnashev. On the reliability function of a binary symmetrical channel with feedback. *Problemy Perdachi Informatsii*, 24, No. 1:3–10, 1988.

[3] 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.

[4] A. G. D'yachkov. Upper bounds on the error probability for discrete memoryless channels with feedback. *Problemy Perdachi Informatsii*, Vol 11, No. 4:13–28, 1975.

[5] R.G. Gallager. A simple derivation of the coding theorem and some applications. *Information Theory, IEEE Transactions on*, 11, No. 1:3–18, 1965.

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

[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.

[8] K. Sh. Zigangirov. Upper bounds for the error probability for channels with feedback. *Problemy Perdachi Informatsii*, Vol 6, No. 2:87–92, 1970.