Upper Bounds to Error Probability with Feedback

Barış Nakiboğlu
Research Laboratory of Electronics at
Massachusetts Institute of Technology
Cambridge, MA 02139
nakib@mit.edu

Lizhong Zheng
Research Laboratory of Electronics at
Massachusetts Institute of Technology
Cambridge, MA 02139
lizhong@mit.edu

***Abstract*— A new analysis technique is suggested for bounding the error probability of fixed length block codes with feedback on discrete memoryless channels from above. Error analysis is inspired by Gallager's error analysis for block codes without feedback. Using Burnashev-Zigangirov-D'yachkov encoding scheme analysis recovers previously known best results on binary symmetric channels and improves up on the previously known best results on k-ary symmetric channels and binary input channels.**

I. Introduction

Shannon showed [9] that capacity of the discrete memoryless channels (DMCs) does not increase with feedback. Later Dobrushin [4] showed that the exponential decay rate of the error probability of fixed length block codes can not exceed sphere packing exponent in symmetric channels.[1] In other words for the rates above the critical rate, at least for symmetric channels, 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 pressing open question in this stream of research.[2]

The first work on the error analysis of block codes with feedback was by Berlekamp, [1]. He obtained a closed form expression of the error exponent at zero rate for binary symmetric channels (BSCs). Later Zigangirov [10] proposed an encoding scheme, for BSCs which reaches sphere packing exponent for all rate larger than a critical rate R_{Zcrit}.[3] Furthermore at zero rate Zigangirov's encoding scheme reaches optimal error exponent, which is derived by Berlekamp in [1]. Later D'yachkov [5] proposed a generalization of the encoding scheme of Zigangirov, and obtained a coding theorem for general DMCs. However the optimization problem resulting from his coding theorem, is quite involved and does not allow for simplifications that will lead to conclusions about the error exponents of general DMCs. In [5] 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 families of channels.

In [8] we have derived an upper bound on error probability for general DMCs, using an analysis technique similar to Gallager's in [6]. However like D'yachkov's expression in [5] our expression was hard to compute, so we have focused on examples: k-ary symmetric channels and binary input channels. We showed that the analysis technique proposed was able to recover D'yachkov's and Zigangirov's result on binary input channels and improve D'yachkov's results on k-ary input channels.

However Burnashev [2] had already improved Zigangirov's results [10] in binary symmetric channels. In this work we suggest a modification to the analysis technique we have presented in [8] to get the improvement corresponding to the Burnashev's in general DMCs. We keep track of the likelihoods of the messages using a stopping time in order to avoid making a worst case assumption like the one we did in [8], at least in some part of the block. Furthermore in order to accommodate different tilting factors in the encoding at different times we use a weighted maximum likelihood decoder instead of a maximum likelihood decoder. However resulting optimization problem is again hard. Thus we focus on k-ary symmetric and binary input channels to demonstrate the improvements we have established.

We start by introducing the channel model and the notation in section II. After that in section III we do the first part of our error analysis inspired by Gallager's analysis in [6] and Burnashev's analysis in [2]. Then we specify the feedback encoding scheme in section IV. After that we come back to our error analysis and derive a parametric expression for the achievable error exponent in section V. These expressions improves upon the previously known best results reported before, [8], in all channels.[4] Finally in section VI we mention the aspects of the problem we are investigating currently

II. Channel Model and Notation

We have a discrete memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$. Channel transition probabilities are given by a $|\mathcal{X}|$-by-$|\mathcal{Y}|$ matrix $W(y|x)$. In addition we

[1] After that Haratounian [7] established an upper bound for the error exponent for non-symmetric channels as a generalization of Dobrushin's result, but his upper bound is strictly larger than the sphere packing exponent for non-symmetric channels.

[2] 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 because understanding those variants will not help the reader much in understanding the work at hand.

[3] 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.

[4] Evidently with the exception BSCs. For BSCs best results are Burnashev's in [2] we are merely recovering his results for BSCs.

assume that a noiseless, delay free feedback link exists from the receiver to the transmitter.

Receiver sends the channel output at time t, Y_t and an additional random variable of its choice, Λ_t, to the transmitter at each time t. The transmitter receives the feedback link symbol for time t, $Z_t = (Y_t, \Lambda_t)$ before the transmission of the input symbol at time $t+1$. A feedback encoding scheme[5] Ψ is a sequence $(\Psi_1, \Psi_2, \ldots, \Psi_{\mathbf{n}})$ of mappings such that

$$\Psi_t(\cdot) : \mathcal{M} \times \mathcal{Z}^{t-1} \to \mathcal{X} \qquad \forall t \in \{1, 2, \ldots, \mathbf{n}\}$$

The input letter for the message $m \in \mathcal{M}$ at time t given $z^{t-1} \in \mathcal{Z}^{t-1}$ is $\Psi_t(m, z^{t-1})$. Note that when there is no feedback $\Psi_t(m, z^{t-1}) = \Psi_t(m)$, $\forall z^{t-1} \in \mathcal{Z}^{t-1}$. The probability of observing Z^t conditioned on message θ is,

$$\mathbf{P}\left\{Z^t \middle| \theta\right\} = \prod_{j=1}^{t} W(Y_j|\Psi_j(\theta, Z^{j-1})) \cdot \mathbf{P}\left\{\Lambda_j \middle| Z^{j-1}, Y_j\right\}.$$

A decoding rule is for a mapping of the form

$$\Phi(\cdot) : \mathcal{Z}^{\mathbf{n}} \to \mathcal{M}.$$

We denote the set of all $|\mathcal{M}|$-long sequence of non-negative numbers by $\mathcal{Q}^{|\mathcal{M}|}$. Thus at any time t both the posterior probability distribution or the likelihood vector of the messages are in $\mathcal{Q}^{|\mathcal{M}|}$.

III. Error Analysis Part I: Stopping Time and Weighted Maximum-Likelihood Decoding

Two main drawbacks of the error analysis we have introduced in [8] were, the use of worst case bound over the space of possible posterior probability distributions and use of fixed tilting factor η for the encoding throughout the block. In order to be able to address the first issue we keep track of the likelihoods of the messages, using a stopping time. In order to be able address the second issue we use weighted maximum likelihood decoding instead of maximum likelihood decoding. In order to accommodate these changes we need to modify the analysis technique we proposed in [8] this section is devoted to that end.

Let τ be a stopping time with respect to the stochastic sequence $Y_1, (\Lambda_1, Y_2), (\Lambda_2, Y_3), \ldots$, i.e. with respect to the receivers observation. For each t let ζ_t be a high probability subset set of $\mathcal{Z}^t$ to be determined later.[6] Let ζ^τ be the set of Z^τ such that all subsequences are in the corresponding high probability subset and $\overline{\zeta^\tau}$ be its complement, i.e.

$$\zeta^\tau = \{Z^\tau : \forall t \le \tau, Z^t \in \zeta_t\} \tag{1}$$

$$\overline{\zeta^\tau} = \{Z^\tau : Z^\tau \notin \zeta^\tau\}. \tag{2}$$

[5]Indeed this additional random variable is not necessary, in the sense that any performance achievable using a feedback symbol of the form $Z_t = (Y_t, \Lambda_t)$ is also achievable, using a feedback symbol of the form $Z_t = Y_t$. However introducing this extra random variable simplifies the analysis a great deal.

[6]Say with probability $\mathbf{P}\{\zeta_t\} = 1 - e^{-\mathbf{n}^2}$.

Evidently $\mathbb{I}_{\{\zeta^\tau\}} + \mathbb{I}_{\{\overline{\zeta^\tau}\}} = 1$. Thus,

$$\begin{aligned} P_{\mathbf{e}} &= \mathbf{E}\left[\mathbb{I}_{\{\hat{\theta}\neq\theta\}}\mathbb{I}_{\{\overline{\zeta^\tau}\}}\right] + \mathbf{E}\left[\mathbb{I}_{\{\hat{\theta}\neq\theta\}}\mathbb{I}_{\{\zeta^\tau\}}\right] \\ &\le \mathbf{P}\left\{\overline{\zeta^\tau}\right\} + \mathbf{E}\left[\mathbb{I}_{\{\hat{\theta}\neq\theta\}}\mathbb{I}_{\{\zeta^\tau\}}\right] \end{aligned} \tag{3}$$

Note that $\mathbb{I}_{\{\tau>\mathbf{n}\}} + \sum_{t=1}^{\mathbf{n}} \mathbb{I}_{\{\tau=t\}} = 1$, thus

$$\mathbf{E}\left[\mathbb{I}_{\{\hat{\theta}\neq\theta\}}\mathbb{I}_{\{\zeta^\tau\}}\right] \le P_{\mathbf{en}}^* + \sum_{t=1}^{\mathbf{n}} P_{\mathbf{e}t} \tag{4}$$

where

$$P_{\mathbf{en}}^* = \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau>\mathbf{n}\}}\mathbb{I}_{\{\hat{\theta}(Z^{\mathbf{n}})\neq\theta\}}\right] \tag{5}$$

$$P_{\mathbf{e}t} = \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau=t\}}\mathbb{I}_{\{\hat{\theta}(Z^{\mathbf{n}})\neq\theta\}}\right] \tag{6}$$

Thus

$$P_{\mathbf{e}} \le \mathbf{P}\left\{\overline{\zeta^\tau}\right\} + P_{\mathbf{en}}^* + \sum_{t=1}^{\mathbf{n}} P_{\mathbf{e}t} \tag{7}$$

Clearly best error performance is obtained by maximum likelihood decoder for any encoding scheme. However, when used in conjunction with the bounds we employ, it is not at all clear why same should hold. Indeed allowing for a weighted maximum likelihood decoder as described below gives us better performance,

$$\hat{\theta}(Z^{\mathbf{n}}) = \arg\max_{\theta} \mathbf{P}\left\{Z_{\tau+1}^{\mathbf{n}} \middle| \theta, Z^\tau\right\} \mathbf{P}\left\{Z^\tau \middle| \theta\right\}^{\alpha}. \tag{8}$$

For the decoder given in equation (8), for any $Z^{\mathbf{n}}$, $\eta > 0$ and $\rho \ge 0$ we have,

$$\mathbb{I}_{\{\hat{\theta}(Z^{\mathbf{n}})\neq\theta\}} \le \left(\sum_{m\neq\theta} \frac{\mathbf{P}\left\{Z_{\tau+1}^{\mathbf{n}}\middle|m, Z^\tau\right\}^{\eta}\mathbf{P}\{Z^\tau|m\}^{\alpha\eta}}{\mathbf{P}\left\{Z_{\tau+1}^{\mathbf{n}}\middle|\theta, Z^\tau\right\}^{\eta}\mathbf{P}\{Z^\tau|\theta\}^{\alpha\eta}}\right)^{\rho}. \tag{9}$$

For our later convenience instead of α we work with $\beta = \alpha\eta$. For any $\eta > 0$, α can take any positive value, consequently so does β. Let us define the weighted likelihood of the messages for a given $\eta > 0$, $\beta > 0$, and τ as

$$\phi(Z^t|m) = \begin{cases} \mathbf{P}\left\{Z^t \middle| m\right\}^{\beta} & t \le \tau \\ \mathbf{P}\left\{Z_{\tau+1}^t \middle| m, Z^\tau\right\}^{\eta} \mathbf{P}\left\{Z^\tau \middle| m\right\}^{\beta} & t > \tau \end{cases} \tag{10}$$

Above upper bound on the indicator function of the error event still depends on the transmitted message θ, i.e. it is not known at the receiver. In order to get rid of that dependence let us define ξ_t,

$$\xi_t = \mathbf{E}\left[\left(\sum_{m\neq\theta} \frac{\phi(Z^t|m)}{\phi(Z^t|\theta)}\right)^{\rho} \middle| Z^{t-1}, Y_t\right]$$

Thus we get

$$P_{\mathbf{en}}^* \le \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau>\mathbf{n}\}}\xi_{\mathbf{n}}\right] \tag{11}$$

$$\begin{aligned} P_{\mathbf{e}t} &\le \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau=t\}}\xi_{\mathbf{n}}\right] \\ &= \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau=t\}}\mathbf{E}\left[\xi_{\mathbf{n}}\middle| Z^{t-1}, Y_t\right]\right]. \end{aligned} \tag{12}$$

Let us assume for the moment that there exist an encoding scheme such that

$$\mathbf{E}\left[\xi_{t+1}\middle| Z^{t-1}, Y_t\right] \le G(\rho, \eta)\xi_t \qquad \forall Z^{t-1}, Y_t \text{ s.t. } t \ge \tau.$$

Then for any $(Z^{\tau-1}, Y_\tau)$ such that $\tau \le \mathbf{n}$ we have

$$\mathbf{E}\left[\xi_{\mathbf{n}} \middle| Z^{\tau-1}, Y_\tau\right] \le G(\rho,\eta)^{(\mathbf{n}-\tau)}\xi_\tau. \tag{13}$$

Using equations (12) and (13) we get the following upper bound on $P_{\mathbf{e}t}$

$$\begin{aligned} P_{\mathbf{e}t} &\le \mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau=t\}}G(\rho,\eta)^{(\mathbf{n}-\tau)}\xi_\tau\right] \\ &\le G(\rho,\eta)^{(\mathbf{n}-t)}\mathbf{E}\left[\mathbb{I}_{\{\zeta^\tau\}}\mathbb{I}_{\{\tau=t\}}\xi_\tau\right]. \end{aligned} \tag{14}$$

In order to bound the error probability further we need to specify the high probability sets ζ_t and the encoding scheme, i.e. the stopping time τ, and the encoding scheme for the interval $[0,\tau]$ and $[\tau+1,\mathbf{n}]$. We do that in the next section and we continue to derive the upper bound after that in section V.

IV. Encoding Scheme:

A. Stopping time:

Note that using the "tilted" likelihood $\mathbf{P}\left\{Z^t\middle|\cdot\right\}^\beta$ we can define "tilted" posterior probability distribution. The stopping time $\tau(\beta,\epsilon)$ is the first time instance at which a message reaches a "tilted" posterior probability higher than ϵ.

$$\tau(\beta,\epsilon) \triangleq \min\left\{t : \max_{m\in\mathcal{M}} \frac{\mathbf{P}\left\{Z^t\middle| m\right\}^\beta}{\sum_{\tilde m}\mathbf{P}\left\{Z^t\middle|\tilde m\right\}^\beta} \ge \epsilon\right\} \tag{15}$$

Note that for any t receiver chooses Λ_t after observing (Z^{t-1}, Y_t) but without knowing the transmitted message θ, thus given (Z^{t-1}, Y_t), Λ_t is independent of the transmitted message θ,

$$\mathbf{P}\left\{Z^t\middle| m\right\} = \mathbf{P}\left\{Z^{t-1}, Y_t\middle| m\right\}\cdot\mathbf{P}\left\{\Lambda_t\middle| Z^{t-1}, Y_t\right\}.$$

As a result receiver know whether $\tau > t$ or not before it draws Λ_t, i.e. τ is a stopping time with respect to the stochastic sequence $Y_1, (\Lambda_1, Y_2), (\Lambda_2, Y_3), \ldots$ as we have assumed. Note that the largest tilted posterior distribution of the message can be smaller than ϵ in first $\mathbf{n}$ times for some $Z^{\mathbf{n}}$, in this case $\tau(\beta,\epsilon) > \mathbf{n}$.

B. Encoding in $[0,\tau]$: Random Coding

Recall that for each history Z^{t-1} the encoding scheme at time t is a mapping of messages to the input letters. For all (Z^{t-2}, Y_{t-1}) such that $\tau \ge t$ we use Λ_{t-1} to choose this mapping randomly. For each (Z^{t-2}, Y_{t-1}) each message is assigned to an input letter x for time t with probability $P(x)$. Furthermore given (Z^{t-2}, Y_{t-1}) assignments of the messages are independent of one another, i.e.

$$\begin{aligned} &\mathbf{P}\left\{\Psi_t(\cdot, Z^{t-1}) = \mathbb{X}\middle| Z^{t-2}, Y_{t-1}\right\} \\ &\quad= \prod_{m\in\mathcal{M}}\mathbf{P}\left\{\Psi_t(m, Z^{t-1}) = \mathbb{X}_m\middle| Z^{t-2}, Y_{t-1}\right\} \\ &\quad= \prod_{m\in\mathcal{M}} P(\mathbb{X}_m) \qquad \forall\mathbb{X}\in\mathcal{X}^{\mathcal{M}} \end{aligned} \tag{16}$$

For such an encoding if $t \le \tau$ then for almost all Λ_{t-1}'s

$$\sum_{m:\Psi_t(m,Z^{t-1})=x}\phi(Z^{t-1}|m) \approx P(x)\sum_m \phi(Z^{t-1}|m)$$

The main idea here is that if $\tau \ge t$ then tilted posterior probability of each message is small, i.e. less than ϵ, and there are many of them, i.e. $|\mathcal{M}|$. Thus if we assign each one of them to the input letter x with probability $P(x)$ independently the total "tilted" posterior probability of the messages that are assigned to input letter x will be very close $P(x)$ most of the time. More precisely,

Lemma 1: Let ζ_{t-1}, $\zeta_{t-1}^{(a)}$, $\zeta_{t-1}^{(b)}$ be

$$\zeta_{t-1} = \zeta_{t-1}^{(a)} \cup \zeta_{t-1}^{(b)}$$

$$\zeta_{t-1}^{(a)} = \left\{Z^{t-1} : \left|\frac{\sum_{m:\Psi_t(m,Z^{t-1})=x}\phi(Z^{t-1}|m)}{\sum_m \phi(Z^{t-1}|m)} - P(x)\right| \le \frac{P(x)}{\mathbf{n}} \quad \forall x\in\mathcal{X}; \tau \ge t\right\}$$

$$\zeta_{t-1}^{(b)} = \left\{Z^{t-1} : \tau < t\right\}$$

then

$$\mathbf{P}\left\{Z^{t-1}\notin\zeta_{t-1}\middle| Z^{t-2}, Y_{t-1}\right\} \le 2|\mathcal{X}|e^{-\frac{\mathbf{n}^2}{2}} \qquad \tau \ge t \tag{17a}$$

$$\mathbf{P}\left\{Z^{t-1}\notin\zeta_{t-1}\middle| Z^{t-2}, Y_{t-1}\right\} = 0 \qquad \tau > t \tag{17b}$$

Proof:

Note that (17b) is trivially follows the definition of $\zeta_{t-1}^{(b)}$, so we focus on (17a). Let a_m, $\overline{a_m}(x)$ and $\sigma(a_m(x))$ be

$$a_m(x) = \mathbb{I}_{\{\Psi_t(m,Z^{t-1})=x\}}\frac{\phi(Z^{t-1}|m)}{\mathbf{P}\{\Lambda_{t-1}|Z^{t-2},Y_{t-1}\}^\beta} \tag{18a}$$

$$\overline{a_m}(x) = \mathbf{E}\left[a_m(x)\middle| Z^{t-2}, Y_{t-1}\right] \tag{18b}$$

$$\sigma(a_m(x))^2 = \mathbf{E}\left[(a_m(x)-\overline{a_m}(x))^2\middle| Z^{t-2}, Y_{t-1}\right] \tag{18c}$$

Then

$$\overline{a_m}(x) = \frac{\phi(Z^{t-1}|m)}{\mathbf{P}\{\Lambda_{t-1}|Z^{t-2},Y_{t-1}\}^\beta}P(x) \tag{19a}$$

$$\begin{aligned} \sigma(a_m(x))^2 &= \left(\frac{\phi(Z^{t-1}|m)}{\mathbf{P}\{\Lambda_{t-1}|Z^{t-2},Y_{t-1}\}^\beta}\right)^2 P(x)(1-P(x)) \\ &\le \frac{\phi(Z^{t-1}|m)\epsilon\sum_{\tilde m}\phi(Z^{t-1}|\tilde m)P(x)(1-P(x))}{\mathbf{P}\{\Lambda_{t-1}|Z^{t-2},Y_{t-1}\}^{2\beta}}. \end{aligned} \tag{19b}$$

where the inequality follows from the fact that $\forall(Z^{t-2}, Y_{t-1})$ such that $\tau \ge t$ $\max_m \phi(Z^{t-1}|m) < \epsilon\sum_m\phi(Z^{t-1}|m)$. As result of [3, Theorem 5.3] we have,

$$\frac{\mathbf{P}\left\{\left|\sum_m \tilde a_m(x)\right| \ge \lambda\middle| Z^{t-2}, Y_{t-1}\right\}}{2} \le e^{-\frac{\lambda^2}{2\sum_m\sigma(a_m(x))^2}}$$

If we choose $\lambda = \mathbf{n}^{-1}\sum_m\overline{a_m}(x)$, $\epsilon = \mathbf{n}^{-4}\min_x P(x)$ apply union bound over $x\in\mathcal{X}$, lemma 1 follows from equations (18) and (19).

QED

Evidently using lemma 1 we can bound the probability of $\overline{\zeta^\tau}$ defined in equations (1), (2) from above as follows

$$\mathbf{P}\left\{\overline{\zeta^\tau}\right\} \le 2|\mathcal{X}|\mathbf{n}e^{-\frac{\mathbf{n}^2}{2}} \tag{20}$$

Note that on the other hand $P_{\mathbf{e}t}$ terms contributing to the upper bound in equation (7) are themselves upper bounded by equation (14). Below we bound ξ_τ from above for $Z^\tau\in\zeta^\tau$ to bound (14). Note that for all $Z^{t-1}\in\zeta_{t-1}$ and $Y_t\in\mathcal{Y}$ we have

$$\left|\frac{\sum_m\mathbf{P}\left\{Z^{t-1},Y_t\middle| m\right\}^\beta}{\sum_m\mathbf{P}\{Z^{t-1}|m\}^\beta} - \sum_x W(Y_t|x)^\beta P(x)\right| \le \frac{\sum_x W(Y_t|x)^\beta P(x)}{\mathbf{n}}$$

Thus for all $Z^\tau \in \zeta^\tau$ we have

$$\begin{aligned}
\tfrac{\sum_m \mathbf{P}\{Z^\tau|m\}^\beta}{\mathbf{P}\{Z^\tau|\theta\}^\beta} &\leq \left(1+\tfrac{1}{\mathbf{n}}\right) \tfrac{\sum_x W(Y_\tau|x)^\beta P(x)}{\mathbf{P}\{Y_\tau|\theta,Z^{\tau-1}\}^\beta} \tfrac{\sum_m \mathbf{P}\left\{Z^{\tau-1}\middle|m\right\}^\beta}{\mathbf{P}\{Z^{\tau-1}|\theta\}^\beta} \\
&\leq e^{\mathbf{n}R}\left(1+\tfrac{1}{\mathbf{n}}\right)^\tau \prod_{t=1}^{\tau} \tfrac{\sum_x W(Y_t|x)^\beta P(x)}{\mathbf{P}\{Y_t|\theta,Z^{t-1}\}^\beta} \\
&\overset{(a)}{\leq} e^{\mathbf{n}R+1} \prod_{t=1}^{\tau} \tfrac{\sum_x W(Y_t|x)^\beta P(x)}{\mathbf{P}\{Y_t|\theta,Z^{t-1}\}^\beta} \\
&\leq e^{\mathbf{n}R+1}\Gamma_\tau(\theta) \qquad (21)
\end{aligned}$$

where (a) follows $(1+1/\mathbf{n})^\tau \leq (1+1/\mathbf{n})^{\mathbf{n}} \leq e$ and $\Gamma_\tau(\theta)$ is the value of $\Gamma_t(m)$ at $t=\tau$ and $m=\theta$ which is defined as

$$\Gamma_t(m) = \prod_{\ell=1}^{t} \tfrac{\sum_x W(Y_\ell|x)^\beta P(x)}{\mathbf{P}\{Y_\ell|m,Z^{\ell-1}\}^\beta}. \qquad (22)$$

Evidently for any Z^τ,

$$\sum_{m\neq\theta} \mathbf{P}\left\{Z^\tau\middle| m\right\}^\beta \leq \sum_m \mathbf{P}\left\{Z^\tau\middle| m\right\}^\beta$$

Consequently for all $Z^\tau \in \zeta^\tau$

$$\begin{aligned}
\xi_\tau &\leq \mathbf{E}\left[\left(\tfrac{\sum_m \mathbf{P}\{Z^\tau|m\}}{\mathbf{P}\{Z^\tau|\theta\}}\right)^\rho \middle| Z^\tau\right] \\
&= \mathbf{E}\left[\Gamma_\tau(\theta)^\rho\middle| Z^\tau\right] e^{\rho(\mathbf{n}R+1)} \qquad (23)
\end{aligned}$$

Evidently above analysis implies for $Z^{\mathbf{n}} \in \zeta^{\mathbf{n}}$, $\tau > \mathbf{n}$ that

$$\xi_{\mathbf{n}} \leq \mathbf{E}\left[\Gamma_{\mathbf{n}}(\theta)^\rho\middle| Z^{\mathbf{n}}\right] e^{\rho(\mathbf{n}R+1)} \qquad (24)$$

C. Encoding in $[\tau+1, \mathbf{n}]$:

In section III we have assumed that there exists a feedback encoding scheme such that

$$\mathbf{E}\left[\xi_{t+1}\middle| Z^{t-1}, Y_t\right] \leq G(\rho,\eta)\xi_t \quad \forall Z^{t-1}, Y_t \text{ s.t. } t \geq \tau.$$

In this section we derive achievable values for $G(\rho,\eta)$. These achievable values are used in section V while we are deriving an upper bound to the error probability $P_{\mathbf{e}}$. In subsection IV-C.1 we formally describe an achievable value for $G(\rho,\eta)$ which we have not been able to simplify to get a single letter expression. In subsection IV-C.2 we describe a modified Zigangirov-D'yachkov(*Z-D*) encoding and obtain an expression for the resulting $G(\rho,\eta)$.

1) An Upper Bound on $G(\rho,\eta)$: Recall again that given Z^t encoding at time $(t+1)$, $\Psi(Z^t)$ is simply a mapping of messages to the input letters. Using the definition of $\phi(Z^t|m)$ given in equation (10) we get,

$$\begin{aligned}
\frac{\mathbf{E}\left[\xi_{t+1}\middle| Z^t\right]}{\xi_t} &= \frac{\mathbf{E}\left[\left(\frac{\sum_{m\neq\theta}\phi(Z^{t+1}|m)}{\phi(Z^{t+1}|\theta)}\right)^\rho\middle|Z^t\right]}{\mathbf{E}\left[\left(\frac{\sum_{m\neq\theta}\phi(Z^{t}|m)}{\phi(Z^{t}|\theta)}\right)^\rho\middle|Z^t\right]} \\
&= \upsilon_{\eta,\rho}(\mathbf{P}\left\{Z^t\middle|\cdot\right\}, \phi(Z^t|\cdot), \Psi_{t+1}(\cdot, Z^t))
\end{aligned}$$

where υ is defined for $\rho \geq 0$, $\eta \geq 0$, $q \in \mathcal{Q}^{|\mathcal{M}|}$, $p \in \mathcal{Q}^{|\mathcal{M}|}$ and $\mathbb{X} \in \mathcal{X}^{|\mathcal{M}|}$ as follows

$$\upsilon_{\eta,\rho}(q,p,\mathbb{X}) = \frac{\sum_{y,m} W(y|\mathbb{X}_m)^{1-\rho\eta} q_m p_m^{-\rho}\left(\sum_{\tilde{m}\neq m} W(y|\mathbb{X}_m)^\eta p_{\tilde{m}}\right)^\rho}{\sum_m q_m p_m^{-\rho}\left(\sum_{\tilde{m}\neq m} p_{\tilde{m}}\right)^\rho} \qquad (25)$$

Thus $\frac{\mathbf{E}\left[\xi_{t+1}|Z^t\right]}{\xi_t}$ is only function of the pair (η,ρ), "tilted" weighted likelihoods of the messages, i.e. $\phi(Z^t|\cdot)$, likelihoods of the messages, i.e. $\mathbf{P}\left\{Z^t\middle|\cdot\right\}$, and the mapping, $\Psi_{t+1}(\cdot,Z^t)$. Note that $\Psi_{t+1}(\cdot,Z^t)$ can depend on both $\phi(Z^t|\cdot)$ and $\mathbf{P}\left\{Z^t\middle|\cdot\right\}$ because given Z^t their values are known. If we chose the mapping at time $(t+1)$ to be,

$$\Psi_{t+1}(\cdot, Z^t) = \operatorname*{argmin}_{\mathbb{X}} \upsilon_{\eta,\rho}(\mathbf{P}\left\{Z^t\middle|\cdot\right\}, \phi(Z^t|\cdot), \mathbb{X})$$

we get

$$\frac{\mathbf{E}\left[\xi_{t+1}\middle| Z^t\right]}{\xi_t} = \min_{\mathbb{X}} \upsilon_{\eta,\rho}(\mathbf{P}\left\{Z^t\middle|\cdot\right\}, \phi(Z^t|\cdot), \mathbb{X}).$$

Evidently for any Z^t this value is upper bounded by the worst case value over $\mathcal{Q}^{|\mathcal{M}|} \times \mathcal{Q}^{|\mathcal{M}|}$. Then

$$\frac{\mathbf{E}\left[\xi_{t+1}\middle| Z^t\right]}{\xi_t} \leq \max_{q,p}\min_{\mathbb{X}} \upsilon_{\eta,\rho}(q,p,\mathbb{X}) \qquad \forall Z^t \qquad (26)$$

Note that although expression in equation (26) is a valid upper bound it is not a single letter expression.

2) Z-D Encoding Scheme: In this subsection we use *Z-D* encoding scheme via "tilted" weighted likelihoods, $\phi(Z^t|\cdot)$, to get explicit single letter upper bounds to $G(\rho,\eta)$. This encoding scheme was first described by Zigangirov [10] for binary symmetric channels then generalized by D'yachkov [5] to general DMCs. We have previously used this encoding on "tilted" likelihoods in [8].

Consider a probability distribution a $P(\cdot)$ on input alphabet $\mathcal{X}$ and a $p \in \mathcal{Q}^{|\mathcal{M}|}$. Without loss of generality we can assume that[7] $\forall m, \tilde{m} \in \mathcal{M}$, if $m \leq \tilde{m}$ then $p_m \geq p_{\tilde{m}}$. Now we can define mapping $\mathbb{X}$ for a given p and $P(\cdot)$ iteratively as follows:

$$\begin{aligned}
\gamma_0(x) &= 0 \qquad \forall x \in \mathcal{X} \\
\mathbb{X}_m &= \operatorname*{argmin}_{x\in\text{supp}(P)} \tfrac{\gamma_{m-1}(x)}{P(x)} \\
\gamma_m(x) &= \sum_{1\leq\tilde{m}\leq m:\mathbb{X}_{\tilde{m}}=x} p_{\tilde{m}}
\end{aligned}$$

For assigning $m \in \mathcal{M}$ 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, $\gamma_{m-1}(x)$. Then we divide $\gamma_{m-1}(x)$'s by the corresponding $P(x)$ values and assign the message $m \in \mathcal{M}$ to the $x \in \mathcal{X}$, for which $P(x) > 0$ and $\frac{\gamma_{m-1}(x)}{P(x)}$ 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.

A *Z-D* encoding scheme with $P(\cdot)$, will satisfy,

$$\chi_m = \tfrac{p_{\mathbb{X}_m}-p_m}{P(\mathbb{X}_m)} \leq \tfrac{p_x}{P(x)} \quad \forall x \in \mathcal{X} \quad \forall m \in \mathcal{M} \qquad (27)$$

where $p_x = \gamma_{|\mathcal{M}|}(x)$. 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 have at least the same mass as the last one, they will satisfy the property given in equation (27) too. Thus for any $p \in \mathcal{Q}^{|\mathcal{M}|}$

[7] If this is not the case for a p, we can rearrange the messages $m \in \mathcal{M}$, according to their p_m in decreasing order. If two or more messages have same mass, p, we order according to their indices.

and any input distribution $P(x)$, the mapping created by a *Z-D* encoding scheme, satisfies

$$p_x - P(x)\chi_m \geq 0 \quad \forall x \in \mathcal{X} \quad \forall m \in \mathcal{M} \tag{28}$$

Thus

$$\frac{\sum_{\tilde{m}\neq m} \mathbb{I}_{\{\mathbb{X}_{\tilde{m}}=x\}} p_{\tilde{m}}}{\sum_{\tilde{m}\neq m} p_m} = \frac{\chi_m P(x)}{\sum_{\tilde{m}\neq m} p_m} + \frac{\sum_{x\neq\mathbb{X}_m} \mathbb{I}_{\{\mathbb{X}_{\tilde{m}}=x\}} (p_x - \chi_m P(x))}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}}$$

In other words, with *Z-D* encoding scheme, the mass of the p 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 $x' \neq \mathbb{X}_m$. Using this decomposition of the input distribution together with the convexity of the function z^ρ for $\rho \geq 1$ and Jensen's inequality we get,

$$\begin{aligned}
&\left[\frac{\sum_{k\neq m} W(y|\mathbb{X}_k)^\eta q_k}{\sum_{k\neq m} q_k}\right]^\rho \\
&= \left[\frac{\sum_x W(y|x)^\eta \sum_{\tilde{m}\neq m} \mathbb{I}_{\{\mathbb{X}_{\tilde{m}}=x\}} p_{\tilde{m}}}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}}\right]^\rho \\
&= \left[\sum_x W(y|x)^\eta \left\{\frac{\chi_m P(x)}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}} + \sum_{x\neq\mathbb{X}_m} \frac{\mathbb{I}_{\{\mathbb{X}_{\tilde{m}}=x\}} (p_x - \chi_m P(x))}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}}\right\}\right]^\rho \\
&\leq \frac{\chi_m \left[\sum_x W(y|x)^\eta P(x)\right]^\rho}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}} + \sum_{x\neq\mathbb{X}_m} \frac{(p_x - \chi_m P(x)) W(y|x)^{\eta\rho}}{\sum_{\tilde{m}\neq m} p_{\tilde{m}}}
\end{aligned} \tag{29}$$

Using equation (27) in the definition of $\upsilon_{\eta,\rho}(q,p,\mathbb{X})$ given in equation (25), for $\rho \geq 1$ for all input distributions P and for all $\eta > 0$, we get

$$\upsilon_{\eta,\rho}(q,p,\mathbb{X}) \leq \max_{x\in \text{supp}\, P(\cdot)} \max\{\mu_x(P,\rho,\eta), \lambda_x(\rho\eta)\}$$

where $\forall x \in \mathcal{X}$

$$\mu_x(P,\rho,\eta) = \sum_y W(y|x)^{(1-\rho\eta)} \left(\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^\eta\right)^\rho$$

$$\lambda_x(\rho\eta) = \max_{\tilde{x}\neq x, \tilde{x}\in\text{supp}(P)} \sum_y W(y|x)^{(1-\rho\eta)} W(y|\tilde{x})^{\rho\eta}$$

When $\rho \in [0,1]$, we do random coding instead of $Z - D$, using concavity of s^ρ for $s \geq 0$ we get

$$\upsilon_{\eta,\rho}(q,p,\mathbb{X}) \leq \sum_x P(x)\mu_x(P,\rho,\eta)$$

V. Error Analysis Part II:

In this section we continue the error analysis we have started in section III, for the encoding scheme specified in section IV. We will start with bounding $P_{\mathbf{en}}^*$, then we bound $P_{\mathbf{et}}$ and proceed to combining the bounds we have established to bound the overall error probability, $P_{\mathbf{e}}$.

A. Bounding $P_{\mathbf{en}}^*$:

First recall equation (21): $\forall Z^t \in \zeta^t$, $t \leq \tau$, $m \in \mathcal{M}$, $\mathbf{P}\left\{Z^t|\, m\right\}^\beta / \sum_{\tilde{m}} \mathbf{P}\left\{Z^t|\, \tilde{m}\right\}^\beta \geq e^{-\mathbf{n}R}/\Gamma_t(m)e$. Thus,

$$\begin{aligned}
\mathbb{I}_{\{\tau>\mathbf{n}\}}\mathbb{I}_{\{Z^\mathbf{n}\in\zeta^\mathbf{n}\}} &\leq \mathbb{I}_{\{\max_m \frac{e^{-\mathbf{n}R}}{e\Gamma_t(m)} \leq \epsilon\}} \\
&\leq \mathbb{I}_{\{\frac{e^{-\mathbf{n}R}}{e\Gamma_t(\theta)} \leq \epsilon\}}
\end{aligned} \tag{30}$$

Using equations (11), (24) and (30) we get

$$\begin{aligned}
P_{\mathbf{en}}^* &\leq \mathbf{E}\left[\mathbb{I}_{\{\frac{e^{-\mathbf{n}R}}{e\Gamma_\mathbf{n}(\theta)} \leq \epsilon\}} \Gamma_\mathbf{n}(\theta)^\rho e^{\rho(\mathbf{n}R+1)}\right] \\
&\leq \mathbf{E}\left[(\epsilon e\Gamma_\mathbf{n}(\theta)e^{\mathbf{n}R})^{\lambda_0} \Gamma_\mathbf{n}(\theta)^\rho e^{\rho(\mathbf{n}R+1)}\right] \\
&\leq \epsilon^{\lambda_0} e^{\rho+\lambda_0} e^{\mathbf{n}R(\lambda_0+\rho)} H_0(\rho,\beta,\lambda_0)^\mathbf{n}
\end{aligned} \tag{31}$$

where

$$H_0(\rho,\beta,\lambda_0) = \sum_{y,x} P(x)W(y|x)\left(\frac{\sum_{\tilde{x}} P(\tilde{x})W(y|\tilde{x})^\beta}{W(y|x)^\beta}\right)^{\rho+\lambda_0} \tag{32}$$

Indeed there is a slight abuse of notation in the above array of inequalities. The expectations that leads to (31) are not the expectations resulting from the encoding scheme described in section IV in which after τ encoder stops using random coding and switches to *Z-D* encoding scheme. In the expressions leading to (31) it is assumed that random coding continues after τ. However for bounding the probability of $\tau \geq \mathbf{n}$ what happens after τ does not make a difference. This is why we can use the modified probability measure in the bounds. To put it more explicitly equations (11), (24) and (30) holds for expectations calculated in either way; consequently expectations calculated in either way gives us an upper bound on $P_{\mathbf{en}}^*$.

B. Bounding $P_{\mathbf{et}}$:

For $Z^t \in \zeta^t$, $t \leq \tau$, $m \in \mathcal{M}$, using an analysis very similar to the one leading to equation (21) we can lower bound $\mathbf{P}\left\{Z^t|\, m\right\}^\beta / \sum_{\tilde{m}} \mathbf{P}\left\{Z^t|\, \tilde{m}\right\}^\beta$ in terms of $\Gamma_t(m)e^{\mathbf{n}R}$, i.e:

$$\frac{\mathbf{P}\left\{Z^t|m\right\}^\beta}{\sum_{\tilde{m}} \mathbf{P}\{Z^t|\tilde{m}\}^\beta} \leq \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \qquad \forall Z^t \in \zeta^t, m \in \mathcal{M}, t \leq \tau \tag{33}$$

If $\tau = t$ then $\max_m \mathbf{P}\left\{Z^t|\, m\right\}^\beta / \sum_{\tilde{m}} \mathbf{P}\left\{Z^t|\, \tilde{m}\right\}^\beta \geq \epsilon$, thus

$$\mathbb{I}_{\{\tau=t\}}\mathbb{I}_{\{Z^t\in\zeta^t\}} \leq \mathbb{I}_{\{\max_m \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \tag{34}$$

Note that (34) is an implicit lower bound on t. i.e.

$$\epsilon \leq \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \leq 4e^{-\mathbf{n}R}\left(\max_{x,y} \frac{W(y|x)^\beta}{\sum_{\tilde{x}} W(y|\tilde{x})^\beta P(\tilde{x})}\right)^t$$

Thus if $Z^\tau \in \zeta^\tau$ and $t = \tau$ than $t \geq t_0$ where,

$$t_0 = \left\lceil \frac{\mathbf{n}R + \ln\epsilon - \ln 4}{\max_{x,y} \log \frac{W(y|x)^\beta}{\sum_{\tilde{x}} W(y|\tilde{x})^\beta P(\tilde{x})}} \right\rceil \tag{35}$$

Using equations (14), (23) and (34) we get,

$$P_{\mathbf{et}} \leq \mathbf{E}\left[\mathbb{I}_{\{\max_m \frac{4}{e^{\mathbf{n}R}\Gamma_t(m)} \geq \epsilon\}} \Gamma_t(\theta)^\rho\right] G(\rho,\eta)^{\mathbf{n}-t} e^{\rho\mathbf{n}R} e^\rho$$

We analyze and bound the two case $\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} \geq \epsilon$ and $\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon$ separately.

$$P_{\mathbf{et}} \leq P_{\mathbf{e}ta} + P_{\mathbf{e}tb} \tag{36}$$

$$P_{\mathbf{e}ta} = \mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} \geq \epsilon; \max_m \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \left(\frac{e^{-\mathbf{n}R}}{\Gamma_t(\theta)}\right)^{-\rho}\right] G(\rho,\eta)^{\mathbf{n}-t} e^\rho$$

$$P_{\mathbf{e}tb} = \mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon; \max_m \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \left(\frac{e^{-\mathbf{n}R}}{\Gamma_t(\theta)}\right)^{-\rho}\right] G(\rho,\eta)^{\mathbf{n}-t} e^\rho$$

Note that (36) and the definitions of $P_{\mathbf{e}ta}$ and $P_{\mathbf{e}tb}$ have the slight abuse of notation, like the one in (31). As before encoding scheme is assumed to continue to employ the random coding after τ in the expectations.

Let us start with bounding $P_{\mathbf{e}ta}$. Note that if $4e^{-\mathbf{n}R}/\Gamma_t(\theta) \geq \epsilon$ then the second condition of the indicator function is always satisfied, thus

$$\begin{aligned} P_{\mathbf{e}ta} &= \mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} \geq \epsilon\}} \left(\tfrac{e^{-\mathbf{n}R}}{\Gamma_t(\theta)}\right)^{-\rho}\right] G(\rho,\eta)^{\mathbf{n}-t} e^{\rho} \\ &\leq \mathbf{E}\left[\left(\tfrac{4e^{-\mathbf{n}R}}{\epsilon\Gamma_t(\theta)}\right)^{\lambda_a} \left(\tfrac{e^{-\mathbf{n}R}}{\Gamma_t(\theta)}\right)^{-\rho}\right] G(\rho,\eta)^{\mathbf{n}-t} e^{\rho} \\ &= e^{\rho} \left(\tfrac{4}{\epsilon}\right)^{\lambda_a} e^{\mathbf{n}R(\rho-\lambda_a)} H_a(\rho,\beta,\lambda_a)^t G(\rho,\eta)^{\mathbf{n}-t} \end{aligned} \tag{37}$$

where

$$H_a(\rho,\beta,\lambda_a) = \sum_{y,x} P(x) W(y|x) \left[\tfrac{\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^{\beta}}{W(y|x)^{\beta}}\right]^{\rho-\lambda_a} \tag{38}$$

For bounding $P_{\mathbf{e}tb}$, note that,

$$\begin{aligned} &\mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon; \max_m \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \Gamma_t(\theta)^{\rho}\right] \\ &= \mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon\}} \mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \Gamma_t(\theta)^{\rho}\right] \\ &= \mathbf{E}\left[\mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon\}} \mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \Gamma_t(\theta)^{\rho} \middle| Y^t\right]\right] \end{aligned} \tag{39}$$

Note that given Y^t, $\Gamma_t(m)$'s are independent of each other, consequently

$$\begin{aligned} &\mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon\}} \mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \Gamma_t(\theta)^{\rho} \middle| Y^t\right] \\ &= \mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon\}} \Gamma_t(\theta)^{\rho} \middle| Y^t\right] \mathbf{E}\left[\mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \middle| Y^t\right] \end{aligned} \tag{40}$$

Let us first bound the first term in the product in (40):

$$\mathbf{E}\left[\mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(\theta)} < \epsilon\}} \Gamma_t(\theta)^{\rho} \middle| Y^t\right] \leq \mathbf{E}\left[\left(\tfrac{\epsilon}{4e^{-\mathbf{n}R}}\right)^{\lambda_{b1}} \Gamma_t(\theta)^{\rho+\lambda_{b1}} \middle| Y^t\right] \tag{41}$$

For bounding the second term in (40) note that:

$$\mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} = \min\{1, \textstyle\sum_{m\neq\theta} \mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}}\}$$

Taking the expectation of both sides we get,

$$\begin{aligned} &\mathbf{E}\left[\mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \middle| Y^t\right] \\ &\qquad \leq \min\left\{1, \mathbf{E}\left[\sum_{m\neq\theta} \mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \middle| Y^t\right]\right\} \end{aligned} \tag{42}$$

Note that

$$\mathbf{E}\left[\sum_{m\neq\theta} \mathbb{I}_{\{\frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \geq \epsilon\}} \middle| Y^t\right] \leq \mathbf{E}\left[\sum_{m\neq\theta} \left(\tfrac{4e^{-\mathbf{n}R}}{\epsilon\Gamma_t(m)}\right)^{\lambda_{b2}} \middle| Y^t\right] \tag{43}$$

For any $m \neq \theta$ distribution of the input letter at any time ℓ is $P(x)$ and it is independent of Y_ℓ. Using this one can show that $\mathbf{E}\left[\Gamma_t(m)^{-1} \middle| Y^t\right] = 1$. Thus using this fact together with equations (42) and (43) we can see that the minimum in equation (42) buys us at most a factor of $4\epsilon^{-1}$, when we optimize over λ_{2b}.

Thus for bounding the second term in the product in (40) instead of (42) and (43) we simply use

$$\mathbf{E}\left[\mathbb{I}_{\{\max_{m\neq\theta} \frac{4e^{-\mathbf{n}R}}{\Gamma_t(m)} \leq \epsilon\}} \middle| Y^t\right] \leq \mathbf{E}\left[\sum_{m\neq\theta} \left(\tfrac{4e^{-\mathbf{n}R}}{\epsilon\Gamma_t(m)}\right)^{\lambda_{2b}} \middle| Y^t\right]. \tag{44}$$

Using equations (39), (41) and (44) we get,

$$\begin{aligned} P_{\mathbf{e}tb} &\leq \left(\tfrac{4e}{\epsilon}\right)^{\rho} \mathbf{E}\left[\left(\tfrac{\epsilon\Gamma_t(\theta)}{4e^{-\mathbf{n}R}}\right)^{\lambda_{b1}+\rho} \sum_{m\neq\theta} \left(\tfrac{\epsilon\Gamma_t(m)}{4e^{-\mathbf{n}R}}\right)^{-\lambda_{b2}}\right] G(\rho,\eta)^{\mathbf{n}-t} \\ &\leq \tfrac{e^{\mathbf{n}R(1+\rho+\lambda_{b1}-\lambda_{b2})}}{e^{-\rho}(4\epsilon^{-1})^{\lambda_{b1}-\lambda_{b2}}} H_b(\rho,\eta,\lambda_{b1},\lambda_{b2})^t G(\rho,\eta)^{\mathbf{n}-t} \end{aligned} \tag{45}$$

where

$$\begin{aligned} &H_b(\rho,\eta,\lambda_{b1},\lambda_{b2}) \\ &= \sum_{x,y} P(x) W(y|x) \tfrac{[\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^{\beta}]^{\rho+\lambda_{b1}}}{W(y|x)^{\beta(\rho+\lambda_{b1})}} \tfrac{\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^{\beta\lambda_{b2}}}{[\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^{\beta}]^{\lambda_{b2}}} \end{aligned} \tag{46}$$

C. Parametric Error Bounds:

Using equations (7), (20) and (36) we get

$$P_{\mathbf{e}} \leq 2|\mathcal{X}|\mathbf{n} e^{-\frac{\mathbf{n}^2}{2}} + P_{\mathbf{e}\mathbf{n}}^{*} + \sum_{t=t_0}^{\mathbf{n}} (P_{\mathbf{e}ta} + P_{\mathbf{e}tb}) \tag{47}$$

where the expression for t_0 is given in (35) and bounds for $P_{\mathbf{e}\mathbf{n}}^{*}$, $P_{\mathbf{e}ta}$ and $P_{\mathbf{e}tb}$ are given in parametric form in equation (31), (37) and (45) in terms of $G(\rho,\eta)$, $\epsilon = \mathbf{n}^{-4} \min_{x \in suppP(x)} P(x)$, $\lambda_0 \geq 0$, $\lambda_a \geq 0$, $\lambda_{b1} \geq 0$, $\lambda_{b2} \geq 0$, $\beta \geq 0$, $\rho \geq 0$, $\eta \geq 0$.

In order to recover the extension of the results of [2], we chose $\lambda_0 = 0$, $\lambda_a = \lambda$, $\lambda_{b1} = 0$, $\lambda_{b2} = 1+\lambda$, $\beta = \frac{1}{1+\rho}$. Note that using the definition of $H_0(\rho,\beta,\lambda_0)$ in (32) we get

$$H_0(\rho, \tfrac{1}{1+\rho}, 0) = e^{-E_0(P,\rho)}$$

where

$$E_0(P,\rho) = -\ln \sum_{y} \left[\sum_{x} P(x) W(y|x)^{\frac{1}{1+\rho}}\right]^{1+\rho}$$

Similarly definitions of $H_a(\rho,\beta,\lambda_a)$ and $H_b(\rho,\beta,\lambda_{b1},\lambda_{b2})$ in (38) and (46) leads to,

$$H_a(\rho, \tfrac{1}{1+\rho}, \lambda) = e^{-H(\rho,\lambda)}$$
$$H_b(\rho, \tfrac{1}{1+\rho}, 0, (1+\lambda)) = e^{-H(\rho,\lambda)}$$

where

$$H(\rho,\lambda) = -\ln \sum_{x,y} P(x) W(y|x) \left[\tfrac{\sum_{\tilde{x}} P(\tilde{x}) W(y|\tilde{x})^{\frac{1}{1+\rho}}}{W(y|\tilde{x})^{\frac{1}{1+\rho}}}\right]^{\rho-\lambda}$$

Plugging these and (35),(31), (37), (45) in (47) we can see that $\forall \rho \geq 0$, $\lambda \geq 0$ and $R \in [0, \mathbf{C}]$ the error probability is upper bounded as

$$P_{\mathbf{e}} \leq 2|\mathcal{X}|\mathbf{n} e^{-\frac{\mathbf{n}^2}{2}} + \tfrac{3\mathbf{n}(4\mathbf{n}^4)^{\lambda} e^{\rho}}{(\min_{x \in suppP(x)} P(x))^{\lambda}} e^{-\mathbf{n}F(R,\rho,\lambda)} \tag{48}$$

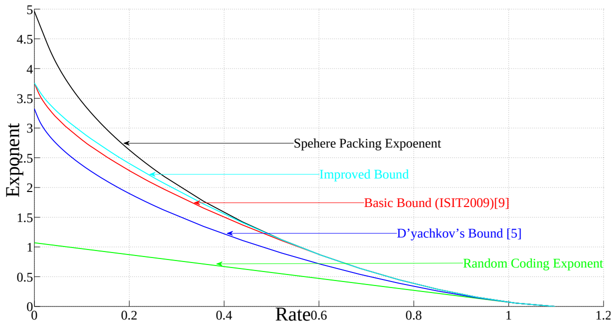

Fig. 1. Spherepacking exponent, exponent resulting from equation (48), exponent we derived in [8], D'yachkov's exponent expression in [5] and random coding exponent are plotted for a ternary symmetric channel with $\delta = 10^{-4}$.

where

$$
\begin{aligned}
F(R,\rho,\lambda) &= \min\{F_1(R,\rho), F_2(R,\rho,\eta,\lambda)\} \\
F_1(R,\rho) &= E_0(\rho) - \rho R \\
F_2(R,\rho,\eta,\lambda) &= \min_{\alpha_0 \le \alpha \le 1} \alpha H(\rho,\lambda) + (1-\alpha) G^*(\rho) - \rho R \\
G^*(\rho) &= -\min_{\eta} \ln G(\rho,\eta) \\
\alpha_0 &= \frac{R}{\max_{x,y} \ln \frac{W(y|x)^\beta}{\sum_{\tilde{x}} W(y|\tilde{x})^\beta P(\tilde{x})}}
\end{aligned}
$$

In order to see the gains of the modification we have discussed over the original scheme we described in [8] we consider k-ary symmetric channels, i.e. channels with $W(y|x)$ of the forms

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

Figure 1 compares the error exponent achievable with the current scheme with the error exponents resulting from previous studies.

VI. Current Work:

We are currently working on the optimal solution of the parametric bounds on the error probability and its implications for general DMCs. Furthermore we are seeking alternative encoding schemes to the random coding for the first phase and a good single letter bound on $G(\rho,\eta)$ function for general DMCs. Finally we are also investigating the possible gains of multi-phase schemes.

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] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math*, 2006.

[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] 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.

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

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

[8] B. Nakiboğlu and L. Zheng. Upper bounds to error probability with feedback. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 1515–1519, 28 2009-July 3 2009.

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

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