

Some Fundamental Limits of Unequal Error Protection

Shashi Borade Barış Nakiboğlu Lizhong Zheng

EECS, MIT, Cambridge, MA 02139. Email: { spb, nakib, lizhong } @mit.edu

Abstract—Various formulations are considered where some information is more important than other and needs better protection. Our information theoretic framework in terms of exponential error bounds provides some fundamental limits and optimal strategies for such problems of unequal error protection. Even for data-rates approaching the channel capacity, it shows how a crucial part of information can be protected with exponential reliability. Channels without feedback are analyzed first, which is useful later in analyzing channels with feedback. A new channel parameter, called the *Red-Alert Exponent*, is fundamentally important in such problems.

I. INTRODUCTION

Classical theoretical framework for communication assumes that all information is equally important. In this framework, the communication system aims to provide a uniform error protection to all messages: any particular message being mistaken as any other is viewed to be equally costly. With such uniformity assumptions, the reliability of a communication scheme is measured by either the average or the worst case probability of error, over all possible messages to be transmitted. In information theory literature, a communication scheme is said to be *reliable* if this error probability can be made small. Communication schemes designed with this framework turn out to be optimal in sending any source over any channel, provided that long enough codes can be employed. This homogeneous view of information motivates the universal interface of “bits” between any source and any channel [1], which is often viewed as Shannon’s most significant contribution.

In many communication scenarios, such as wireless networks, interactive systems, and control applications, where sufficient error protection becomes a luxury, providing such a uniform protection for all the information may be either a wasteful or an infeasible approach. Instead, it is more efficient here to protect a (crucial) part of information better than the rest. For example in a wireless network, control signals including channel state, power control, and scheduling information are often more important than the payload data, and should be protected more carefully. Similarly for the Internet, packet headers are more important and need better protection. Another example is when a multiple resolution source code is transmitted over a wireless channel, it makes sense to protect the coarse resolution better so the user can at least have some crude reconstruction after bad channel realizations. For such situations of heterogeneous information, unequal error protection (UEP) is a natural generalization of the conventional content-blind information processing.

The simplest method of unequal error protection is to allocate different channels for different types of data. For example, wireless systems allocate a separate “control channel”, often with short codes and low spectral efficiency, to transmit control signals with high reliability. The well known Gray code, assigning similar bit strings to close by constellation points, can be viewed as UEP: even if there is some error in identifying the transmitted symbol, there is a good chance that some of the bits are correctly received. More systematic

designs for UEP can be found in [7] and references therein. For erasure channels, this problem is known as “priority encoded transmission” (PET) [6]. For wireless channels, [8] analyzes this problem in terms of diversity-multiplexing tradeoffs. Most of these approaches focus on designing good codes for specific channel models. The optimality of these designs was established in only limited cases. This paper aims to provide a general information theoretic framework for understanding fundamental limits in UEP.

A general formulation of the unequal error protection problem requires some different definitions of decoding error than the commonly used ones. Consider a channel encoder which takes the input of l information bits, $\mathbf{b} = [b_1, b_2, \dots, b_l]$, which is equivalent to a random variable M taking values from the set $\{1, 2, 3, \dots, 2^l\}$. Each message in this set corresponds to a particular value of the bit-sequence \mathbf{b} . This set of possible values of M are referred to as “messages”. After a message is encoded and transmitted over the channel, a decoding error is defined as the event that the receiver decodes to a different message than the transmitted one. In most information theory texts, when a decoding error occurs, the entire bit sequence \mathbf{b} is rejected. That is, errors in decoding the message and in decoding the information bits are treated similarly.

In the existing formulations of unequal error protection codes, the information bits are divided into subsets, and the decoding errors in different subsets of bits are viewed as different kinds of errors. For example, one might want to provide a better protection to one subset of bits by ensuring that errors in these bits are less probable than the other bits. We call such problems “bit-wise UEP”. Previous examples of packet headers, multiple resolution codes, etc. belong to this category of UEP.

However, in some situations, instead of *bits* one might want to provide a better protection to a subset of *messages* by ensuring a lower probability of error when one of these special messages is transmitted/decoded. For example, one might consider embedding a special message in a normal l -bit code, i.e., transmitting one of $2^l + 1$ messages, where the extra message has a special meaning and requires a smaller error probability. We call such problems as “message-wise UEP”.

For example, this special message could indicate some system emergency, which is too costly to be missed. Borrowing from hypothesis testing, we call the conditional error probability of a special message as its *missed-detection* probability, which is the probability of missing to detect the special message when it was transmitted. Note that the decoding error conditioned on the special message is not associated to error in any particular bit. Instead, it corresponds to a particular bit-sequence (corresponding to the special message) being decoded as some other bit-sequence.

Alternatively, the special messages could demand small *false-alarm* probability: the event when the receiver erroneously chooses that message although some other message was sent. For example consider the reboot message for a

remote-controlled system such as a robot or satellite. Its false-alarm could cause unnecessary shutdowns and other system troubles. For brevity however, we will not discuss false-alarms anymore and only focus on avoiding missed-detections in message-wise UEP.

In conventional data communication, there is no need to distinguish between bit-errors and message-errors, as all information is “created equal”, its meaning (and importance) is separated from the engineering problem of communication [1]. In the UEP problems however, bits and messages are different as some are labeled as “special” or “high priority”. Now it becomes necessary to differentiate the two notions of special information: special bits and special messages.

The main contribution of this paper is a set of results, identifying the performance limits and optimal coding strategies, for a variety of UEP scenarios. We will focus on a few simplified notions of UEP, most with immediate practical applications, and try to illustrate the main insights for them. One can imagine using these UEP strategies for embedding protocol information within the actual data. By eliminating a separate control channel, this can enhance the overall bandwidth and/or energy efficiency.

For conceptual clarity, this article focuses on situations where the data-rate is essentially equal to the channel capacity¹. By this analysis, we will be addressing UEP issues for scenarios where data rate is a crucial system resource that can not be compromised. In these cases, no positive error exponent in the conventional sense can be achieved. That is, if we aim to protect the entire information uniformly well, neither bit-wise nor message-wise error probabilities can decay exponentially fast with increasing block length. We ask the question then “can we make the error probability of a particular bit, or a particular message, decay exponentially fast with block length?”

The question of fundamental limits of UEP was clearly of interest in previous works on code designs for UEP. To the best of our knowledge, however, there was no general characterization of these limits in terms of error exponents; partially due to the difficulty in proving converses. In this paper and [5], we develop such converses as well as optimal strategies. More importantly, the notion of message-wise UEP was essentially never addressed in the past (except in the paper Joint Source-Channel Error Exponent by Csiszár, [9]).

When we break away from the conventional framework and start to provide better protection to selected parts of information, these parts of information need not be only bits. A general formulation of UEP could be an arbitrary combination of protection demands from messages, where each message demands better protection against some specific kinds of errors. In this general definition of UEP, bit-wise UEP and message wise UEP are simply two particular ways of specifying which kinds of errors are too costly compared to others.

In the following, Section II discusses bit-wise UEP and message-wise UEP for the no-feedback case. Theorem 1 shows that for data-rates approaching capacity, even a single bit cannot achieve any positive error exponent. Thus in bit-wise

UEP, the data-rate must back-off from capacity for achieving any error exponent even for a single bit. On the contrary, in message-wise UEP, positive error exponents can be achieved even at capacity. If only one message in a capacity achieving code was special and demanded an error exponent, Theorem 2 shows its optimal value is equal to a new fundamental channel parameter called the Red-Alert Exponent. We then consider situations where an exponentially large subset of messages is special and demands a positive error exponent. Theorem 3 shows a surprising result that these special messages can achieve the same exponent as if all the other (ordinary) messages were absent. In other words, a capacity achieving code and an error exponent-optimal code below capacity can coexist without affecting each other. These results shed some new light on the structure of capacity achieving codes.

Insights from the no-feedback case become useful in Section III for the case with full feedback, which shows that full feedback creates some fundamental connections between bit-wise UEP and message-wise UEP. Now even for bit-wise UEP, positive error exponent can be achieved at capacity. In fact, Theorem 4 shows that a single special bit can achieve the same exponent as a single special message, which equals the Red-Alert Exponent. For a single special message however, Theorem 5 shows that feedback does not improve the achievable exponent. The case of exponentially many messages is resolved in Theorem 6. Of course, many special messages cannot achieve a better exponent compared to a single special message. We will see that the special messages can achieve the same error exponent with feedback as if all other messages were absent. Lastly, some future directions are discussed in Section IV.

II. ERROR EXPONENTS AT CAPACITY: NO-FEEDBACK CASE

Consider a discrete memoryless channel W from input X to output Y and let \mathcal{X}, \mathcal{Y} denote their alphabets, respectively. The output distribution conditioned on input $i \in \mathcal{X}$ is denoted by $W_{Y|X}(\cdot|i)$ and the channel capacity is denoted by C . We assume that all entries of the channel transition matrix are non-zero, *i.e.*, every output is reachable from every input. Let us first review the classical definition of an error exponent when all information is treated equally [2],[3].

Definition 1: A (n, R, ϵ_n) code denotes a length n code of rate R , which has e^{nR} messages and the average error probability of the overall code equals ϵ_n .

$$\Pr(\hat{M} \neq M) = \frac{1}{e^{nR}} \sum_k \sum_{j \neq k} \Pr(\hat{M} = j | M = k) = \epsilon_n$$

where M and \hat{M} are in $\{1, 2, \dots, e^{nR}\}$ denote the randomly chosen transmitted message and the decoded message, respectively.

At a given rate R , a sequence of (n, R, ϵ_n) codes with increasing block length n is said to achieve an error exponent if its error probability ϵ_n can decay exponentially with n .

Definition 2: The error exponent $E(R)$ at rate R is the maximum value of E such that a sequence of (n, R, ϵ_n) codes exists for which ϵ_n satisfies $\epsilon_n \doteq e^{-nE}$.

We use \doteq as a shorthand notation for

$$E = \lim_{n \rightarrow \infty} \frac{-\log \epsilon_n}{n} \quad (1)$$

¹In another write-up [5], we will analyze similar problems in a more general framework to allow data-rates below capacity.

Reliable communication at capacity means that for arbitrarily small gap to capacity $C - R \triangleq \xi > 0$, a sequence of (n, R, ϵ_n) codes exists for which ϵ_n vanishes for large n . However, ϵ_n cannot decay exponentially in this case. That is, no positive exponent is achievable for this error probability as the gap ξ to capacity vanishes [2].

$$\xi \rightarrow 0 \quad \Rightarrow \quad E(R) = E(C - \xi) \rightarrow 0$$

A. Special bit

We first address the situation where one particular (say the first) information bit out of the total $nR/\log 2$ information bits is a special bit—it needs a superior error protection compared to other bits. If this first bit is denoted as b_0 and its decoded value is denoted by \hat{b}_0 . We require that the error probability for b_0 , $\Pr(\hat{b}_0 \neq b_0)$, to decay exponentially while ensuring reliable communication at capacity for the remaining bits. Let us define its exponent.

Definition 3: Let $E_b(\xi)$ be the largest value such that a sequence of $(n, C - \xi, \epsilon_n)$ codes exists for which ϵ_n vanishes for large n and $\Pr(\hat{b}_0 \neq b_0) \doteq e^{-nE_b(\xi)}$. We define E_b as the infimum of $E_b(\xi)$ over all positive ξ .

$$E_b = \inf_{\xi > 0} E_b(\xi)$$

This is equivalent to this simpler version used later.

Definition 4: E_b is the largest number such that a sequence of $(n, C - \xi, \epsilon_n)$ codes exists for arbitrarily small $\xi > 0$, for which ϵ_n vanishes, and $\Pr(\hat{b}_0 \neq b_0) \doteq e^{-nE_b}$.

As noted earlier, the overall information cannot achieve any positive error exponent $E(R)$ near capacity. However, it is not clear whether a single special bit can steal an error exponent E_b near capacity.

Theorem 1: $E_b = 0$

Intuitive Interpretation: Let the shaded balls in Fig. II-A denote the minimal decoding regions of the $\doteq e^{nC}$ messages. These decoding regions to ensure reliable communication, they essentially denote the typical noise-balls around codewords. The decoding regions on the left of the thick line corresponds to $\hat{b}_1 = 1$ and those on the right correspond to the same when $\hat{b}_1 = 0$. Each of these halves includes half of the decoding regions.

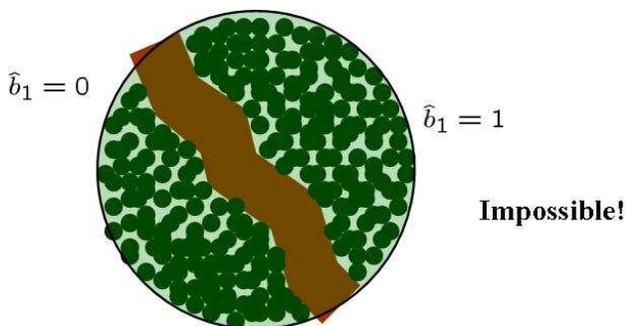


Fig. 1. Impossible: splitting the output space into 2 distant enough clusters.

For achieving a positive error exponent for the special bit, the codewords in the two halves should be sufficiently separated from each other as seen in Fig. II-A. Such separation is

necessary to ensure exponentially small probability of landing in the wrong half. However, above theorem indicates that such a thick patch takes too much volume, and is impossible when we have to fill $\doteq e^{nC}$ typical noise balls in the output space.

B. Special message

Now we focus on situations where one particular message (say $M = 1$) out of the total e^{nR} messages is a special message—it needs a superior error protection. The missed detection (i.e., conditional error) probability $\Pr(\hat{M} \neq 1 | M = 1)$ for this ‘emergency’ message needs to be minimized.

Definition 5: E_m is the largest number such that a sequence of $(n, C - \xi, \epsilon_n)$ codes exists for arbitrary $\xi > 0$, for which ϵ_n vanishes, and $\Pr(\hat{M} \neq 1 | M = 1) \doteq e^{-nE_m}$.

Theorem 2: $E_m = \max_{i \in \mathcal{X}} D(P_Y^*(\cdot) \| W_{Y|X}(\cdot|i)) \triangleq D_{\text{Red}}$, where P_Y^* denotes the capacity achieving output distribution.

Compare this with the corresponding result for classical communication near capacity. If all the messages demand equally small missed detection probability, then no positive error exponent is achievable for them near capacity. This follows from the previous discussion of the classical error exponent $E(R)$. The above theorem shows the improvement in this exponent if we only demand it for a single message instead of all.

Definition 6: Parameter D_{Red} of a channel is defined as its *Red-Alert Exponent*.

$$D_{\text{Red}} = \max_{i \in \mathcal{X}} D(P_Y^*(\cdot) \| W_{Y|X}(\cdot|i)) \quad (2)$$

The input letter achieving this maximum is denoted by x_r .

Notice the relation between D_{Red} and C : the arguments to KL divergence are flipped. It is because Karush-Kuhn-Tucker conditions for achieving capacity imply the following [2],

$$C = \max_{i \in \mathcal{X}} D(W_{Y|X}(\cdot|i) \| P_Y^*(\cdot)).$$

Capacity C represents the best possible data-rate over a channel, whereas Red-Alert Exponent D_{Red} represents the best possible protection of a message for data-rates near capacity.

It is worth mentioning here the ‘‘very noisy’’ channel in [2]. In this formulation [4], the KL divergence is symmetric, which implies $D(P_Y^*(\cdot) \| W_{Y|X}(\cdot|i)) \approx D(W_{Y|X}(\cdot|i) \| P_Y^*(\cdot))$. Hence the Red-Alert Exponent and capacity become essentially equal.

Optimal strategy: The special codeword is a repetition sequence of input x_r . Its decoding region S contains every output sequence with empirical distribution (output type) different than the capacity achieving P_Y^* . For the ordinary codewords, use a capacity achieving code and apply ML decoding over them for output sequences outside S .

For a symmetric channel like BSC, any input letter can be used as x_r . Since the P_Y^* is the uniform distribution (denoted by U_Y) for these channels, $D_{\text{Red}} = D(U_Y(\cdot) \| W_{Y|X}(\cdot|i))$ for any input i . This is the sphere-packing exponent $E_{\text{sp}}(0)$ of this channel at rate 0.

Intuitive Interpretation: Having a large missed detection exponent for the special message corresponds to having a large decoding region $\mathcal{G}(1)$ for the special message. This ensures

that when the special message is transmitted, probability of landing outside $\mathcal{G}(1)$ is exponentially small. In a sense, E_{md} indicates how large $\mathcal{G}(1)$ could be made, while still filling $\doteq e^{nC}$ typical noise balls in the remaining space. The red region in Fig. 2 denotes such a large region. Note that the actual decoding region $\mathcal{G}(1)$ is much larger than this illustration, because it consists of all output types except P_Y^* , whereas the ordinary decoding regions only contain the output type P_Y^* .

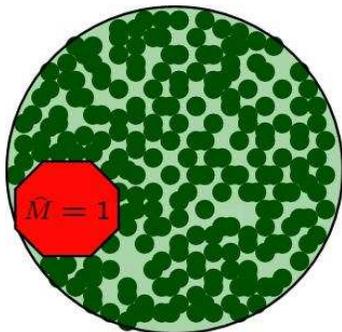


Fig. 2. Avoiding missed-detection

C. Many special messages

Now consider that instead of a single special message, exponentially many of the $e^{n(C-\xi)}$ total messages are special. Let these special messages be the first e^{nr} messages. Define $E_M(r)$ as their missed detection exponent.

Definition 7: For a fixed $r < C$, define $E_M(r)$ as the largest number such that a sequence of $(n, C - \xi, \epsilon_n)$ codes exists for arbitrarily small $\xi > 0$, for which ϵ_n vanishes, and

$$\Pr(\hat{M} \neq k | M = k) \doteq e^{-nE_M(r)}, \quad \forall k \in \{1, 2, \dots, e^{nr}\}$$

If there were only e^{nr} messages in the code (instead of $e^{n(C-\xi)}$), their best missed detection exponent equals $E(r)$. This is the classical exponent defined in Eq. (1) earlier.

Theorem 3: $E_M(r) = E(r) \quad \forall r \in [0, C]$.

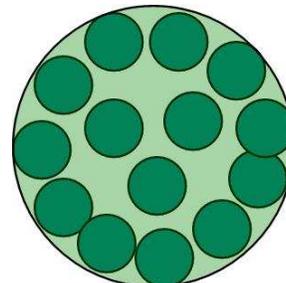
Thus whatever $E(r)$ is achievable for only e^{nr} messages, is also achievable when there are $e^{n(C-\xi)} - e^{nr} \approx e^{nC}$ extra ordinary messages requiring reliable communication.

Optimal strategy: Start with an optimal code-book for e^{nr} messages which achieves error exponent $E(r)$. These codewords are used for the special messages. Now the ordinary codewords are added using random coding. The ordinary codewords which land close to a special codeword may be discarded without essentially any effect on the rate of communication. At the decoder, a two-stage decoding rule is employed. The first stage decides that some special codeword was sent if at least one of the special codewords is within a threshold distance from the received sequence. Otherwise, the first stage decides that an ordinary codeword was sent. Depending on the first stage decision, the second stage ignores all codewords of one kind and applies ML decoding to the rest.

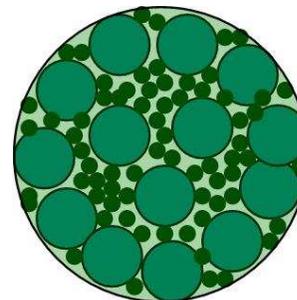
Intuitive Interpretation: This means that we can start with a code of e^{nr} messages, where the decoding regions are large enough to provide a missed detection exponent of $E(r)$. Consider the balls around each codeword with sphere-packing

radius (see Fig. 3(a)). For each message, the probability of going outside its ball decays exponentially with the sphere-packing exponent.

Although, these e^{nr} balls fill up most of the output space, there are still some cavities left between them. These small cavities can still accommodate $\doteq e^{nC}$ typical noise balls for the ordinary messages (see Fig. 3(b)), which are much smaller than the original e^{nr} balls. This is analogous to filling sand particles in a box full of large boulders. This theorem is like saying that the number of sand particles remains unaffected (exponentially) in spite of the large boulders.



(a) Exponent optimal code



(b) Achieving capacity

Fig. 3. “There is always room for capacity.”

III. EFFECTS OF FULL FEEDBACK

Now we revisit the previous problems assuming perfect feedback at the transmitter: it knows all the past outputs before sending a new input symbol. Feedback allows us to use variable time decoding schemes. Similar to Burnashev [10], we focus on block encoding schemes where transmission of a new message begins only after decoding of the old message is finished. Since the decoding time n could be a random variable now, let \bar{n} denote its average.

Definition 8: A $(\bar{n}, R, \epsilon_{\bar{n}})$ feedback code denotes an encoding strategy which has $e^{\bar{n}R}$ messages and error probability $\epsilon_{\bar{n}}$, where \bar{n} equals the average decoding delay assuming uniformly distributed messages.

$$\bar{n} \triangleq \frac{1}{e^{\bar{n}R}} \sum_{k=1}^{e^{\bar{n}R}} E[n|M=k]$$

where $E[n|M=k]$ is average decoding time for message k .

A. Special bit

First consider the situation where the first bit b_0 out of the $\bar{n}R/\log 2$ bits is special. The error exponent for the special bit at capacity is defined as follows.

Definition 9: E_b^f is the largest number such that a sequence of $(\bar{n}, R, \epsilon_{\bar{n}})$ feedback codes exists for arbitrary $\xi > 0$, for which $\epsilon_{\bar{n}}$ vanishes, and $\Pr(\hat{b}_0 = b_0) \doteq e^{-\bar{n}E_b^f}$.

Theorem 4: $E_b^f = D_{\text{Red}}$.

Recall that without feedback, the single bit could not achieve a positive error exponent near capacity. The following strategy shows how feedback connects message-wise UEP with bit-wise UEP: strategy for protecting a special message becomes useful for protecting special bits. This special message is for indicating incorrect decisions at the receiver.

Optimal strategy: We achieve this exponent using the missed detection exponent of D_{Red} for a special message. This special message aims to notify the receiver when \hat{b}_0 is incorrect. More specifically, transmitter first conveys b_0 using a short repetition code of length $\sqrt{\bar{n}}$. If \hat{b}_0 is correct after this repetition code, the remaining bits are conveyed using a capacity achieving code of length $\bar{n} - \sqrt{\bar{n}}$. If \hat{b}_0 is incorrect after the repetition code, transmitter sends a ‘buzzer’ codeword of length $\bar{n} - \sqrt{\bar{n}}$. For this buzzer, we use the same codeword that achieved the missed detection exponent $E_m = D_{\text{Red}}$, which is a repetition of the input symbol x_r .

An erasure is declared (only) if the decoder detects the buzzer in the last $\bar{n} - \sqrt{\bar{n}}$ symbols. Then the encoder retransmits by repeating the same strategy afresh. The erasure probability is vanishingly small, which ensures the effective rate of communication approaches capacity in spite of such retransmissions. Decoded \hat{b}_0 is wrong if buzzer is not detected, which happens with the missed-detection exponent D_{Red} .

Remark: Similar scheme is useful when $\bar{n}r/\log 2$ bits (instead of one) are special and achieves $(1 - r/C)D_{\text{Red}}$ as their exponent.

B. Special message

Now one particular message (say $M = 1$) requires small missed-detection probability. Similar to the no-feedback case, define E_m^f as the missed-detection exponent near capacity which implies $\Pr(\hat{M} \neq 1 | M = 1) \doteq \exp(-\bar{n}E_m^f)$.

Theorem 5: Feedback does not improve the missed detection exponent for a single special message: $E_m^f = E_m = D_{\text{Red}}$.

If Red-Alert Exponent was defined as the best protection of a special message (for data-rates near capacity), then this result could be thought of as an analog the ‘‘feedback does not increase capacity’’ for Red-Alert Exponent. Also note that with feedback, E_m^f for the special message and E_b^f for the special bit became equal.

C. Many special messages

Now let us reconsider the problem where the first $e^{\bar{n}r}$ messages are special. We will now require that average decoding delay $\mathbb{E}[n | M = k]$ to be equal across all messages—special and ordinary—and hence equals \bar{n} . This uniformity constraint reflects a system requirement for ensuring a robust delay performance, which is invariant of the transmitted message. As in the no-feedback case, define $E_M^f(r)$ capacity such that $\Pr(\hat{M} \neq k | M = k) \doteq e^{-\bar{n}E_M^f(r)}$ for every special message k .

Theorem 6: Let $D_{\text{max}} \equiv \max_{i,j} D(W_{Y|X}(\cdot|i) || W_{Y|X}(\cdot|j))$,

$$E_M^f(r) = \min\{D_{\text{Red}}, (1 - r/C)D_{\text{max}}\}, \quad \forall r < C.$$

Thus $E_M^f(r)$ is the minimum of D_{Red} and the Burnashev exponent at rate r . For r at which $D_{\text{Red}} \leq (1 - r/C)D_{\text{max}}$, all $e^{\bar{n}r}$ special messages achieve the best missed detection exponent D_{Red} for a single special message. For larger r where $D_{\text{Red}} > (1 - r/C)D_{\text{max}}$, the special messages achieve the Burnashev exponent as if the ordinary messages were absent.

The optimal strategy is based on transmitting a special bit first. It again shows how feedback connects bit-wise UEP with message-wise UEP: now however the strategy for protecting a special bit is used for protecting special messages (which is the exact opposite of the scheme for achieving E_b^f)

Optimal strategy: We combine the strategy for achieving D_{Red} for a special bit and the Yamamoto-Itoh strategy for achieving Burnashev exponent [11]. In the first phase, an indicator bit b_0 is sent with a repetition code of $\sqrt{\bar{n}}$ symbols. This is an indicator bit for special messages: it is 1 when a special message is to be sent and 0 otherwise. If it is decoded incorrectly as $\hat{b}_0 = 0$, then a missed detection buzzer is sent for the remaining $\bar{n} - \sqrt{\bar{n}}$ symbols. If it is decoded correctly as $\hat{b}_0 = 0$, then the ordinary codeword is sent using a capacity achieving code.

If it is decoded as $\hat{b}_0 = 1$, then the particular special message is sent using the Yamamoto-Itoh scheme: transmit it at capacity using $\approx \frac{\bar{n}r}{C}$ symbols and confirm the decoded \hat{M} in the remaining $\approx \bar{n}(1 - \frac{r}{C})$ symbols.

IV. FUTURE DIRECTIONS

This framework provides a large set of fundamental problems to be studied. For example, many fundamental limits for UEP at rates below capacity need to be understood. Effects of allowing erasures and list decoding also need to be studied. Designing efficient codes for achieving these tradeoffs is another open area. Information networks (e.g., two-way channels, broadcast and relay channels) provide another rich dimension of problems in information theory and optimization.

ACKNOWLEDGMENT

Authors are indebted to Bob Gallager for his insights and encouragement for this work in general. In particular, Theorem 3 was mainly inspired from his remarks. Helpful comments from David Forney are also gratefully acknowledged.

REFERENCES

- [1] C. E. Shannon, ‘‘A mathematical theory of communication,’’ *Bell System Technical Journal*, vol. 27, pp. 379-423 and 623-656, July and Oct., 1948.
- [2] R. Gallager, *Info. Theory and Reliable Communication*, Wiley, 1968.
- [3] D. Forney, ‘‘On exponential error bounds for random codes on the BSC,’’ unpublished manuscript.
- [4] S. Borade and L. Zheng, ‘‘Euclidean information theory,’’ Allerton Conference, Monticello, Sept. 2007.
- [5] S. Borade, B. Nakiboglu, L. Zheng, ‘‘Fundamental limits of UEP: data-rates below capacity,’’ in preparation.
- [6] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, Priority encoding transmission, *IEEE Trans. Inform. Theory*, vol. 42, pp. 1737-1744, Nov. 1996.
- [7] A. R. Calderbank, N. Seshadri Multilevel codes for unequal error protection, *IEEE Trans. Info. Theory*, vol. 39, pp. 1234-1248, July 1993.
- [8] S. Diggavi, D. Tse, ‘‘On successive refinement of diversity,’’ Allerton Conference, Illinois, September 2004.
- [9] I. Csiszar, ‘‘Joint source-channel error exponent,’’ *Problems of Control and Information Theory*, vol. 9 (5), pp. 315-328, 1980.
- [10] M. Burnashev, ‘‘Data transmission over a discrete channel with feedback, random trans. time’’, *Problems Info. Trans.*, vol. 12, pp. 10-30, 1976.
- [11] H. Yamamoto and K. Itoh, Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback, *IEEE Trans. Info. Theory*, vol. 25, pp. 729733, November 1979.