Bit-wise Unequal Error Protection for Variable Length Blockcodes with Feedback

Siva K. Gorantla
U.I.U.C.

Barış Nakiboğlu
M.I.T.

Todd P. Coleman
U.I.U.C.

Lizhong Zheng
M.I.T.

***Abstract*—*Bit-wise* unequal error protection problem with two layers is considered for variable length block-codes with feedback. Inner and outer bounds are derived for achievable performance for finite expected decoding time. These bounds completely characterize the error exponent of the special bits as a function of overall rate R, overall error exponent E and the rate of the special bits R_s. Single message *Message-wise* unequal protection problem is also solved as a step on the way.**

I. Introduction

In a recent work [2] an information theoretic frame work has been introduced for unequal error protection (*UEP*) problems. Like conventional *UEP*, error events are grouped into different classes[1] and protection against these different classes of error events are promoted differently, and like the conventional information theory these different levels of protection are measured with exponential decay rates of probabilities of these classes.

Different choices of these classes of error events lead to different problems. The convention introduced in [2] is that, if the class of the error events in consideration can be expressed solely in terms of error events associated with individual messages, then the problem is called *message-wise UEP* problem. For example one can choose these classes to be the missed detections of the messages in disjoint sets like it was done in [5]. Similarly, if the class of the error events in consideration can be expressed solely in terms of error events of different groups of bits then the problem is called *bit-wise UEP* problem. As noted in [2], there are many *UEP* problem of practical importance that are neither *message-wise UEP* nor *bit-wise UEP* problems. Yet studying these special classes seems to be a good starting point. In this manuscript we focus on two closely related *UEP* problems, *bit-wise UEP* problem and *message-wise UEP* problem with single special message.

In *bit-wise UEP* there are special bits which need a better protection than the rest. We determine best error exponent these bits can get, $E_{\text{bits}}(R, E, R_s)$, when overall rate is R, overall error exponent is E and rate of special bits is R_s. In *message-wise UEP* problem with single special message we determine the best error exponent a message can have $E_{\text{md}}(R, E)$ in a rate R code with error exponent E. Our results generalize corresponding results in [2] which were derived for the case when overall rate is assumed to be (very close to) the channel capacity. For many *UEP* problems the channel model and family of codes in consideration makes a big difference. Problems we are considering here are no exception. In this work we focus solely on variable length block-codes on DMCs.

We start by introducing the channel model and variable length block-codes in Section II. Then in Section III we state two *UEP* problems we will consider. We present the achievability results and converses in Sections IV and V respectively. We combine these results to obtain analytical expressions for error exponents $E_{\text{bits}}(R, E, R_s)$ and $E_{\text{md}}(R, E)$ in Section VI.

II. Model and Preliminaries

A. Channel Model

We consider a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probability matrix $W(\cdot|\cdot)$. The input and output letters at time t, up to time t and between time t_1 and t_2 are denoted by X_t, Y_t, X^t, Y^t, $X_{t_1}^{t_2}$ and $Y_{t_1}^{t_2}$ respectively. Consequently

$$\Pr\left[Y_t | \, X^t, Y^{t-1}\right] = W(Y_t|X_t).$$

For the reasons that will become clear shortly we assume that $\min_{x,y} W(y|x) = \lambda > 0$. We denote the output distribution $W(\cdot|x)$ associated with the input letters x also by $W_x(\cdot)$.

B. Variable Length Block-codes

A variable length block-code is a (τ, Φ, Ψ) triple where τ is the decoding time, Φ is the encoding scheme and Ψ is the decoding rule. Decoding time τ is a stopping time with respect to the receivers observation.[2] For each $y^t \in \mathcal{Y}^t$ such that $t < \tau$, encoding scheme $\Phi(\cdot, y^t)$ determines the input letter at time $(t+1)$ for each message in the message set $\mathcal{M}$.

$$\Phi(\cdot, y^t) : \mathcal{M} \to \mathcal{X} \qquad \forall y^t : \tau > t.$$

The decoding rule is a mapping from the set of output sequences at τ to the message set $\mathcal{M}$

$$\Psi(\cdot) : \mathcal{Y}^\tau \to \mathcal{M}.$$

At time zero a message M, chosen uniformly at random from $\mathcal{M}$, is given to the transmitter. The transmitter uses the codeword associated with this message, i.e. $\Phi(M, \cdot)$, to convey it until the decoding time τ is reached. Then the receiver decodes the message $\hat{M} = \Psi(Y^\tau)$. The error probability and the rate of a variable length block-code are given by

$$P_e = \Pr\left[\hat{M} \neq M\right] \qquad \text{and} \qquad R = \tfrac{\ln|\mathcal{M}|}{E[\tau]}.$$

Variable length block-codes on DMCs can also be interpreted as trees, for a more detailed discussion of that interpretation readers may go over [1, Section II].

[1] However unlike the conventional *UEP* these classes are not assumed to be of some particular form right away.

[2] In other words given Y^t receiver knows the value of $\mathbb{1}_{(\tau > t)}$ where $\mathbb{1}_{(\cdot)}$ is the indicator function.

C. Reliable Sequences

A sequence of variable length block-codes $\mathcal{Q}$ is reliable iff $\lim_{n\to\infty} P_e{}^{(n)} = 0$. The rate and the error exponent of a reliable sequence, $\mathcal{Q}$ is defined as,

$$R_{\mathcal{Q}} = \liminf_{n\to\infty} \frac{\ln|\mathcal{M}^{(n)}|}{E\left[\tau^{(n)}\right]} \qquad E_{\mathcal{Q}} = \liminf_{n\to\infty} \frac{-\ln P_e{}^{(n)}}{E\left[\tau^{(n)}\right]}.$$

Furthermore the reliability function of the variable length block-codes is defined as

$$E(R) = \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R} E_{\mathcal{Q}}.$$

Variable length block codes were first considered by Csiszár in [4]. Later Burnashev, [3], studied them on DMCs and proved that: If all entries of $W(\cdot|\cdot)$ are positive then for all $R \leq C$

$$E(R) = \left(1 - \tfrac{R}{C}\right) D.$$

where D is the maximum KL divergence between any two input letters and x_a, x_r are the maximizing input letters:

$$D = \max_{i,j\in\mathcal{X}} \mathsf{D}\left(W_i \| W_j\right) = \mathsf{D}\left(W_{x_a} \| W_{x_r}\right). \tag{1}$$

If there are one or more zero entries in $W(\cdot|\cdot)$ then, $D = \infty$ and for all $R < C$ for large enough $E[\tau]$ there are rate R variable length block-codes which are error free i.e. $P_e = 0$. In this situation all of the messages and bits can have zero error probability simultaneously. This is why we focused on the channels for which all entries of $W(\cdot|\cdot)$ are positive.

III. Problem Statement

In the conventional setting either the average or the maximum of the message error probabilities is studied. However in many situations different kinds of error events have different costs. For example some part of the message, i.e. some bits of the message, can be far more important than the rest. In that situation we need to study the error events associated with these special bits separately. To put it more explicitly, message set $\mathcal{M}$ can be of the form $\mathcal{M} = \mathcal{M}_s \times \mathcal{M}_o$ and the demand on error probability of special bits $\Pr\left[\hat{\mathcal{M}}_s \neq \mathcal{M}_s\right]$ can be far more stringent than the one on the error probability of ordinary bits $\Pr\left[\hat{\mathcal{M}}_o \neq \mathcal{M}_o\right]$.

We characterize the trade-off between these two error events in terms of the trade-off between the exponential decay rates of their probabilities with $E[\tau]$.

Definition 1: For any reliable sequence $\mathcal{Q}$ with message sets $\mathcal{M}^{(n)}$ of the form $\mathcal{M}^{(n)} = \mathcal{M}_s^{(n)} \times \mathcal{M}_o^{(n)}$, the rate and the error exponents of the special bits are defined as,

$$R_{\mathcal{Q},s} \triangleq \liminf_{n\to\infty} \frac{\ln|\mathcal{M}_s^{(n)}|}{E\left[\tau^{(n)}\right]} \qquad E_{\mathcal{Q},s} \triangleq \liminf_{n\to\infty} \frac{-\ln \Pr{}^{(n)}\left[\hat{M}_s \neq M_s\right]}{E\left[\tau^{(n)}\right]}$$

Then for any rate $0 \leq R \leq C$, exponent $0 \leq E \leq (1-\frac{R}{C})D$, special bit rate $0 \leq R_s \leq R$, the special bit error exponent $E_{\text{bits}}(R, E, R_s)$ is defined as

$$E_{\text{bits}}(R, E, R_s) \triangleq \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R,\ E_{\mathcal{Q}} \geq E,\ R_{\mathcal{Q},s} \geq R_s} E_{\mathcal{Q},s} \tag{2}$$

In characterizing the trade-off for *bit-wise UEP* problems, *message wise UEP* problem with single special message plays a key role. In single message *message wise UEP* problem the trade-off between the exponential decay rates of $\min_{m\in\mathcal{M}} \Pr\left[\hat{M} \neq M \middle| M = m\right]$ and $\Pr\left[\hat{M} \neq M\right]$ is studied. Similar to the *bit-wise UEP* problem let us first give the operational definition of $E_{\text{md}}(R, E)$ in terms of reliable sequences.

Definition 2: For any reliable sequence $\mathcal{Q}$ missed detection exponent is defined as,

$$E_{\text{md},\mathcal{Q}} \triangleq \liminf_{n\to\infty} \frac{-\ln \Pr{}^{(n)}\left[\hat{M} \neq 1 \middle| M = 1\right]}{E\left[\tau^{(n)}\right]} \tag{3}$$

Then for any rate $0 \leq R \leq C$, exponent $0 \leq E \leq (1-\frac{R}{C})D$, missed detection exponent $E_{\text{md}}(R, E)$ is defined as

$$E_{\text{md}}(R, E) \triangleq \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R,\ E_{\mathcal{Q}} \geq E} E_{\text{md},\mathcal{Q}}. \tag{4}$$

IV. Inner Bounds: Achievablities

We start by considering a family of fixed length block-codes without feedback and we establish an inner bound to the achievable rate, missed detection exponent pairs. The codes achieving this trade-off have a positive missed detection exponent but their overall error exponent is zero, i.e. as we consider longer and longer codes average error probability decays to zero but subexponentially in the block length. We append a control phase to these codes like the one used by Yamamoto and Itoh in [9] to obtain a positive error exponent. These fixed length block-codes with feedback and erasures are then used as the building block for the variable length block-codes for the *UEP* problems we are interested in. This encoding scheme can be seen as a generalization of an encoding scheme first suggested by Kudrayshov [7]. The key feature of the encoding scheme in [7] is the tacit acceptance and explicit rejection strategy, which was also used in [2]. We combine this strategy with a classic control phase with explicit acknowledgments to get a positive error exponent for all messages/bits. The outer bounds we derive in Section V reveals that such schemes are optimal.

A. An Achievable Scheme without Feedback

Let us first consider a parametric family of codes in terms of two input-letter-input-distribution pairs $(x_1, \mathsf{P}_{X,1})$ and $(x_2, \mathsf{P}_{X,2})$ and a time sharing constant α. Let us denote the output distributions resulting from $\mathsf{P}_{X,k}$ on $W(\cdot|\cdot)$ by $\mathsf{P}_{Y,k}$

$$\mathsf{P}_{Y,k}(j) = \sum_x W(j|x)\mathsf{P}_{X,k}(x) \quad k = 1, 2.$$

Lemma 1: For any block-length $\mathbf{n}$, time sharing constant $0 \leq \alpha \leq 1$, input distribution-input letter pairs $(x_1, \mathsf{P}_{X,1})$ and $(x_2, \mathsf{P}_{X,2})$ there exist a fixed length block-code such that

$$|\mathcal{M}| \geq e^{\mathbf{n}(\alpha \mathsf{I}(\mathsf{P}_{X,1};W) + (1-\alpha)\mathsf{I}(\mathsf{P}_{X,2};W) - \epsilon_1(\mathbf{n}))}$$

$$P_{e,1} \leq e^{-\mathbf{n}\left(\alpha \mathsf{D}\left(\mathsf{P}_{Y,1}\|W_{x_1}\right) + (1-\alpha)\mathsf{D}\left(\mathsf{P}_{Y,2}\|W_{x_2}\right) - \epsilon_2(\mathbf{n})\right)}$$

$$P_{e,m} \leq \epsilon_3(\mathbf{n}) \qquad m = 2, 3, \ldots, |\mathcal{M}|$$

where $\epsilon_i(\mathbf{n}) \geq 0$ and $\lim_{\mathbf{n}\to\infty} \epsilon_i(\mathbf{n}) = 0$ for $i = 1, 2, 3$.

Proof: The codeword of message 1, $x^{\mathbf{n}}(1)$, is concatenation of $\lfloor \mathbf{n}\alpha \rfloor$ x_1's and $(\mathbf{n} - \lfloor \mathbf{n}\alpha \rfloor)$ x_2's. The decoding region of the first message is all the $y^{\mathbf{n}}$'s that are not typical with $(\alpha, \mathsf{P}_{Y,1}, \mathsf{P}_{Y,2})$,

$$\mathcal{G}(1) = \left\{ y^{\mathbf{n}} : \lfloor \mathbf{n}\alpha \rfloor \Delta_{(\mathsf{Q}_{(y_1^{\lfloor \mathbf{n}\alpha \rfloor})}, \mathsf{P}_{Y,1})} + (\mathbf{n} - \lfloor \mathbf{n}\alpha \rfloor) \Delta_{(\mathsf{Q}_{(y^{\mathbf{n}}_{\lfloor \mathbf{n}\alpha \rfloor + 1})}, \mathsf{P}_{Y,2})} \geq \mathbf{n}^{\frac{3}{4}} \right\}$$

where $\mathsf{Q}_{(y_t^{t'})}$ is the empirical distribution of $y_t^{t'}$ and $\Delta_{(\cdot,\cdot)}$ is the total variation between the two distributions. Then

$$P_{e1} \leq e^{-\mathbf{n}(\alpha \mathsf{D}(\mathsf{P}_{Y,1} \| W_{x_1}) + (1-\alpha)\mathsf{D}(\mathsf{P}_{Y,2} \| W_{x_2}) - \epsilon_2(\mathbf{n}))}$$

The codewords of the remaining messages are specified using the classical random coding argument with the empirical typicality. Consider an ensemble of codes in which first $\mathbf{n}_1$ entries of all the codewords are independent and identically distributed (i.i.d.) with input distribution $\mathsf{P}_{X,1}$ and the rest of the entries are i.i.d. with the input distribution $\mathsf{P}_{X,2}$. Decoding region of each message m is the set $y^{\mathbf{n}}$'s for which $(x^{\mathbf{n}}(m), y^{\mathbf{n}})$ is jointly typical with $(\alpha, \mathsf{P}_{X,1}W, \mathsf{P}_{X,2}W)$ and for which no other $(x^{\mathbf{n}}(\tilde{m}), y^{\mathbf{n}})$ is jointly typical with $(\alpha, \mathsf{P}_{X,1}W, \mathsf{P}_{X,2}W)$. Following typicality arguments as usual we get lemma 1. A more detailed discussion is in [6]. ■

Note that given the channel and the target rate $R \leq C$ one can optimize over time-sharing constant α, and the input-letter-input-distribution pairs $(x_1, \mathsf{P}_{X,1})$ and $(x_2, \mathsf{P}_{X,2})$ to obtain the best missed detection exponent achievable for a given rate with the above architecture. In order to characterize this trade off let us define $\mathcal{J}(R)$ as follows

$$\mathcal{J}(R) \triangleq \max_{\substack{\alpha, x_1, x_2, \mathsf{P}_{X,1}, \mathsf{P}_{X,2}: \\ \alpha \mathsf{I}(\mathsf{P}_{X,1}; W) + (1-\alpha)\mathsf{I}(\mathsf{P}_{X,2}; W) \geq R}} \alpha \mathsf{D}(\mathsf{P}_{Y1} \| W_{x_1}) + (1-\alpha)\mathsf{D}(\mathsf{P}_{Y2} \| W_{x_2}) \quad (5)$$

B. Error-and-Erasure Decoding

The codes described in Lemma 1 have large missed detection exponent, but their overall error exponent is zero. We append them with a control phase and allow erasures to give them a positive error exponent, like it was done in [9].

Lemma 2: For any block length $\mathbf{n}$, rate $0 < R < C$ and error exponent $E < (1 - \frac{R}{C})D$, there exists a block-code with erasure probability $P_{\mathbf{x}}$ such that,

$$\begin{aligned} |\mathcal{M}| &\geq e^{\mathbf{n}(R - \epsilon_1(\mathbf{n}))} \\ P_{e,1} &\leq e^{-\mathbf{n}(E + (1-E/D)\mathcal{J}(\frac{R}{1-E/D}) + \epsilon_2(\mathbf{n}))} \\ P_{e,m} &\leq \epsilon_4(\mathbf{n}) \min\{1, e^{-\mathbf{n}(E - \epsilon_2(\mathbf{n}))}\} \qquad m = 2, 3, \ldots, |\mathcal{M}| \\ P_{\mathbf{x},m} &\leq \epsilon_5(\mathbf{n}) \end{aligned}$$

where $\epsilon_i(\mathbf{n}) \geq 0$ and $\lim_{\mathbf{n} \to \infty} \epsilon_i(\mathbf{n}) = 0$ for $i = 1, 2, 4, 5$.

Proof: We use a two phase block-code to achieve this performance. In the first phase a length $\mathbf{n}_1 = \mathbf{n} - \lfloor \frac{E}{D}\mathbf{n} \rfloor$, rate $\frac{\mathbf{n}}{\mathbf{n}_1}R$ code with high missed detection exponent is used to convey the message, like the one described in Lemma 1. At the end of this phase a tentative decision is made. If the tentative decision is correct the accept letter x_a is sent for remaining $\lfloor \frac{E}{D}\mathbf{n} \rfloor$ time units; if the tentative decision is wrong the reject letter letter x_r is sent. Letters x_a and x_r are the ones described in equation (1). At the end of the second phase an erasure is declared if the output sequence in the second phase is not typical with W_{x_a}, if it is typical tentative decision becomes the final. A more detailed discussion is in [6]. ■

C. Single Special Message:

We get a variable length block-code by using above fixed length block-code with erasures repetitively until a non-erasure decoding occurs. Expected decoding time and average error probability of this new code and the error probability of its special message are given by

$$E[\tau] = \frac{\mathbf{n}}{1 - P_{\mathbf{x}}} \qquad \tilde{P_e}' = \frac{P_e}{1 - P_{\mathbf{x}}} \qquad \tilde{P}_{e1}' = \frac{P_{e1}}{1 - P_{\mathbf{x}1}} \qquad (6)$$

D. Special Bits:

Lemma 3: For any block length $\mathbf{n}$, rate $0 < R < C$ and error exponent $E \leq (1 - \frac{R}{C})D$, special bit rate $R_s \leq R$, there exists a block-code with erasure probability $P_{\mathbf{x}}$ such that,

$$\begin{aligned} |\mathcal{M}_s| &\geq e^{\mathbf{n}(R_s - \epsilon_1(\mathbf{n}))} \\ |\mathcal{M}_o| &\geq e^{\mathbf{n}(R - R_s - \epsilon_1(\mathbf{n}))} \\ P_{e,m}^{\ s} &\leq e^{-\mathbf{n}(E + (1 - \frac{R_s}{C} - \frac{E}{D})\mathcal{J}(\frac{R - R_s}{1 - R_s/C - E/D}) + \epsilon_2(\mathbf{n}))} && m \in \mathcal{M}_s \\ P_{e,m}^{\ o} &\leq \epsilon_6(\mathbf{n}) \min\{1, e^{-\mathbf{n}(E - \epsilon_2(\mathbf{n}))}\} && m \in \mathcal{M}_o \\ P_{\mathbf{x},m} &\leq \epsilon_7(\mathbf{n}) && m \in \mathcal{M} \end{aligned}$$

where $\epsilon_i(\mathbf{n}) \geq 0$ and $\lim_{\mathbf{n} \to \infty} \epsilon_i(\mathbf{n}) = 0$ for $i = 1, 2, 6, 7$.

Proof: Consider a three phase block-code. In the first phase a length $\mathbf{n}_1 = \frac{R}{C}\mathbf{n}$, rate C code is used to convey M_s, i.e. $M_I = M_s$. At the end of the first phase a tentative decision, $\tilde{M}_I$ is made. In the second phase, a code with a special message, like the one described in Lemma 1, is used. If $\tilde{M}_I \neq M_I$ the special message is sent, i.e. $M_{II} = 1$. If $\tilde{M}_I = M_I$ then M_o is sent, i.e. $M_{II} = M_o + 1$. The code in the second phase is of length $\mathbf{n}_2 = \mathbf{n}(1 - \frac{R_s}{C} - \frac{E}{D})$ and of rate $\frac{\mathbf{n}}{\mathbf{n}_2}(R - R_s)$. At the end of the second phase a tentative decision, $\tilde{M}_{II}$ is made about M_{II}. If the tentative decision $\tilde{M}_{II}$ is correct the accept letter x_a is sent through out the third phase; if the tentative decision $\tilde{M}_{II}$ is wrong the reject letter letter x_r is sent. At the end of the third phase an erasure is declared if

- Y sequence in the third phase is not typical with W_{x_a} or
- $\tilde{M}_{II} = 1$

Otherwise, tentative decisions become the final, i.e. $\hat{M}_s = \tilde{M}_I$ and $\hat{M}_o = \tilde{M}_{II} - 1$. A more detailed discussion is in [6]. ■

Using a similar repetition argument we can turn this errors-and-erasures code to a variable length code such that,

$$E[\tau] = \frac{\mathbf{n}}{1 - P_{\mathbf{x}}} \qquad \tilde{P_e}' = \frac{P_e}{1 - P_{\mathbf{x}}} \qquad \tilde{P_e}^{s\prime} = \frac{P_e^{\ s}}{1 - P_{\mathbf{x}}} \qquad (7)$$

We show in Section V that this lower bound is tight and the encoding scheme leading to it is optimal. But even before that one can show that within the three phase encoding scheme like the one considered in Lemma 3 particular choice of relative duration of phases are optimal.

V. Outer Bounds: Converses

In this section we first prove a rather technical lemma for variable length block-codes. This Lemma bounds the expected value of a particular conditional error probability from below in terms of the rate of decrease of the conditional entropy of the messages in different intervals. Lower bounds for both of the *UEP* problems in consideration follows from this lemma.

A. Missed Detection Probability and Decay Rate of Entropy:

Like similar lemmas in [1] and [2], Lemma 4 bounds the error probability for a question that is asked at a stopping time. But unlike them Lemma 4 works with the missed detection probability instead of average error probability. Putting it more explicitly Lemma 4 bounds the expected value of $\Pr\left[\hat{M} \notin \mathcal{A}_i \middle| Y^{\tau_i}, M \in \mathcal{A}_i\right]$ for variable length block-codes with feedback, where $\mathcal{A}_i$ is a subset of the message set that is uniquely determined by Y^{τ_i}.

Lemma 4: For any variable length block-code with decoding time τ, and k stopping times such that $0 \leq \tau_1 \leq \tau_2 \leq \cdots \tau_k \leq \tau$ with the associated functions $(\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k)$ of the form[3] $\mathcal{A}_i(\cdot) : \mathcal{Y}^{\tau_i} \to \mathcal{P}(\mathcal{M})$ we have

$$\ln \frac{1}{E\left[\Pr\left[\hat{M} \notin \mathcal{A}_i \middle| Y^{\tau_i}, M \in \mathcal{A}_i\right]\right]} \leq \frac{\ln 2 + \sum_{j=i}^{k} E[\tau_{j+1}-\tau_j] \mathcal{J}\left(\frac{E\left[\mathcal{H}(M|Y^{\tau_j}) - \mathcal{H}(M|Y^{\tau_{j+1}})\right]}{E[\tau_{j+1}-\tau_j]}\right)}{1 - P_e - \Pr[M \in \mathcal{A}_i]} \tag{8}$$

Bound on $E\left[\Pr\left[\hat{M} \notin \mathcal{A}_i \middle| Y^{\tau_i}, M \in \mathcal{A}_i\right]\right]$ depends only on $\Pr[M \in \mathcal{A}_i]$ and the rate of decrease of conditional entropy in the intervals $(\tau_j, \tau_{j+1}]$ for $j \geq i$. Particular choice of $\mathcal{A}_j$ for $j \neq i$ has no effect on the bound. This property of the bound is its main merit over bounds resulting from the previously suggested techniques.

Proof: We prove the lemma for $i = 1$, clearly this implies the bound $\forall i \in \{1, 2, \ldots, k\}$. Let $\mathcal{G}(\mathcal{A}_1)$ be the decoding region for the messages in $\mathcal{A}_1$, i.e $\mathcal{G}(\mathcal{A}_1) = \{y^\tau : \hat{M}(y^\tau) \in \mathcal{A}_1\}$. Using data processing inequality for KL-divergence and the fact that $h(x) = x \ln \frac{1}{x} + (1-x) \ln \frac{1}{1-x} \leq \ln 2 \; \forall x \in [0,1]$ we get

$$E\left[\ln \frac{\Pr[Y^\tau | Y^{\tau_1}]}{\Pr[Y^\tau | Y^{\tau_1}, M \in \mathcal{A}_1]}\right] \geq \ln \tfrac{1}{2} + \{1 - \Pr[\mathcal{G}(\mathcal{A}_1)]\} \ln \frac{1}{E\left[\Pr\left[\hat{M} \notin \mathcal{A}_1 \middle| Y^{\tau_1}, M \in \mathcal{A}_1\right]\right]}.$$

Note that $\Pr[\mathcal{G}(\mathcal{A}_1)] \leq P_e + \Pr[M \in \mathcal{A}_1]$ thus we have

$$\frac{E\left[\ln \frac{2\Pr[Y^\tau | Y^{\tau_1}]}{\Pr[Y^\tau | Y^{\tau_1}, M \in \mathcal{A}_1]}\right]}{1 - P_e - \Pr[M \in \mathcal{A}_1]} \geq \ln \frac{1}{E\left[\Pr\left[\hat{M} \notin \mathcal{A}_1 \middle| Y^{\tau_1}, M \in \mathcal{A}_1\right]\right]} \tag{9}$$

Now we bound $E\left[\ln \frac{\Pr[Y^\tau | Y^{\tau_1}]}{\Pr[Y^\tau | Y^{\tau_1}, M \in \mathcal{A}_1]}\right]$ from above. For that let us consider the stochastic sequence

$$S_t = \left[\sum_{j=\tau_1+1}^{t} \mathcal{J}\left(\mathcal{I}\left(M; Y_j \middle| Y^{j-1}\right)\right) - \ln \frac{\Pr[Y^t | Y^{\tau_1}]}{\Pr[Y^t | Y^{\tau_1}, M \in \mathcal{A}_1]}\right] \mathbb{1}_{\{t > \tau_1\}} \tag{10}$$

where $\mathcal{I}\left(M; Y_t \middle| Y^{t-1}\right) = E\left[\ln \frac{\Pr[Y_t | M, Y^{t-1}]}{\Pr[Y_t | Y^{t-1}]} \middle| Y^{t-1}\right]$.

Note that given Y^{t-1} random variables $M - X_t - Y_t$ form a Markov chain, thus $\mathcal{I}(M; Y_{t+1} | Y^t) \leq \mathcal{I}(X_t; Y_{t+1} | Y^t)$. Consequently using the fact that $\mathcal{J}(\cdot)$ is a decreasing function and the definition of $\mathcal{J}(\cdot)$ given in (5) we get $E[S_{t+1} | Y^t] \geq S_t$. Since $\min W_{i,j} = \lambda$ and $|\mathcal{J}(\cdot)| \leq D$ we have $E[|S_{t+1} - S_t| | Y^t] \leq \ln \frac{1}{\lambda} + D$ thus S_t is a sub-martingale. Furthermore for stopping times $E[\tau_1] \leq E[\tau_2] < \infty$, we can use [8, Theorem 2, p487] to get $E[S_{\tau_2}] \geq E[S_{\tau_1}] = 0$, i.e.

$$E\left[\ln \frac{\Pr[Y^{\tau_2} | Y^{\tau_1}]}{\Pr[Y^{\tau_2} | Y^{\tau_1}, M \in \mathcal{A}_1]}\right] \leq E\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{J}\left(\mathcal{I}\left(M; Y_t \middle| Y^{t-1}\right)\right)\right] \tag{11}$$

Using the concavity of $\mathcal{J}(\cdot)$ and Jensen's inequality we get

$$\frac{E\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{J}\left(\mathcal{I}\left(M; Y_t \middle| Y^{t-1}\right)\right)\right]}{E[\tau_2 - \tau_1]} \leq \mathcal{J}\left(\frac{E\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{I}\left(M; Y_t \middle| Y^{t-1}\right)\right]}{E[\tau_2 - \tau_1]}\right) \tag{12}$$

In order to find the sum within $\mathcal{J}(\cdot)$ in equation (12) consider the stochastic sequence,

$$V_t = \mathcal{H}(M | Y^t) + \sum_{i=1}^{t} \mathcal{I}\left(M; Y_i \middle| Y^{i-1}\right) \tag{13}$$

Note that $E[V_{t+1} | Y^t] = V_t$ and $E[|V_t|] < \infty$, thus V_t is a martingale. Furthermore, $E[|V_{t+1} - V_t| | Y^t] < \infty$ and $E[\tau_1] \leq E[\tau_2] < \infty$, thus using Doob's optimal stopping theorem, [8, Theorem 2, p487], we get $E[V_{\tau_1}] = E[V_{\tau_2}]$ i.e.,

$$E\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{I}\left(M; Y_t \middle| Y^{t-1}\right)\right] = \mathcal{H}(M | Y^{\tau_1}) - \mathcal{H}(M | Y^{\tau_2}) \tag{14}$$

Using equations (11), (12) and (14)

$$E\left[\ln \frac{\Pr[Y^{\tau_2} | Y^{\tau_1}]}{\Pr[Y^{\tau_2} | Y^{\tau_1}, M \in \mathcal{A}_1]}\right] \leq E[\tau_2 - \tau_1] \, \mathcal{J}\left(\frac{\mathcal{H}(M | Y^{\tau_1}) - \mathcal{H}(M | Y^{\tau_2})}{E[\tau_2 - \tau_1]}\right)$$

Repeating same arguments for the intervals $[\tau_i + 1, \tau_{i+1}]$ and using equation (9) we get equation (8) for τ_1. ∎

B. Single Special Message:

Lemma 5: For any variable length block-code with rate $0 < R < C$, error exponent $0 < E < (1 - R/C)D$, decoding time τ and $0 < \delta < 0.5$

$$-\frac{\ln \Pr\left[\hat{M} \neq i \middle| M = i\right]}{E[\tau]} \leq E + \left(1 - \frac{E - \tilde{\epsilon}}{D}\right) \mathcal{J}\left(\frac{R - \tilde{\epsilon}}{1 - (E - \tilde{\epsilon})/D}\right) \tag{15}$$

where $\tilde{\epsilon} = \frac{\tilde{\epsilon}_1 D + \tilde{\epsilon}_2}{1 - \tilde{\epsilon}_1}$, $\tilde{\epsilon}_1 = P_e + \delta + \frac{P_e}{\delta} + \frac{1}{|\mathcal{M}|}$ and $\tilde{\epsilon}_2 = \frac{\ln 2 - \ln \lambda\delta}{E[\tau]}$.
Lemma 5 is a generalization of [2, Theorem 8] to the case where $E > 0$, however unlike [2, Theorem 8] proof presented below does not use the previous results like [1, Lemma 1].

Proof: Let $\tau_1 = 0$, $\mathcal{A}_1 = i$. Let τ_2 and $\mathcal{A}_2$ be

$$\tau_2 = \inf\{t : \max_m \Pr\left[M = m \middle| Y^t\right] \geq 1 - \delta \text{ or } t = \tau\} \tag{16a}$$

$$\mathcal{A}_2 = \{m \in \mathcal{M} : \Pr[M = m | Y^{\tau_2}] < (1 - \delta)\}. \tag{16b}$$

Note that $\Pr[M \in \mathcal{A}_2 | Y^{\tau_2}] \geq \lambda\delta$. Thus

$$E\left[\Pr\left[\hat{M} \notin \mathcal{A}_2 \middle| Y^{\tau_2}, M \in \mathcal{A}_2\right]\right] \leq \frac{P_e}{\lambda\delta}. \tag{17}$$

If $\Pr\left[\hat{M} \neq M \middle| \mathcal{A}_2 = \mathcal{M}\right] \geq \delta$. Thus using Markov inequality for P_e we get

$$\Pr[\mathcal{A}_2 = \mathcal{M}] \leq \frac{P_e}{\delta}. \tag{18}$$

[3] Recall that for any set $\mathcal{M}$ the power set $\mathcal{P}(\mathcal{M})$ is the set of all possible subsets of $\mathcal{M}$.

Using the fact that $\Pr[M \in \mathcal{A}_2 | \mathcal{A}_2 \neq \mathcal{M}] \leq \delta$ and (18)

$$\Pr[M \in \mathcal{A}_2] \leq \delta + \tfrac{P_e}{\delta}. \tag{19}$$

Note that $\mathcal{A}_2$ has at least $(|\mathcal{M}|-1)$ messages. Thus Fano's inequality together with equation (18) implies

$$\mathcal{H}(M|Y^{\tau_2}) \leq -\ln 2 + (\delta + \tfrac{P_e}{\delta}) \ln |\mathcal{M}|. \tag{20}$$

Using equations (17-20), Lemma 4 and $\mathcal{J}(R) \leq D$ we get

$$\frac{1}{E[\tau]} \ln \frac{1}{\Pr\left[\hat{M} \neq i \middle| M=i\right]} \leq \frac{\eta \mathcal{J}\left(\frac{R(1-\tilde{\epsilon}_1)-\tilde{\epsilon}_2}{\eta}\right) + (1-\eta)D + \tilde{\epsilon}_2}{1-\tilde{\epsilon}_1} \tag{21a}$$

$$\frac{1}{E[\tau]} \ln \frac{1}{P_e} \leq \frac{(1-\eta)D+\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1} \tag{21b}$$

where $\eta = \frac{E[\tau_2]}{E[\tau]}$. $\mathcal{J}(R)$ is a concave function thus it lies below its tangents. Thus for any $0 \leq R \leq C$ and $\eta \leq \frac{R}{C}$

$$\frac{d}{d\eta}\left[\eta \mathcal{J}\left(\frac{R}{\eta}\right) + (1-\eta)D\right] = \mathcal{J}\left(\frac{R}{\eta}\right) - \frac{R}{\eta}\mathcal{J}'\left(\frac{R}{\eta}\right) - D \geq 0$$

Thus the bound in equation (21a) has its maximum value for the maximum value of η. Furthermore equation (21b) gives an upper bound on η. These two observations leads to Lemma 5. ∎

C. Special Bits:

Lemma 6: For any variable length block-code with rate $0 < R < C$, error exponent $0 < E < (1 - R/C)D$, special bit rate $0 < R_s < R$, decoding time τ and $0 < \delta < 0.5$

$$\frac{-\ln \Pr\left[\hat{M}_s \neq M_s\right]}{E[\tau]} \leq E + (1 - \tfrac{R_s}{C} - \tfrac{E}{D} - \tilde{\epsilon})\mathcal{J}\left(\frac{R-R_s}{1-\frac{R_s}{C}-\frac{E}{D}-\tilde{\epsilon}}\right) \tag{22}$$

where $\tilde{\epsilon} = \frac{\tilde{\epsilon}_4}{1-\tilde{\epsilon}_3}\frac{C+D}{CD}$, $\tilde{\epsilon}_3 = P_e + \delta + \frac{P_e}{\delta}$ and $\tilde{\epsilon}_4 = \frac{\ln 2 - \ln \lambda\delta}{E[\tau]}$.

Proof: Let τ_1 and $\mathcal{A}_1$ be

$$\tau_1 = \inf\{t : \max_{m_s} \Pr\left[M_s = m_s | Y^t\right] \geq 1 - \delta \text{ or } t = \tau\} \tag{23a}$$

$$\mathcal{A}_1 = \{(m_s, m_o) \in \mathcal{M} : \Pr[M_s = m_s | Y^{\tau_1}] < 1 - \delta\}. \tag{23b}$$

and τ_2 and $\mathcal{A}_2$ be same as Lemma 5, i.e. equation (16). Using an analysis very similar to the one for τ_2 and $\mathcal{A}_2$ we can get the equivalents of equations (17), (19), (20) for τ_1 and $\mathcal{A}_1$

$$\frac{\Pr\left[\hat{M}_s \neq M_s\right]}{\lambda\delta} \geq E\left[\Pr\left[\hat{M} \notin \mathcal{A}_1 \middle| Y^{\tau_1}, M \in \mathcal{A}_1\right]\right] \tag{24a}$$

$$\Pr[M \in \mathcal{A}_1] \leq \delta + \tfrac{P_e}{\delta} \tag{24b}$$

$$\mathcal{H}(M|Y^{\tau_1}) \leq -\ln 2 + \ln \tfrac{|\mathcal{M}|}{|\mathcal{M}_1|} + (\delta + \tfrac{P_e}{\delta}) \ln |\mathcal{M}_1| \tag{24c}$$

Using equation (14) and $\mathcal{I}\left(M; Y^t \middle| y^{t-1}\right) \leq C$ we get

$$C E[\tau_1] \geq E[\mathcal{H}(M|Y^{\tau_0}) - \mathcal{H}(M|Y^{\tau_1})] \tag{25}$$

Let us introduce the short hand:

$$\eta_1 = \tfrac{E[\tau_1]}{E[\tau]}, \quad \eta_2 = \tfrac{E[\tau_2 - \tau_1]}{E[\tau]}, \quad f_1 = \tfrac{E[\mathcal{H}(M|Y^{\tau_1})]}{E[\tau]}, \quad f_2 = \tfrac{E[\mathcal{H}(M|Y^{\tau_2})]}{E[\tau]}$$

Using lemma 4, equations (17), (19), (20), (24), (25) and $\mathcal{J}(\cdot) \leq D$ implies that if a $\Pr\left[\hat{M}_s \neq M\right]$ is achievable for a (R, E, R_s) triple then equation (26) is satisfied for some $(\eta_1, \eta_2, f_1, f_2)$

$$\frac{-\Pr\left[\hat{M}_s \neq M_s\right]}{E[\tau]} \leq \frac{\tilde{\epsilon}_4}{1-\tilde{\epsilon}_3} + \frac{\eta_2}{1-\tilde{\epsilon}_3}\mathcal{J}\left(\frac{f_1 - f_2}{\eta_2}\right) + \frac{(1-\eta_1-\eta_2)D}{1-\tilde{\epsilon}_3} \tag{26a}$$

$$(1-\tilde{\epsilon}_3)E \leq \tilde{\epsilon}_4 + (1 - \eta_1 - \eta_2)D \tag{26b}$$

$$f_1 \leq \tilde{\epsilon}_4 + R - (1-\tilde{\epsilon}_3)R_s \tag{26c}$$

$$f_2 \leq \tilde{\epsilon}_4 + \tilde{\epsilon}_3 R \tag{26d}$$

$$f_1 \geq R - C\eta_1 \tag{26e}$$

one can show that largest value of the inequality constraint in equation (26a) happens when equations (26b), (26c), (26d), (26e) are satisfied with equality. Using this fact we obtain equation (22). ∎

VI. Results and Conclusions

Using Lemma 5 for $\delta = \frac{-1}{\ln P_e}$ together with equation (6) and Lemma 2 we get

Theorem 1: For any rate $0 < R \leq C$ and error exponent $0 \leq E \leq (1 - R/C)D$ missed detection exponent is

$$E_{\text{md}}(R, E) = E + (1 - E/D)\,\mathcal{J}\left(\frac{R}{1-E/D}\right). \tag{27}$$

Using lemma 6 for $\delta = \frac{-1}{\ln P_e}$ together with equation (7) and lemma 3 we get

Theorem 2: For any rate $0 \leq R \leq C$, error exponent $0 \leq E \leq (1 - \frac{R}{C})D$ and special bit rate $0 \leq R_s \leq R$ the special bit error exponent $E_{\text{bits}}(R, E, R_s)$ is

$$E_{\text{bits}}(R, E, R_s) = E + \left(1 - \tfrac{R_s}{C} - \tfrac{E}{D}\right)\mathcal{J}\left(\frac{R - R_s}{1 - \frac{R_s}{C} - \frac{E}{D}}\right) \tag{28}$$

Using the very same machinery, we have characterized the achievable region for a k layer bit-wise *UEP* problem, [6]. One important observation is that unlike the case when total rate is capacity, multi-layer systems do not necessarily have a successive cancellation property. We have derived the necessary conditions for having successive cancellation property. Arguably the most important contribution of this work is the new technique for establishing outer bounds, Lemma 4.

References

[1] P. Berlin, B. Nakiboğlu, B. Rimoldi, and E. Telatar. A simple converse of burnashev's reliability function. *Information Theory, IEEE Transactions on*, 55(7):3074–3080, July 2009.

[2] S. Borade, B. Nakiboğlu, and L. Zheng. Unequal error protection: An information-theoretic perspective. *Information Theory, IEEE Transactions on*, 55(12):5511–5539, Dec. 2009.

[3] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problemy Perdachi Informatsii*, 12(4):10–30, 1976.

[4] I. Csiszár. On the capacity of noisy channels with arbitrary signal costs. *Problems of Control and Information Theory*, 2:283–304, 1973.

[5] I. Csiszár. Joint source-channel error exponent. *Problems of Control and Information Theory*, Vol. 9, Iss.5:315–328, 1980.

[6] S. K. Gorantla, B. Nakiboğlu, T. P. Coleman, and L. Zheng. Multi layer bit-wise unequal error protection for variable length blockcodes with feedback.

[7] B. D. Kudryashov. On message transmission over a discrete channel with noiseless feedback. *Problemy Perdachi Informatsii*, 15(1):3–13, 1973.

[8] Albert N. Shiriaev. *Probability*. Springer-Verlag Inc., New York, NY, USA, 1996.

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