

MIT Open Access Articles

Unequal Error Protection: An Information-Theoretic Perspective

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation	Borade, S., B. Nakiboglu, and Lizhong Zheng. "Unequal Error Protection: An Information-Theoretic Perspective." Information Theory, IEEE Transactions on 55.12 (2009): 5511-5539. © 2009 Institute of Electrical and Electronics Engineers
As Published	http://dx.doi.org/10.1109/TIT.2009.2032819
Publisher	Institute of Electrical and Electronics Engineers
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/52346
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

DSpace@MIT

Unequal Error Protection: An Information-Theoretic Perspective

Shashi Borade, *Student Member, IEEE*, Barış Nakiboğlu, and Lizhong Zheng, *Member, IEEE*

***Abstract*—An information-theoretic framework for unequal error protection is developed in terms of the exponential error bounds. The fundamental difference between the *bit-wise* and *message-wise* unequal error protection (*UEP*) is demonstrated, for fixed-length block codes on discrete memoryless channels (DMCs) without feedback. Effect of feedback is investigated via variable-length block codes. It is shown that, feedback results in a significant improvement in both *bit-wise* and *message-wise UEPs* (except the single message case for missed detection). The distinction between false-alarm and missed-detection formalizations for *message-wise UEP* is also considered. All results presented are at rates close to capacity.**

***Index Terms*—Block codes, blowing-up lemma, error exponents, false alarm, feedback, missed detection, unequal error protection (UEP), variable-length block coding.**

I. Introduction

CLASSICAL theoretical framework for communication [35] 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, 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 [35], and is often viewed as Shannon's most significant contribution.

In many communication scenarios, such as wireless networks, interactive systems, and control applications, where uniformly good error protection becomes a luxury; providing such a protection to the entire information might be wasteful, if not infeasible. Instead, it is more efficient here to protect a crucial part of information better than the rest; see, for example, the following.

- In a wireless network, control signals like channel state, power control, and scheduling information are often more important than the payload data, and should be protected more carefully. Thus, even though the final objective is delivering the payload data, the physical layer should provide a better protection to such protocol information. Similarly for the Internet, packet headers are more important for delivering the packet and need better protection to ensure that the actual data gets through.
- Another example is transmission of a multiple resolution source code. The coarse resolution needs a better protection than the fine resolution so that the user at least obtains some crude reconstruction after bad noise realizations.
- Controlling unstable plants over noisy communication link [33] and compressing unstable sources [34] provide more examples where different parts of information need different reliability.

In contrast with the classical homogeneous view, these examples demonstrate the heterogeneous nature of information. Furthermore, the practical need for unequal error protection (UEP) due to this heterogeneity demonstrated in these examples is the reason why we need to go beyond the conventional content-blind information processing.

Consider a message set $\mathcal{M} = \{1, 2, 3, \ldots, 2^k\}$ for a block code. Note that members of this set, i.e., "messages," can also be represented by length k strings of information bits, $\boldsymbol{b} = [b_1, b_2, \ldots b_k]$. A block code is composed of an encoder which maps the messages $M \in \mathcal{M}$ into channel inputs and a decoder which maps channel outputs to decoded message $\hat{M} \in \mathcal{M}$. An error event for a block code is $\{\hat{M} \neq M\}$. In most information theory texts, when an error occurs, the entire bit sequence $\boldsymbol{b}$ is rejected. That is, errors in decoding the message and in decoding the information bits are treated similarly. We avoid this, and try to figure out what can be achieved by analyzing the errors of different subsets of bits separately.

In the existing formulations of unequal error protection codes [38] in coding theory, the information bits are partitioned 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 as "bit-wise UEP." Previous examples of packet headers, multiple resolution codes, etc. , belong to this category of UEP.

Manuscript received March 17, 2008; revised June 11, 2009. Current version published November 20, 2009. This work was supported by DARPA ITMANET project and by AFOSR under Grant FA9550-06-0156. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, (ISIT) Toronto, ON, Canada, July 2008.

S. Borade was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139 USA (e-mail: spb@mit.edu; spb@alum.mit.edu).

B. Nakiboğlu and L. Zheng are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge MA 02139 USA (e-mail: nakib@mit.edu; lizhong@mit.edu).

Communicated by A. Grant, Associate Editor for Communications.

Color versions of Figures 1–5 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2032819

0018-9448/$26.00 © 2009 IEEE

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

However, in some situations, instead of *bits* one might want to provide a better protection to a subset of *messages*. For example, one might consider embedding a special message in a normal k-bit code, i.e., transmitting one of $2^k + 1$ messages, where the extra message has a special meaning and requires a smaller error probability. Note that the error event for the special message is not associated to the error in any particular bit or set of bits. Instead, it corresponds to a particular bit-sequence (i.e., message) being decoded as some other bit sequence. Borrowing from hypothesis testing, we can define two kinds of errors corresponding to a special message.

- *Missed detection* of a message i occurs when transmitted message M is i and decoded message $\hat{M}$ is some other message $j \neq i$. Consider a special message indicating some system emergency which is too costly to be missed. Clearly, such special messages demand a small missed detection probability. Missed detection probability of a message is simply the conditional error probability after its transmission.
- *False alarm* of a message i occurs when transmitted message M is some other message $j \neq i$ and decoded message $\hat{M}$ is i. Consider the reboot message for a remote-controlled system such as a robot or a satellite or the "disconnect" message to a cell phone. Its false alarm could cause unnecessary shutdowns and other system troubles. Such special messages demand small false-alarm probability.

We call such problems as "message-wise UEP." In conventional framework, every bit is as important as every other bit and every message is as important as every other message. In short, in a conventional framework it is assumed that all the information is "created equal." In such a framework there is no reason to distinguish between bit-wise or message-wise error probabilities because message-wise error probability is larger than bit-wise error probability by an insignificant factor, in terms of exponents. However, in the UEP setting, it is necessary to differentiate between message errors and bit errors. We will see that in many situations, error probability of special bits and messages have behaved very differently.

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 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 paper focuses exclusively on situations where the data rate is essentially equal to the channel capacity. These situation can be motivated by the scenarios where data rate is a crucial system resource that cannot be compromised. In these situations, 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 code 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?"

When we break away from the conventional framework and start to provide better protection against certain kinds of errors, there is no reason to restrict ourselves by assuming that those errors are erroneous decoding of some particular *bits* or missed detections or false alarms associated with some particular messages. A general formulation of UEP could be an arbitrary combination of protection demands 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, we start by specifying the channel model and giving some basic definitions in Section II. Then, in Section III, we discuss bit-wise UEP and message-wise UEP for block codes without feedback. 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 positive error exponent even for a single bit. On the contrary, in message-wise UEP, positive error exponents can be achieved even at capacity. We first consider the case when there is only one special message and show, in Theorem 2, that the optimal (missed-detection) error exponent for the special message is equal to the *red-alert exponent*, which is defined in Section III-B. We then consider situations where an exponentially large number of messages are special and each special message demands a positive (missed detection) error exponent. (This situation has previously been analyzed in [12], and a result closely related to our has been reported there.) Theorem 3 shows a surprising result that these special messages can achieve the same exponent as if all the other (nonspecial) messages are absent. In other words, a capacity-achieving code and an error exponent-optimal code below capacity can coexist without hurting each other. These results also shed some new light on the structure of capacity-achieving codes.

Insights from the block codes without feedback become useful in Section IV, where we investigate similar problems for variable-length block codes with feedback. Feedback together with variable decoding time creates some fundamental connections between bit-wise UEP and message-wise UEP. Now even for bit-wise UEP, a positive error exponent can be achieved at capacity. Theorem 5 shows that a single special bit can achieve the same exponent as a single special message—the red-alert exponent. As the number of special bits increases, the achievable exponent for them decays linearly with their rate as shown in Theorem 6. Then, Theorem 7 generalizes this result to the case when there are multiple levels of specialty—most special, second-most special, and so on. It uses a strategy similar to onion-peeling and achieves error exponents which are successively refinable over multiple layers. For single special message case, however, Theorem 8 shows that feedback does not improve the optimal missed detection exponent. The case of exponentially many messages is resolved in Theorem 9. Evidently, many special messages cannot achieve an exponent higher than that of a single special message, i.e., red-alert exponent. However, it turns out that the special messages can reach red-alert exponent at rates below a certain threshold, as if all the other special messages were absent. Furthermore, for the rates above the very same threshold, special messages reach

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

the corresponding value of Burnashev's exponent, as if all the ordinary messages were absent.

Section V then addresses message-wise UEP situations, where special messages demand small probability of false alarms instead of missed detections. It considers the case of fixed-length block codes without feedback as well as variable-length block codes with feedback. This discussion for false alarms was postponed from earlier sections to avoid confusion with the missed-detection results in earlier sections. Some future directions are discussed briefly in Section VI.

After discussing each theorem, we will provide a brief description of the optimal strategy, but refrain from detailed technical discussions. Proofs can be found in later sections. In Sections VII and VIII we will present the proofs of the results in Section III, on block codes without feedback, and Section IV, on variable-length block codes with feedback, respectively. Finally, in Section IX we discuss the proofs for the false-alarm results of Section V. Before going into the presentation of our work let us give a very brief overview of earlier work on the problem, in different fields.

A. Earlier Work and Contribution

The simplest method of unequal error protection is to allocate different channels for different types of data. For example, many wireless systems allocate a separate "control channel," often with short codes with low rate 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. But clearly this approach is far from addressing the problem in any effective way.

The first systematic consideration of problems in coding theory was within the frame work of linear codes. In [24], Masnick and Wolf suggested techniques which protect different parts (bits) of the message against different number of channel errors (channel symbol conversions). This frame work has been extensively studied over the years in [7], [8], [16], [21], [22], [26], [27], and in many others. Later issue is addressed within the framework of low-density parity-check (LDPC) codes as well [29]–[32], [39], and [28].

"Priority encoded transmission" (PET) was suggested by Albenese *et al.* [2] as an alternative model of the problem, with packet erasures. In this approach, guarantees are given not in terms of channel errors but packet erasures. Coding and modulation issues are addressed simultaneously in [10]. For wireless channels, [15] analyzes this problem in terms of diversity–multiplexing tradeoffs.

In contrast with the above mentioned work, we pose and address the problem within the information-theoretic frame work. We work with the error probabilities and refrain from making assumptions about the particular block code used while proving our converse results. This is the main difference between our approach and the prevailing approach within the coding theory community.

In [3], Bassalygo *et al.* considered the error-correcting codes whose messages are composed of two group of bits, each of which required a different level of protection against channel errors and provided inner and outer bounds to the achievable performance, in terms of Hamming distances and rates. Unlike other works within coding theory frame work, they do not make any assumption about the code. Thus, their results can indeed be reinterpreted in our framework as a result for bit-wise UEP, on binary-symmetric channels.

Some of the the UEP problems have already been investigated within the framework of information theory as well. Csiszár studied message-wise UEP with many messages in [12]. Moreover, results in [12] are not restricted to the rates close to capacity, like ours. Also messages-wise UEP with a single special message was dealt with in [23] by Kudryashov. In [23], a UEP code with single special message is used as a subcode within a variable-delay communication scheme. The scheme proposed in [23] for the single special message case is a key building block in many of the results in Section IV. However, the optimality of the scheme was not proved in [23]. We show that it is indeed optimal.

The main contribution of the current work is the proposed framework for UEP problems within information theory. In addition to the particular results presented on different problems and the contrasts demonstrated between different scenarios, we believe the proof techniques used in [1] Sections VII-A, VIII-B.2, and VIII-D.2 are novel and promising for future work in the field.

II. Channel Model and Notation

A. DMCs and Block Codes

We consider a discrete memoryless channel (DMC) $W_{Y|X}$, with input alphabet $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$ and output alphabet $\mathcal{Y} = \{1, 2, \ldots, |\mathcal{Y}|\}$. The conditional distribution of output letter Y when the channel input letter X equals $i \in \mathcal{X}$ is denoted by $W_{Y|X}(\cdot|i)$

$$\Pr[Y = j | X = i] = W_{Y|X}(j|i), \quad \forall i \in \mathcal{X}, \forall j \in \mathcal{Y}.$$

We assume that all the entries of the channel transition matrix are positive, that is, every output letter is reachable from every input letter. This assumption is indeed a crucial one. Many of the results we present in this paper change when there are zero-probability transitions.

A length n block code without feedback with message set $\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$ is composed of two mappings, encoder mapping and decoder mapping. Encoder mapping assigns a length n codeword[2]

$$\bar{x}^n(k) \triangleq (\bar{x}_1(k), \bar{x}_2(k) \ldots, \bar{x}_n(k)), \quad \forall k \in \mathcal{M}$$

where $\bar{x}_t(k)$ denotes the input at time t for message k. Decoder mapping, $\hat{M}$, assigns a message to each possible channel output sequence, i.e., $\hat{M} : \mathcal{Y}^n \to \mathcal{M}$.

[1]The key idea in Section VIII-B.2 is a generalization of the approach presented in [4].

[2]Unless mentioned otherwise, lower case letters (e.g., x) denote a particular value of the corresponding random variable denoted in capital letters (e.g., X).

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

At time zero, the transmitter is given the message M, which is chosen from $\mathcal{M}$ according to a uniform distribution. In the following n time units, it sends the corresponding codeword. After observing Y^n, the receiver decodes a message. The error probability P_e and rate R of the code is given by

$$P_e \triangleq \Pr\left[\hat{M} \neq M\right] \quad \text{and} \quad R \triangleq \frac{\ln|\mathcal{M}|}{n}.$$

B. Different Kinds of Errors

While discussing message-wise UEP, we consider the conditional error probability for a particular message $i \in \mathcal{M}$

$$\Pr\left[\hat{M} \neq i \middle| M = i\right].$$

Recall that this is the same as the missed detection probability for message i.

On the other hand, when we are talking about bit-wise UEP, we consider message sets that are of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$. In such cases, message M is composed of two submessages: $M = (M_1, M_2)$. The first submessage M_1 corresponds to the high-priority bits while the second submessage M_2 corresponds to the low-priority bits. The uniform choice of M from $\mathcal{M}$ implies the uniform and independent choice of M_1 and M_2 from $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. Error probability of a submessage M_j is given by

$$\Pr\left[\hat{M}_j \neq M_j\right], \quad j = 1, 2.$$

Note that the overall message M is decoded incorrectly when either M_1 or M_2 or both are decoded incorrectly. The goal of bit-wise UEP is to achieve best possible $\Pr\left[\hat{M}_1 \neq M_1\right]$ while ensuring a reasonably small $P_e = \Pr\left[\hat{M} \neq M\right]$.

C. Reliable Code Sequences

We focus on systems where reliable communication is achieved in order to find exponentially tight bounds for error probabilities of special parts of information. We use the notion of code sequences to simplify our discussion.

A sequence of codes indexed by their block lengths is called *reliable* if

$$\lim_{n\to\infty} {P_e}^{(n)} = 0.$$

For any reliable code sequence $\mathcal{Q}$, the rate $R_{\mathcal{Q}}$ is given by

$$R_{\mathcal{Q}} \triangleq \liminf_{n\to\infty} \frac{\ln|\mathcal{M}^{(n)}|}{n}.$$

The (conventional) error exponent of a reliable sequence is then

$$E_{\mathcal{Q}} \triangleq \liminf_{n\to\infty} \frac{-\ln {P_e}^{(n)}}{n}.$$

Thus, the number of messages in $\mathcal{Q}$ is[3] $\doteq e^{nR_{\mathcal{Q}}}$ and their average error probability decays like $e^{-nE_{\mathcal{Q}}}$ with block length. Now we can define error exponent $E(R)$ in the conventional sense, which is equivalent to the ones given in [13], [17], [20], [25], [36].

Definition 1: For any $R \leq C$ the error exponent $E(R)$ is defined as

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

As mentioned previously, we are interested in UEP when operating at capacity. We already know, [36], that $E(C) = 0$, i.e., the overall error probability cannot decay exponentially at capacity. In the following sections, we show how certain parts of information can still achieve a positive exponent at capacity. In doing that, we are focusing only on the reliable sequences whose rates are equal to C. We call such reliable code sequences *capacity-achieving sequences*.

Throughout the text we denote Kullback–Leibler (KL) divergence between two distributions $\alpha_X(\cdot)$ and $\beta_X(\cdot)$ as $D(\alpha_X(\cdot)\|\beta_X(\cdot))$.

$$D(\alpha_X(\cdot)\|\beta_X(\cdot)) = \sum_{i\in\mathcal{X}} \alpha_X(i) \ln \frac{\alpha_X(i)}{\beta_X(i)}.$$

Similarly, conditional KL divergence between $V_{Y|X}(\cdot|\cdot)$ and $W_{Y|X}(\cdot|\cdot)$ under $P_X(\cdot)$ is given by

$$D\left(V_{Y|X}(\cdot|X)\middle\|W_{Y|X}(\cdot|X)\middle|P_X\right) = \sum_{i\in\mathcal{X}} P_X(i) \sum_{j\in\mathcal{Y}} V_{Y|X}(j|i) \ln \frac{V_{Y|X}(j|i)}{W_{Y|X}(j|i)}.$$

The output distribution that achieves the capacity is denoted by P_Y^* and a corresponding input distribution is denoted by P_X^*.

III. UEP at Capacity: Block Codes Without Feedback

A. Special Bit

We first address the situation where one particular bit (say the first) out of the total $\log_2|\mathcal{M}|$ bits is a special bit—it needs a much better error protection than the overall information. The error probability of the special bit is required to decay as fast as possible while ensuring reliable communication at capacity, for the overall code. The single special bit is denoted by M_1 where $\mathcal{M}_1 = \{0, 1\}$ and over all message M is of the form $M = (M_1, M_2)$ where $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$. The optimal error exponent E_{b} for the special bit is then defined as follows.[4]

Definition 2: For a capacity-achieving sequence $\mathcal{Q}$ with message sets $\mathcal{M}^{(n)} = \mathcal{M}_1 \times \mathcal{M}_2^{(n)}$ where $\mathcal{M}_1 = \{0, 1\}$, the special bit error exponent is defined as

$$E_{\mathrm{b},\mathcal{Q}} \triangleq \liminf_{n\to\infty} \frac{-\ln \Pr^{(n)}\left[\hat{M}_1 \neq M_1\right]}{n}.$$

Then E_{b} is defined as $E_{\mathrm{b}} \triangleq \sup_{\mathcal{Q}} E_{\mathrm{b},\mathcal{Q}}$.

Thus, if $\Pr^{(n)}\left[\hat{M}_1 \neq M_1\right] \doteq \exp(-nE_{\mathrm{b},\mathcal{Q}})$ for a reliable sequence $\mathcal{Q}$, then E_{b} is the supremum of $E_{\mathrm{b},\mathcal{Q}}$ over all capacity-achieving $\mathcal{Q}$'s.

[3]The $\doteq$ sign denotes equality in the exponential sense. For a sequence $a^{(n)}$

$$a^{(n)} \doteq e^{nF} \Leftrightarrow F = \liminf_{n\to\infty} \frac{\ln a^{(n)}}{n}$$

[4]Appendix A discusses a different but equivalent type of definition and shows why it is equivalent to this one. These two types of definitions are equivalent for all the UEP exponents discussed in this paper.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

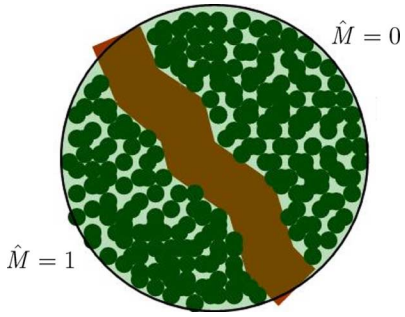


Fig. 1. Splitting the output space into two distant enough clusters.

Since $E(C)=0$, it is clear that the entire information cannot achieve any positive error exponent at capacity. However, it is not clear whether a single special bit can steal a positive error exponent $E_{\rm b}$ at capacity.

Theorem 1:

$$E_{\rm b}=0.$$

This implies that, if we want the error probability of the messages to vanish with increasing block length and the error probability of at least one of the bits to decay with a positive exponent with block length, the rate of the code sequence should be strictly smaller than the capacity.

Proof of the theorem is heavy in calculations, but the main idea behind is the "blowing up lemma" [13]. Conventionally, this lemma is only used for strong converses for various capacity theorems. It is also worth mentioning that the conventional converse techniques like Fano's inequality are not sufficient to prove this result.

Intuitive Interpretation: Let the shaded balls in Fig. 1 denote the minimal decoding regions of the messages. These decoding regions ensure reliable communication, they are essentially the typical noise balls ([11]) around codewords. The decoding regions on the left of the thick line correspond to $\hat{M}_1=1$ and those on the right correspond to the same when $\hat{M}_1=0$. Each of these halves includes half of the decoding regions. Intuitively, the blowing up lemma implies that if we try to add slight extra thickness to the left clusters in Fig. 1, it blows up to occupy almost all the output space. This strange phenomenon in high-dimensional spaces leaves no room for the right cluster to fit. Infeasibility of adding even slight extra thickness implies zero error exponent the special bit.

B. Special Message

Now consider situations where one particular message (say $M=1$) out of the $\doteq e^{nC}$ total messages is a special message—it needs a superior error protection. The missed detection probability for this "emergency" message needs to be minimized. The best missed detection exponent $E_{\rm md}$ is defined as follows.[5]

[5]Note that the definition obtained by replacing $\Pr^{(n)}[\hat{M}\neq 1|M=1]$ by $\min_j \Pr^{(k)}[\hat{M}\neq j|M=j]$ is equivalent to the one given above, since we are taking the supremum over $\mathcal{Q}$ anyway. In short, the message j with smallest conditional error probability could always be relabeled as message 1.

Definition 3: For a capacity-achieving sequence $\mathcal{Q}$, missed detection exponent is defined as

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

Then $E_{\rm md}$ is defined as $E_{\rm md}\triangleq \sup_{\mathcal{Q}} E_{{\rm md},\mathcal{Q}}$.

Compare this with the situation where we aim to protect all the messages uniformly well. If all the messages demand equally good missed detection exponent, then no positive exponent is achievable at capacity. This follows from the earlier discussion about $E(C)=0$. The following theorem shows the improvement in this exponent if we only demand it for a single message instead of all.

Definition 4: The parameter $\tilde{C}$ is defined[6] as the red-alert exponent of a channel

$$\tilde{C}\triangleq \max_{i\in\mathcal{X}} D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|i)\right).$$

We will denote the input letter achieving above maximum by x_r.

Theorem 2:

$$E_{\rm md}=\tilde{C}.$$

Recall that Karush–Kuhn–Tucker (KKT) conditions for achieving capacity imply the following expression for capacity[20, Theorem 4.5.1]:

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

Note that simply switching the arguments of KL divergence within the maximization for C gives us the expression for $\tilde{C}$. The capacity C represents the best possible data rate over a channel, whereas red-alert exponent $\tilde{C}$ represents the best possible protection achievable for a message at capacity.

It is worth mentioning here the "very noisy" channel in [20]. In this formulation [6], the KL divergence is symmetric, which implies

$$D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|i)\right)\approx D\left(W_{Y|X}(\cdot|i)\middle\|\, P_Y^*(\cdot)\right).$$

Hence, the red-alert exponent and capacity become roughly equal. For a symmetric channel such as a binary-symmetric channel (BSC), all inputs can be used as x_r. Since P_Y^* is the uniform distribution for these channels, $\tilde{C}=D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|i)\right)$ for any input letter i. This also happens to be the sphere-packing exponent $E_{\rm sp}(0)$ of this channel [36] at rate 0.

Optimal Strategy: Codewords of a capacity-achieving code are used for the ordinary messages. Codeword for the special message is a repetition sequence of the input letter x_r. For all the output sequences special message is decoded, except for the output sequences with empirical distribution (type) approximately equal to P_Y^*. For the output sequences with empirical distribution approximately P_Y^*, the decoding scheme of the original capacity-achieving code is used.

[6]The authors would like to thank Krishnan Eswaran of University of California, Berkeley for suggesting this name.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

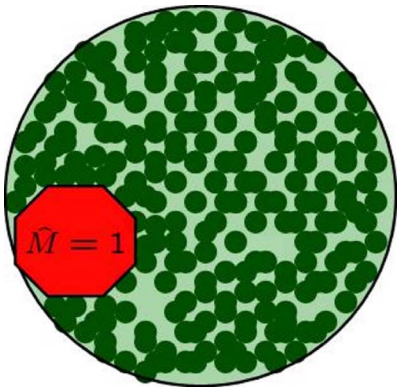


Fig. 2. Avoiding missed-detection.

Indeed, Kudryashov [23] had already suggested the encoding scheme described above, as a subcode for his non-block-variable delay coding scheme. However, discussion in [23] does not make any claims about the optimality of this encoding scheme.

Intuitive Interpretation: Having a large missed detection exponent for the special message corresponds to having a large decoding region for the special message. This ensures that when $M = 1$, i.e., when the special message is transmitted, probability of $\hat{M} \neq 1$ is exponentially small. In a sense $E_{\rm md}$ indicates how large the decoding region of the special message could be made, while still filling $\doteq e^{nC}$ typical noise balls in the remaining space. The hexagon in Fig. 2 denotes such a large region. Note that the actual decoding region of the special message is much larger than this illustration, because it consists of all output types except the ones close to P_Y^*, whereas the ordinary decoding regions only contain the output types close to P_Y^*.

Utility of this result is two-fold: first, the optimality of such a simple scheme was not obvious before; second, as we will see later, protecting a single special message is a key building block for many other problems when feedback is available.

C. Many Special Messages

Now consider the case when instead of a single special message, exponentially many of the total $\doteq e^{nC}$ messages are special. Let $\mathcal{M}_s^{(n)} \subseteq \mathcal{M}^{(n)}$ denote this set of special messages

$$\mathcal{M}_s^{(n)} = \{1, 2, \ldots, \lceil e^{nr} \rceil\}.$$

The best missed detection exponent, achievable simultaneously for all of the special messages, is denoted by $E_{\rm md}(r)$.

Definition 5: For a capacity-achieving sequence $\mathcal{Q}$, the missed detection exponent achieved on sequence of subsets $\mathcal{M}_s$ is defined as

$$E_{\mathrm{md},\mathcal{Q},\mathcal{M}_s} \triangleq \liminf_{n\to\infty} \frac{-\ln \max\limits_{i\in\mathcal{M}_s^{(n)}} \Pr^{(n)}\left[\hat{M} \neq i | M = i\right]}{n}.$$

Then, for a given $r < C$, $E_{\rm md}(r)$ is defined as

$$E_{\rm md}(r) \triangleq \sup_{\mathcal{Q},\mathcal{M}_s} E_{\mathrm{md},\mathcal{Q},\mathcal{M}_s}$$

where maximization is over $\mathcal{M}_s$'s such that

$$\liminf_{n\to\infty} \frac{\ln |\mathcal{M}_s^{(n)}|}{n} = r.$$

This message-wise UEP problem has already been investigated by Csiszár in his paper on joint source–channel coding [12]. His analysis allows for multiple sets of special messages each with its own rate and an overall rate that can be smaller than the capacity.[7]

Essentially, $E_{\rm md}(r)$ is the best value for which missed detection probability of every special message is $\doteq \exp(-nE_{\rm md}(r))$ or smaller. Note that if the only messages in the code are these $\lceil e^{nr} \rceil$ special messages (instead of $|\mathcal{M}^{(n)}| \doteq e^{nC}$ total messages), their best missed detection exponent equals the classical error exponent $E(r)$ discussed earlier.

Theorem 3:

$$E_{\rm md}(r) = E(r), \quad \forall\, r \in [0, C).$$

Thus, we can communicate reliably at capacity and still protect the special messages as if we were only communicating the special messages. Note that the classical error exponent $E(r)$ is yet unknown for the rates below the critical rate (except for zero rate). Nonetheless, this theorem says that whatever $E(r)$ can be achieved for $\lceil e^{nr} \rceil$ messages when they are by themselves in the codebook, can still be achieved when there are $\doteq e^{nC}$ additional ordinary messages requiring reliable communication.

Optimal Strategy: Start with an optimal codebook for $\lceil e^{nr} \rceil$ messages which achieves the 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.

The decoder uses a two-stage decoding rule, in the first stage of which it decides whether or not a special message was sent. If the received sequence is close to one or more of the special codewords, the receiver decides that a special message was sent else it decides an ordinary message was sent. In the second stage, receiver employs maximum-likelihood (ML) decoding either among the ordinary messages or among the special messages depending on its decision in the first stage.

The overall missed detection exponent $E_{\rm md}(r)$ is bottlenecked by the second-stage errors. It is because the first-stage error exponent is essentially the sphere-packing exponent $E_{\rm sp}(r)$, which is never smaller than the second-stage error exponent $E(r)$.

Intuitive Interpretation: This means that we can start with a code of $\lceil e^{nr} \rceil$ 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 $\lceil e^{nr} \rceil$ balls fill up most of the output space, there are still some cavities left between them. These

[7]The authors would like to thank Pulkit Grover of University of California, Berkeley for pointing out this closely related work [12]

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

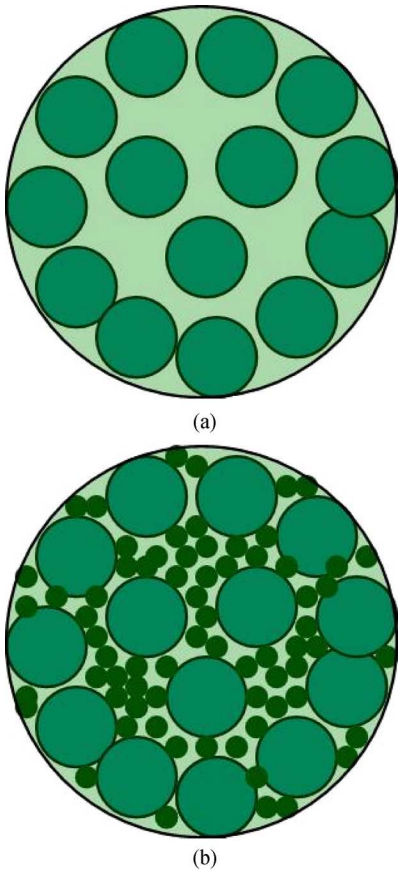


Fig. 3. "There is always room for capacity." (a) Exponent optimal code. (b) Achieving capacity.

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 $\lceil e^{nr} \rceil$ 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 (in terms of the exponent) in spite of the large boulders.

D. Allowing Erasures

In some situations, a decoder may be allowed to declare an erasure when it is not sure about the transmitted message. These erasure events are not counted as errors and are usually followed by a retransmission using a decision feedback protocol like Hybrid-ARQ (hybrid automatic repeat request). This subsection extends the earlier result for $E_{\rm md}(r)$ to the cases when such erasures are allowed.

In decoding with erasures, in addition to the message set $\mathcal{M}$, the decoder can map the received sequence Y^n to a virtual message called "erasure." Let $P_{\rm erasure}$ denote the average erasure probability of a code

$$P_{\rm erasure} = \Pr\left[\hat{M} = \text{erasure}\right].$$

Previously, when there was no erasures, errors were not detected. For errors and erasures decoding, erasures are detected errors, the rest of the errors are undetected errors and P_e denotes the undetected error probability. Thus, average and conditional (undetected) error probabilities are given by

$$P_e = \Pr\left[\hat{M} \neq M, \hat{M} \neq \text{erasure}\right]$$

$$P_e(i) = \Pr\left[\hat{M} \neq M, \hat{M} \neq \text{erasure} \middle| M = i\right].$$

An infinite sequence $\mathcal{Q}$ of block codes with errors and erasures decoding is *reliable*, if its average error probability and average erasure probability, both vanish with n.

$$\lim_{n\to\infty} {P_e}^{(n)} = 0 \quad \text{and} \quad \lim_{n\to\infty} {P_{\rm erasure}}^{(n)} = 0.$$

If the erasure probability is small, then average number of retransmissions needed is also small. Hence, this condition of vanishingly small ${P_{\rm erasure}}^{(n)}$ ensures that the effective data rate of a decision feedback protocol remains unchanged in spite of retransmissions. We again restrict ourselves to reliable sequences whose rate equal C.

We could redefine all previous exponents for decision feedback (df) scenarios, i.e., for reliable codes with erasure decoding. But resulting exponents do not change with the provision of erasures with vanishing probability for single bit or single message problems, i.e., decision feedback protocols such as Hybrid-ARQ does not improve $E_{\rm b}$ or $E_{\rm md}$. Thus, we only discuss the decision feedback version of $E_{\rm md}(r)$.

Definition 6: For a capacity-achieving sequence with erasures, $\mathcal{Q}$, the missed detection exponent achieved on sequence of subsets $\mathcal{M}_s$ is defined as

$$E^{\rm df}_{{\rm md},\mathcal{Q}}(r) \triangleq \liminf_{n\to\infty} \frac{-\ln \max\limits_{i\in\mathcal{M}_s^{(n)}} \Pr^{(n)}\left[\hat{M}\neq i, \hat{M}\neq \text{erasure} | M=i\right]}{n}.$$

Then for a given $r < C$, $E^{\rm df}_{{\rm md},\mathcal{Q}}(r)$ is defined as

$$E^{\rm df}_{{\rm md},\mathcal{Q}}(r) \triangleq \sup_{\mathcal{Q},\mathcal{M}_s} E_{{\rm md},\mathcal{Q},\mathcal{M}_s}$$

where maximization is over $\mathcal{M}_s$'s such that

$$\liminf_{n\to\infty} \frac{\ln|\mathcal{M}_s^{(n)}|}{n} = r.$$

The next theorem shows that allowing erasures increases the missed-detection exponent for r below a critical rate, on symmetric channels.

Theorem 4: For symmetric channels

$$E^{\rm df}_{\rm md}(r) \geq E_{\rm sp}(r), \quad \forall\, r \in [0, C).$$

Coding strategy is similar to the no-erasure case. We first start with an erasure code for $\lceil e^{nr} \rceil$ messages like the one in [18]. Then add randomly generated ordinary codewords to it. Again, a two-stage decoding is performed where the first stage decides between the set of ordinary codewords and the set of special codewords using a threshold distance. If this first stage chooses special codewords, the second stage applies the decoding rule in [18] amongs special codewords. Otherwise, the second stage uses the ML decoding among ordinary codewords.

The overall missed detection exponent $E^{\rm df}_{\rm md}(r)$ is bottlenecked by the first-stage errors. That is because the first-stage error exponent $E_{\rm sp}(r)$ is smaller than the second-stage error

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

exponent $E_{\rm sp}(r)+C-r$. This is in contrast to the case without erasures.

IV. UEP at Capacity: Variable Length Block Codes With Feedback

In the preceding section, we analyzed bit-wise and message-wise UEP problems for fixed-length block codes (without feedback) operating at capacity. In this section, we will revisit the same problems for variable-length block codes with perfect feedback, operating at capacity. Before going into the discussion of the problems, let us recall variable-length block codes with feedback briefly.

A variable-length block code with feedback is composed of a coding algorithm and a decoding rule. Decoding rule determines the decoding time and the message that is decoded then. Possible observations of the receiver can be seen as leaves of $|\mathcal{Y}|$-ary tree, as in [4]. In this tree, all nodes at length 1 from the root denote all $|\mathcal{Y}|$ possible outputs at time $t=1$. All non-leaf nodes among these split into further $|\mathcal{Y}|$ branches in the next time $t=2$ and the branching of the non-leaf nodes continue like this ever after. Each node of depth t in this tree corresponds to a particular sequence, y^t, i.e., a history of outputs until time t. The parent of node y^t is its prefix y^{t-1}. Leaves of this tree form a prefix-free source code, because decision to stop for decoding has to be a casual event. In other words, the event $\{\tau = t\}$ should be measurable in the σ-field generated by Y^t. In addition we have $\Pr[\tau<\infty]=1$ thus decoding time τ is a Markov stopping time with respect to receivers observation. The coding algorithm, on the other hand, assigns an input letter $X_{t+1}(y^t;i)$ to each message $i\in\mathcal{M}$ at each non-leaf node y^t of this tree. The encoder stops transmission of a message when a leaf is reached, i.e., when the decoding is complete.

The codes we consider are block codes in the sense that transmission of each message (packet) starts only after the transmission of the previous one ends. The error probability and rate of the code are simply given by

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

A more thorough discussion of variable-length block codes with feedback can be found in [9] and [4].

Earlier discussion in Section II-B about different kinds of errors is still valid as is but we need to slightly modify our discussion about the reliable sequences. A reliable sequence of variable-length block codes with feedback $\mathcal{Q}$ is any countably infinite collection of codes indexed by integers, such that

$$\lim_{k\to\infty} {P_e}^{(k)}=0.$$

In the rate and exponent definitions for reliable sequences, we replace block length n by the expected decoding time $E[\tau]$. Then, a capacity-achieving sequence with feedback is a reliable sequence of variable -length block codes with feedback whose rate is C

It is worth noting the importance of our assumption that all the entries of the transition probability matrix $W_{Y|X}$ are positive. For any channel with a $W_{Y|X}$ which has one or more zero probability transitions, it is possible to have error-free codes operating at capacity [9]. Thus, all the exponents discussed below are infinite for DMCs with one or more zero probability transitions.

A. Special Bit

Let us consider a capacity-achieving sequence $\mathcal{Q}$ whose message sets are of the form $\mathcal{M}^{(k)}=\mathcal{M}_1\times\mathcal{M}_2^{(k)}$ where $\mathcal{M}_1=\{0,1\}$. Then the error exponent of the M_1, i.e., the initial bit, is defined as follows.

Definition 7: For a capacity-achieving sequence with feedback $\mathcal{Q}$, with message sets $\mathcal{M}^{(k)}$ of the form $\mathcal{M}^{(k)}=\mathcal{M}_1\times\mathcal{M}_2^{(k)}$ where $\mathcal{M}_1=\{0,1\}$, the special bit error exponent is defined as

$$E_{{\rm b},\mathcal{Q}}^f\triangleq\liminf_{k\to\infty}\frac{-\ln\Pr^{(n)}\left[\hat{M}_1\neq M_1\right]}{E\left[\tau^{(k)}\right]}.$$

Then $E_{\rm b}^f$ is defined as $E_{\rm b}^f\triangleq\sup_{\mathcal{Q}}E_{{\rm b},\mathcal{Q}}^f$

Theorem 5:

$$E_{\rm b}^f=\tilde{C}.$$

Recall that without feedback, even a single bit could not achieve any positive error exponent at capacity, Theorem 1. But feedback together with variable decoding time connects the message-wise UEP and the bit-wise UEP and results in a positive exponent for bit-wise UEP. In the following, the described strategy shows how schemes for protecting a special message can be used to protect a special bit.

Optimal Strategy: We use a length $(k+\sqrt{k})$ fixed-length block code with errors and erasures decoding as a building block for our code. The transmitter first transmits M_1 using a short repetition code of length $\sqrt{k}$. If the tentative decision about M_1, $\tilde{M}_1$, is correct after this repetition code, the transmitter sends M_2 with a length k capacity-achieving code. If $\tilde{M}_1$ is incorrect after the repetition code, the transmitter sends the symbol x_r for k time units, where x_r is the input letter i maximizing the $D\left(P_Y^*(\cdot)\|W_{Y|X}(\cdot|i)\right)$. If the output sequence in the second phase, $Y_{\sqrt{k}+1}^{\sqrt{k}+k}$, is not a typical sequence of P_Y^*, an erasure is declared for the block. And the same message is retransmitted by repeating the same strategy afresh. Otherwise, the receiver uses an ML decoder to chose $\hat{M}_2$ and $\hat{M}=(\tilde{M}_1,\hat{M}_2)$.

The erasure probability is vanishingly small, as a result the undetected error probability of M_i in fixed-length erasure code is approximately equal to the error probability of M_i in the variable-length block code. Furthermore, $E[\tau]$ is roughly $(k+\sqrt{k})$ despite the retransmissions. A decoding error for M_1 happens only when $\tilde{M}_1\neq M_1$ and the empirical distribution of the output sequence in the second phase is close to P_Y^*. Note that latter event happens with probability $\doteq e^{-\tilde{C}E[\tau]}$.

B. Many Special Bits

We now analyze the situation where instead of a single special bit there are approximately $E[\tau]\,r/\ln 2$ special bits out of the total $E[\tau]\,C/\ln 2$ (approximately) bits. Hence, we consider

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

the capacity-achieving sequences with feedback having message sets of the form $\mathcal{M}^{(k)} = \mathcal{M}_1^{(k)} \times \mathcal{M}_2^{(k)}$. Unlike in the previous subsection, where the size of $\mathcal{M}_1^{(k)}$ was fixed, we now allow its size to vary with the index of the code. We restrict ourselves to the cases where $\liminf_{k\to\infty} \frac{\ln|\mathcal{M}_1^{(k)}|}{E\left[\tau^{(k)}\right]} = r$. This limit gives us the rate of the special bits. It is worth noting at this point that even when the rate r of special bits is zero, the number of special bits might not be bounded, i.e., $\liminf_{k\to\infty} |\mathcal{M}_1^{(k)}|$ might be infinite. The error exponent $E^f_{\text{bits},\mathcal{Q}}$ at a given rate r of special bits is defined as follows.

Definition 8: For any capacity-achieving sequence with feedback $\mathcal{Q}$ with the message sets $\mathcal{M}^{(k)}$ of the form $\mathcal{M}^{(k)} = \mathcal{M}_1^{(k)} \times \mathcal{M}_2^{(k)}$, $r_{\mathcal{Q}}$ and $E^f_{\text{bits},\mathcal{Q}}$ are defined as

$$r_{\mathcal{Q}} \triangleq \liminf_{k\to\infty} \frac{\ln|\mathcal{M}_1^{(k)}|}{E\left[\tau^{(k)}\right]}$$

$$E^f_{\text{bits},\mathcal{Q}} \triangleq \liminf_{k\to\infty} \frac{-\ln \Pr^{(k)}\left[\hat{M}_1 \neq M_1\right]}{E\left[\tau^{(k)}\right]}.$$

Then $E^f_{\text{bits}}(r)$ is defined as $E^f_{\text{bits}}(r) \triangleq \sup_{\mathcal{Q}:r_{\mathcal{Q}}\geq r} E^f_{\text{bits},\mathcal{Q}}$

The next theorem shows how this exponent decays linearly with rate r of the special bits.

Theorem 6:

$$E^f_{\text{bits}}(r) = \left(1 - \frac{r}{C}\right)\tilde{C}.$$

Notice that the exponent $E^f_{\text{bits}}(0) = \tilde{C}$, i.e., it is as high as the exponent in the single bit case, in spite of the fact that here the number of bits can be growing to infinity with $E[\tau]$. This linear tradeoff between rate and reliability reminds us of Burnashev's result [9].

Optimal Strategy: Like the single bit case, we use a fixed-length block code with erasures as our building block. First, the transmitter sends M_1 using a capacity-achieving code of length $\frac{r}{C}k$. If the tentative decision $\tilde{M}_1$ is correct, the transmitter sends M_2 with a capacity-achieving code of length $(1-\frac{r}{C})k$. Otherwise, the transmitter sends channel input x_r for $(1-\frac{r}{C})k$ time units. If the output sequence in the second phase is not typical with P_Y^*, an erasure is declared and the same strategy is repeated afresh. Else, receiver uses an ML decoder to decide $\hat{M}_2$ and decodes the message M as $\hat{M} = (\tilde{M}_1, \hat{M}_2)$. A decoding error for M_1 happens only when an error happens in the first phase and the output sequence in the second phase is typical with P_Y^* when the reject codeword is sent. But the probability of the later event is $\doteq e^{-(1-\frac{r}{C})\tilde{C}k}$. The factor of $(1-\frac{r}{C})$ arises because of the relative duration of the second phase to the overall communication block. Similar to the single bit case, erasure probability remains vanishingly small in this case. Thus, not only the expected decoding time of the variable-length block code is roughly equal to the block length of the fixed-length block code, but also its error probabilities are roughly equal to the corresponding error probabilities associated with the fixed-length block code.

C. Multiple Layers of Priority

We can generalize this result to the case when there are multiple levels of priority, where the most important layer contains $E[\tau]\, r_1/\ln 2$ bits, the second-most important layer contains $E[\tau]\, r_2/\ln 2$ bits, and so on. For an L-layer situation, message set $\mathcal{M}^{(k)}$ is of the form $\mathcal{M}^{(k)} = \mathcal{M}_1^{(k)} \times \mathcal{M}_2^{(k)} \times \cdots \times \mathcal{M}_L^{(k)}$. We assume, without loss of generality , that the order of importance of the M_i's is $M_1 \succ M_2 \succ \cdots \succ M_L$. Hence, we have $P_e^{M_1} \leq P_e^{M_2} \leq \cdots \leq P_e^{M_L}$.

Then for any L-layer capacity-achieving sequence with feedback, we define the error exponent of the sth layer as

$$E^f_{\text{bits},s,\mathcal{Q}} = \liminf_{k\to\infty} \frac{-\ln \Pr^{(k)}\left[\hat{M}_s \neq M_s\right]}{E\left[\tau^{(k)}\right]}.$$

The achievable error exponent region of the L-layered capacity-achieving sequences with feedback is the set of all achievable exponent vectors $(E^f_{\text{bits},1,\mathcal{Q}}, E^f_{\text{bits},2,\mathcal{Q}}, \ldots, E^f_{\text{bits},L-1,\mathcal{Q}})$. The following theorem determines that region.

Theorem 7: The achievable error exponent region of the L-layered capacity-achieving sequences with feedback, for rate vector $(r_1, r_2, \ldots, r_{L-1})$ is the set of vectors $(E_1, E_2, \ldots, E_{L-1})$ satisfying

$$E_i \leq \left(1 - \frac{\sum_{j=1}^{i} r_j}{C}\right)\tilde{C}, \quad \forall i \in \{1, 2, \ldots, (L-1)\}.$$

Note that the least important layer cannot achieve any positive error exponent because we are communicating at capacity, i.e., $E_L = 0$.

Optimal Strategy: The transmitter first sends the most important layer, M_1, using a capacity-achieving code of length $\frac{r_1}{C}k$. If it is decoded correctly, then it sends the next layer with a capacity-achieving code of length $\frac{r_2}{C}k$. Else, it starts sending the input letter x_r for not only $\frac{r_2}{C}k$ time units but also for all remaining $L-2$ phases. The same strategy is repeated for $M_3, M_4, \ldots, M_L$.

Once the whole block of channel outputs Y^k is observed; the receivers check the empirical distribution of the output in all of the phases except the first one. If they are all typical with P_Y^*, the receiver uses the tentative decisions to decode, $\hat{M} = (\tilde{M}_1, \tilde{M}_2, \ldots \tilde{M}_L)$. If one or more of the output sequences are not typical with P_Y^*, an erasure is declared for the whole block and transmission starts from scratch.

For each layer i, with the above strategy we can achieve an exponent as if there were only two kinds of bits (as in Theorem 6)

- bits in layer i or in more important layers $k < i$ (i.e., special bits);
- bits in less important layers (i.e., ordinary bits).

Hence, Theorem 7 does not only specify the optimal performance when there are multiple layers, but also shows that the performance we observed in Theorem 6 is successively refinable. Fig. 4 shows these simultaneously achievable exponents of Theorem 6, for a particular rate vector $(r_1, r_2, \ldots, r_{L-1})$.

Note that the most important layer can achieve an exponent close to $\tilde{C}$ if its rate is close to zero. As we move to the layers

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

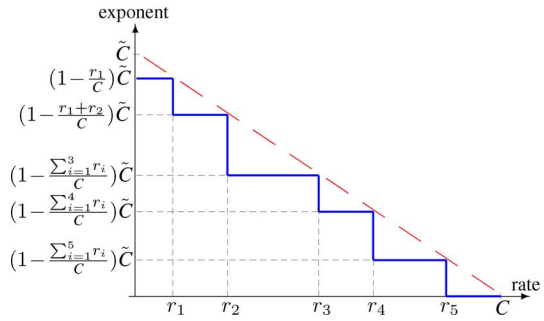


Fig. 4. Successive refinability for multiple layers of priority, demonstrated on an example with six layers; $\sum_{i=1}^{6} r_i = C$.

with decreasing importance, the achievable error exponent decays gradually.

D. Special Message

Now consider one particular message, say the first one, which requires small missed-detection probability. Similar to the no-feedback case, define E_{md}^{f} as its missed-detection exponent at capacity.

Definition 9: For any capacity-achieving sequence with feedback $\mathcal{Q}$, the missed detection exponent is defined as

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

Then E_{md}^{f} is defined as $E_{\mathrm{md}}^{f} \triangleq \sup_{\mathcal{Q}} E_{\mathrm{md},\mathcal{Q}}^{f}$.

Theorem 8:

$$E_{\mathrm{md}}^{f} = \tilde{C}.$$

Theorem 2 and 8 implies following corollary.

Corollary 1: Feedback does not improve the missed detection exponent of a single special message: $E_{\mathrm{md}}^{f} = E_{\mathrm{md}}$.

If the red-alert exponent were defined as the best protection of a special message achievable at capacity, then this result could have been thought of as an analog of "feedback does not increase capacity" for the red-alert exponent. Also note that with feedback, E_{md}^{f} for the special message and E_{b}^{f} for the special bit are equal.

E. Many Special Messages

Now let us consider the problem where the first $\lceil e^{E[\tau]r}\rceil$ messages are special, i.e., $\mathcal{M}_s = \{1,2,\ldots,\lceil e^{E[\tau]r}\rceil\}$. Unlike previous problems, now we will also impose a uniform expected delay constraint as follows.

Definition 10: For any reliable variable length block code with feedback

$$\Gamma \triangleq \frac{\max_{i\in\mathcal{M}} E\left[\tau\middle| M=i\right]}{E\left[\tau\right]}.$$

A reliable sequence with feedback $\mathcal{Q}$ is a uniform delay reliable sequence with feedback if and only if $\lim_{k\to\infty} \Gamma^{(k)} = 1$.

This means that the average $E\left[\tau\middle| M=i\right]$ for every message i is essentially equal to $E\left[\tau\right]$ (if not smaller). This uniformity constraint reflects a system requirement for ensuring a robust delay performance, which is invariant of the transmitted message.[8] Let us define the missed-detection exponent $E_{\mathrm{md}}^{f}(r)$ under this uniform delay constraint.

Definition 11: For any uniform delay capacity-achieving sequence with feedback, $\mathcal{Q}$, the missed detection exponent achieved on sequence of subsets $\mathcal{M}_s$ is defined as

$$E_{\mathrm{md},\mathcal{Q},\mathcal{M}_s}^{f} \triangleq \liminf_{n\to\infty} \frac{-\ln \max_{i\in\mathcal{M}_s^{(k)}} \Pr^{(k)}\left[\hat{M}\neq i | M=i\right]}{E\left[\tau^{(k)}\right]}.$$

Then for a given $r < C$, we define

$$E_{\mathrm{md}}^{f}(r) \triangleq \sup_{\mathcal{Q},\mathcal{M}_s} E_{\mathrm{md},\mathcal{Q},\mathcal{M}_s}^{f}$$

where maximization is over $\mathcal{M}_s$'s such that

$$\liminf_{k\to\infty} \frac{\ln|\mathcal{M}_s^{(k)}|}{E\left[\tau^{(k)}\right]} = r.$$

The following theorem shows that the special messages can achieve the minimum of the red-alert exponent and the Burnashev's exponent at rate r.

Theorem 9:

$$E_{\mathrm{md}}^{f}(r) = \min\left\{\tilde{C}, (1-\frac{r}{C})D_{\max}\right\}, \quad \forall\, r < C.$$

where $D_{\max} \triangleq \max_{i,j\in\mathcal{X}} D\left(W_{Y|X}(\cdot|i)\middle\| W_{Y|X}(\cdot|j)\right)$.

For $r \in [0,(1-\frac{\tilde{C}}{D_{\max}})C]$, each special message achieves the best missed detection exponent $\tilde{C}$ for a single special message, as if the rest of the special messages were absent. For $r \in [(1-\frac{\tilde{C}}{D_{\max}})C, C)$, special messages achieve the Burnashev's exponent as if the ordinary messages were absent.

The optimal strategy is based on transmitting a special bit first. This result demonstrates, yet another time, how feedback connects bit-wise UEP with message-wise UEP. In the optimal strategy for bit-wise UEP with many bits a special message was used, whereas now in message-wise UEP with many messages a special bit is used. The roles of bits and messages, in two optimal strategies are simply swapped between the two cases.

Optimal Strategy: We combine the strategy for achieving $\tilde{C}$ for a special bit and the Yamamoto–Itoh strategy for achieving Burnashev's exponent [40]. In the first phase, a special bit, b, is sent with a repetition code of $\sqrt{k}$ symbols. This is the indicator bit for special messages: it is 1 when a special message is to be sent and 0 otherwise.

If b is decoded incorrectly as $\hat{b}=0$, input letter x_r is sent for the remaining k time unit. If it is decoded correctly as $\hat{b}=0$, then the ordinary message is sent using a codeword from a capacity-achieving code. If the output sequence in the second

[8]Optimal exponents in all previous problems remain unchanged irrespective of this uniform delay constraint.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

phase is typical with P_Y^* receiver use an ML decoder to chose one of the ordinary messages, else an erasure is declared for $(k+\sqrt{k})$ long block.

If $\hat{b}=1$, then a length k two-phase code with errors and erasure decoding, like the one given in [40] by Yamamoto and Itoh, is used to send the message. In the communication phase, a length $\frac{r}{C}k$ capacity-achieving code is used to send the message, M, if $M \in \mathcal{M}_s$. If $M \notin \mathcal{M}_s$, an arbitrary codeword from the length $\frac{r}{C}k$ capacity-achieving code is sent. In the control phase, if $M \in \mathcal{M}_s$ and if it is decoded correctly at the end of communication phase, the accept letter x_a is sent for $(1-\frac{r}{C})k$ time units, else the reject letter, x_d, is sent for $(1-\frac{r}{C})k$ time units. If the empirical distribution in the control phase is typical with $W_{Y|X}(\cdot|x_a)$ then special message decoded at the end of the communication phase becomes the final $\hat{M}$, else an erasure is declared for $(k+\sqrt{k})$ long block.

Whenever an erasure is declared for the whole block, transmitter and receiver applies the above strategy again from scratch. This scheme is repeated until nonerasure decoding is reached.

V. Avoiding False Alarms

In the previous sections while investigating message-wise UEP we have only considered the missed detection formulation of the problems. In this section, we will focus on an alternative formulation of message-wise UEP problems based on false alarm probabilities.

A. Block Codes Without Feedback

We first consider the no-feedback case. When false alarm of a special message is a critical event, e.g., the "reboot" instruction, the false alarm probability $\Pr\left[\hat{M}=1|M\neq 1\right]$ for this message should be minimized, rather than the missed detection probability $\Pr\left[\hat{M}\neq 1|M=1\right]$.

Using Bayes' rule and assuming uniformly chosen messages we get

$$\Pr\left[\hat{M}=1|M\neq 1\right]=\frac{\Pr\left[\hat{M}=1,M\neq 1\right]}{\Pr\left[M\neq 1\right]}=\frac{\sum_{j\neq 1}\Pr\left[\hat{M}=1|M=j\right]}{(|\mathcal{M}|-1)}.$$

In classical error exponent analysis, [20], the error probability for a given message usually means its missed detection probability. However, examples such as the "reboot" message necessitate this notion of false alarm probability.

Definition 12: For a capacity-achieving sequence, $\mathcal{Q}$, such that

$$\limsup_{n\to\infty}\Pr{}^{(n)}\left[\hat{M}\neq 1\middle|\, M=1\right]=0,$$

false alarm exponent is defined as

$$E_{\mathrm{fa},\mathcal{Q}}\triangleq\liminf_{n\to\infty}\frac{-\ln\Pr{}^{(n)}\left[\hat{M}=1\middle|\, M\neq 1\right]}{n}.$$

Then E_{fa} is defined as $E_{\mathrm{fa}}\triangleq\sup_{\mathcal{Q}}E_{\mathrm{fa},\mathcal{Q}}$.

Thus, E_{fa} is the best exponential decay rate of false alarm probability with n. Unfortunately, we do not have the exact expression for E_{fa}. However, the upper bound given below is sufficient to demonstrate the improvement introduced by feedback and variable decoding time.

Theorem 10:

$$E_{\mathrm{fa}}^{\mathrm{l}}\leq E_{\mathrm{fa}}\leq E_{\mathrm{fa}}^{\mathrm{u}}.$$

The upper and lower bounds to the false alarm exponent are given by

$$E_{\mathrm{fa}}^{\mathrm{l}}\triangleq\max_{i\in\mathcal{X}}\min_{\substack{V_{Y|X}:\\(P^*V)_Y(\cdot)=W_{Y|X}(\cdot|i)}}D(V_{Y|X}(\cdot|X)\|\, W_{Y|X}(\cdot|X)|\, P_X^*)$$

$$E_{\mathrm{fa}}^{\mathrm{u}}\triangleq\max_{i\in\mathcal{X}}D\left(W_{Y|X}(\cdot|i)\|\, W_{Y|X}(\cdot|X)|\, P_X^*\right)$$

where $(P^*V)_Y(\cdot)=\sum_j V_{Y|X}(\cdot|j)P_X^*(j)$.

The maximizers of the optimizations for $E_{\mathrm{fa}}^{\mathrm{l}}$ and $E_{\mathrm{fa}}^{\mathrm{u}}$ are denoted by x_{f_l} and x_{f_u}

$$E_{\mathrm{fa}}^{\mathrm{l}}=\min_{\substack{V_{Y|X}:\\(P^*V)_Y(\cdot)=W_{Y|X}(\cdot|x_{f_l})}}D\left(V_{Y|X}(\cdot|X)\|\, W_{Y|X}(\cdot|X)|\, P_X^*\right)$$

$$E_{\mathrm{fa}}^{\mathrm{u}}=D\left(V_{Y|X}(\cdot|x_{f_u})\|\, W_{Y|X}(\cdot|X)|\, P_X^*\right).$$

Strategy to Reach Lower Bound: The codeword for the special message $M=1$ is a repetition sequence of input letter x_{f_l}. Its decoding region is the typical "noise ball" around it, the output sequences whose empirical distribution is approximately equal to $W_{Y|X}(\cdot|x_{f_l})$. For ordinary messages, we use a capacity-achieving codebook, where all codewords have the same empirical distribution (approximately) P_X^*. Then for y^n whose empirical distribution is not in the typical "noise ball" around the special codeword, the receiver performs ML decoding among the ordinary codewords.

Note the contrast between this strategy for achieving $E_{\mathrm{fa}}^{\mathrm{l}}$ and the optimal strategy for achieving E_{md}. For achieving E_{md}, the output sequences of any type other than the ones close to P_Y^* were decoded as the special message; whereas for achieving E_{fa}, only the output sequences of types that are close to $W_{Y|X}(\cdot|x_{f_l})$ are decoded as the special message.

Intuitive Interpretation: A false-alarm exponent for the special message corresponds to having the smallest possible decoding region for the special message. This ensures that when some ordinary message is transmitted, the probability of the event $\{\hat{M}=1\}$ is exponentially small. We cannot make it too small though, because when the special message is transmitted, the probability of the very same event should be almost one. Hence, the decoding region of the special message should at least contain the typical noise ball around the special codeword. The small hexagon in Fig. 5 denotes such a region.

Note that $E_{\mathrm{fa}}^{\mathrm{l}}$ is larger than channel capacity C due to the convexity of KL divergence

$$E_{\mathrm{fa}}^{\mathrm{l}}=\max_{i\in\mathcal{X}}\min_{\substack{V_{Y|X}:\\(P^*V)_Y(\cdot)=W_{Y|X}(\cdot|i)}}D(V_{Y|X}(\cdot|X)\|\, W_{Y|X}(\cdot|X)|\, P_X^*)$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

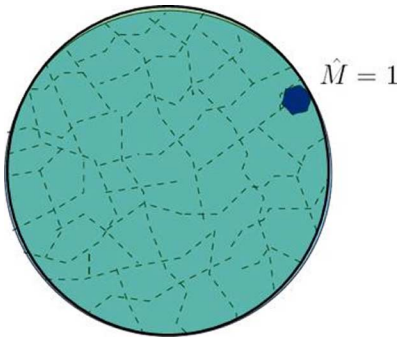


Fig. 5. Avoiding false alarm.

$$\begin{aligned} &> \max_{i\in\mathcal{X}} \min_{\substack{V_{Y|X}:\\ (P^*V)_Y(\cdot)=W_{Y|X}(\cdot|i)}} D\left((P^*V)_Y(\cdot)\|(P^*W)_Y(\cdot)\right) \\ &= \max_{i\in\mathcal{X}} D\left(W_{Y|X}(\cdot|i)\middle\| P_Y^*(\cdot)\right) \\ &= C \end{aligned}$$

where P_Y^* denotes the output distribution corresponding to the capacity-achieving input distribution P_X^* and the last equality follows from KKT condition for achieving capacity we mentioned previously [20, Theorem 4.5.1].

Now we can compare our result for a special message with a similar result for classical situation where all messages are treated equally. It turns out that if every message in a capacity-achieving code demands equally good false-alarm exponent, then this uniform exponent cannot be larger than C. This result seems to be directly connected with the problem of identification via channels [1]. We can prove the achievability part of their capacity theorem using an extension of the achievability part of $E_{\mathrm{fa}}^{\mathrm{l}}$. Perhaps a new converse of their result is also possible using such results. Furthermore, we see that reducing the demand of false-alarm exponent to only one message, instead of all, enhances it from C to at least $E_{\mathrm{fa}}^{\mathrm{l}}$.

B. Variable-Length Block Codes With Feedback

Recall that feedback does not improve the missed-detection exponent for a special message. On the contrary, the false-alarm exponent of a special message is improved when feedback is available and variable decoding time is allowed. We again restrict to uniform delay capacity-achieving sequences with feedback, i.e., capacity-achieving sequences satisfying $\lim_{k\to\infty}\Gamma^{(k)}=1$.

Definition 13: For a uniform delay capacity-achieving sequence with feedback, $\mathcal{Q}$, such that

$$\limsup_{k\to\infty}\Pr{}^{(k)}\left[\hat{M}\neq 1\middle| M=1\right]=0$$

the false-alarm exponent is defined as

$$E_{\mathrm{fa},\mathcal{Q}}^{f}\triangleq\liminf_{k\to\infty}\frac{-\ln\Pr{}^{(k)}\left[\hat{M}=1\middle| M\neq 1\right]}{E\left[\tau^k\right]}.$$

Then E_{fa}^{f} is defined as $E_{\mathrm{fa}}^{f}\triangleq\sup_{\mathcal{Q}}E_{\mathrm{fa},\mathcal{Q}}^{f}$.

Theorem 11:

$$E_{\mathrm{fa}}^{f}=D_{\max}.$$

Note that $D_{\max}>E_{\mathrm{fa}}^{\mathrm{u}}$. Thus, feedback strictly improves the false-alarm exponent, $E_{\mathrm{fa}}^{f}>E_{\mathrm{fa}}$.

Optimal Strategy: We use a strategy similar to the one employed in proving Theorem 9 in Section IV-E. In the first phase, a length $\sqrt{k}$ code is used to convey whether $M=1$ or not, using a special bit $b=\mathbb{I}_{\{M=1\}}$.

- If $\hat{b}=0$, a length k capacity-achieving code with $E_{\mathrm{md}}=\tilde{C}$ is used. If the decoded message for the length k code is 1, an erasure is declared for $(k+\sqrt{k})$ long block. Else the decoded message of length k code becomes the decoded message for the whole $(k+\sqrt{k})$ long block.
- If $\hat{b}=1$,
  - and $M=1$, input symbol x_a is transmitted for k time units.
  - and $M\neq 1$, input symbol x_d is transmitted for k time units.

  If the output sequence, $Y_{\sqrt{k}+1}^{\sqrt{k}+k}$, is typical with $W_{Y|X}(\cdot|x_a)$ then $\hat{M}=1$ else, an erasure is declared for $(k+\sqrt{k})$ long block.

The receiver and the transmitter starts from scratch if an erasure is declared at the end of second phase.

Note that, this strategy simultaneously achieves the optimal missed-detection exponent $\tilde{C}$ and the optimal false-alarm exponent $D_{\max}$ for this special message.

VI. Future Directions

In this paper, we have restricted our investigation of UEP problems to data rates that are essentially equal to the channel capacity. The scenarios we have analyzed provide us with a rich class of problems when we consider data rates below capacity.

Most of the UEP problems has a coding-theoretic version. In these coding-theoretic versions deterministic guarantees, in terms of Hamming distances, are demanded instead of the probabilistic guarantees, in terms of error exponents. As we have mentioned in Section I-A, coding-theoretic versions of bit-wise UEP problems have been extensively studied for the case of linear codes. But it seems that coding-theoretic versions of both message-wise UEP problems and bit-wise UEP problem for nonlinear codes have been scarcely investigated [3], [5].

Throughout this paper, we focused on the channel coding component of communication. However, often times, the final objective is to communicate a source within some distortion constraint. The message-wise UEP problem itself has first come up within this framework [12]. But the source we are trying to convey can itself be heterogeneous, in the sense that some part of its output may demand a smaller distortion than other parts. Understanding optimal methods for communicating such sources over noisy channels present many novel joint source–channel coding problems.

At times, the final objective of communication is achieving some coordination between various agents [14]. In these scenarios, the channel is used for both communicating data and achieving coordination. A new class of problem lends itself to us when we try to figure out the tradeoffs between error exponents of the coordination and data?

We can also actively use UEP in network protocols. For example, a relay can forward some partial information even if

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

it cannot decode everything. This partial information could be characterized in terms of special bits as well as special messages. Another example is two-way communication, where UEP can be used for more reliable feedback and synchronization.

Information-theoretic understanding of UEP also gives rise to some network optimization problems. With UEP, the interface to physical layer is no longer bits. Instead, it is a collection of various levels of error protection. The achievable channel resources of reliability and rate need to be efficiently divided among these levels, which gives rise to many resource allocation problems.

VII. Block Codes Without Feedback: Proofs

In the following sections, we use the following standard notation for entropy, conditional entropy, and mutual information:

$$H(P_X) = \sum_{j\in\mathcal{X}} P_X(j)\log\frac{1}{P_X(j)}$$

$$H(W_{Y|X}|P_X) = \sum_{\substack{j\in\mathcal{X}\\ k\in\mathcal{Y}}} P_X(j)W_{Y|X}(k|j)\log\frac{1}{W_{Y|X}(k|j)}$$

$$I(P,W) = \sum_{\substack{j\in\mathcal{X}\\ k\in\mathcal{Y}}} P_X(j)W_{Y|X}(k|j)\log\frac{W_{Y|X}(k|j)}{(PW)_Y(k)}.$$

where $(PW)_Y(k) = \sum_{i\in\mathcal{X}} W_{Y|X}(k|i)P_X(i)$.

In addition, we denote the decoding region of a message $i \in \mathcal{M}$ by $\mathcal{G}(i)$, i.e.,

$$\mathcal{G}(i) \triangleq \{y^n : \hat{M}(y^n) = i\}.$$

A. Proof of Theorem 1

Proof: We first show that any capacity-achieving sequence $\mathcal{Q}$ with $E_{\mathrm{b},\mathcal{Q}}$ can be used to construct another capacity-achieving sequence, $\mathcal{Q}'$ with $E_{\mathrm{b},\mathcal{Q}'} = \frac{E_{\mathrm{b},\mathcal{Q}}}{2}$, all members of which are fixed composition codes. Then we show that $E_{\mathrm{b},\mathcal{Q}'} = 0$ for any capacity-achieving sequence, $\mathcal{Q}'$ which only includes fixed composition codes.

Consider a capacity-achieving sequence $\mathcal{Q}$ with message sets $\mathcal{M}^{(n)} = \mathcal{M}_1 \times \mathcal{M}_2^{(n)}$, where $\mathcal{M}_1 = \{0,1\}$. As a result of Markov inequality, at least $\frac{4}{5}|\mathcal{M}^{(n)}|$ of the messages in $\mathcal{M}^{(n)}$ satisfy

$$\Pr\left[\hat{M}_1 \neq M_1 \middle| M = i\right] \leq 5\Pr\left[\hat{M}_1 \neq M_1\right]. \tag{1}$$

Similarly, at least $\frac{4}{5}|\mathcal{M}^{(n)}|$ of the messages in $\mathcal{M}^{(n)}$ satisfy

$$\Pr\left[\hat{M} \neq M \middle| M = i\right] \leq 5\Pr\left[\hat{M} \neq M\right]. \tag{2}$$

Thus, at least $\frac{3}{5}|\mathcal{M}^{(n)}|$ of the messages in $\mathcal{M}^{(n)}$ satisfy both (1) and (2). Consequently, at least $\frac{1}{10}|\mathcal{M}^{(n)}|$ messages are of the form $(0, M_2)$ and satisfy (1) and (2). If we group them according to their empirical distribution, at least one of the groups will have more than $\frac{|\mathcal{M}^{(n)}|}{10(n+1)^{|\mathcal{X}|}}$ messages, because the number of different empirical distributions for elements of $\mathcal{X}^n$ is less than $(n+1)^{|\mathcal{X}|}$. We keep the first $\frac{|\mathcal{M}^{(n)}|}{10(n+1)^{|\mathcal{X}|}}$ codewords of this most populous type, denote them by $\bar{\boldsymbol{x}}'_A(\cdot)$, and throw away all of other codewords corresponding to the messages of the form $(0, M_2)$. We do the same for the messages of the form $M = (1, M_2)$ and denote corresponding codewords by $\bar{\boldsymbol{x}}'_B(\cdot)$.

Thus, we have a length n code with message set $\mathcal{M}'$ of the form $\mathcal{M}' = \mathcal{M}_1 \times \mathcal{M}'_2$ where $\mathcal{M}_1 = \{0,1\}$ and $|\mathcal{M}'_2| = \frac{|\mathcal{M}'_2|}{10(n+1)^{|\mathcal{X}|}}$. Furthermore, $\forall i \in \mathcal{M}'$

$$\Pr\left[\hat{M}'_1 \neq M'_1 \middle| M' = i\right] \leq 5\Pr\left[\hat{M}_1 \neq M_1\right]$$
$$\Pr\left[\hat{M}' \neq M' \middle| M' = i\right] \leq 5\Pr\left[\hat{M} \neq M\right].$$

Now let us consider following $2n$-long block code with message set $\mathcal{M}'' = \mathcal{M}_1 \times \mathcal{M}''_2 \times \mathcal{M}''_3$ where $\mathcal{M}''_2 = \mathcal{M}''_3 = \mathcal{M}'_2$. If $M'' = (0, M''_2, M''_3)$, then $\bar{\boldsymbol{x}}(M'') = \bar{\boldsymbol{x}}'_A(M''_2)\bar{\boldsymbol{x}}'_B(M''_3)$. If $M'' = (1, M''_2, M''_3)$ then $\bar{\boldsymbol{x}}(M'') = \bar{\boldsymbol{x}}'_B(M''_2)\bar{\boldsymbol{x}}'_A(M''_3)$. Decoder of this new length $2n$ code uses the decoder of the original length n code first on y^n and then on y_{n+1}^{2n}. If the concatenation of length n codewords corresponding to the decoded halves, is a codeword for an $i \in \mathcal{M}''$, then $\hat{M}'' = i$. Else, an arbitrary message is decoded. One can easily see that the error probability of the length $2n$ code is less than twice the error probability of the length n code, i.e.,

$$\begin{aligned}\Pr\left[\hat{M}'' \neq M'' \middle| M''\right] &\leq \Pr\left[\hat{M}''_2 \neq M''_2 \middle| M''\right] \\ &\quad + \Pr\left[\hat{M}''_3 \neq M''_3 \middle| M''\right] \\ &= \Pr\left[\hat{M}' \neq M' \middle| M' = M''_2\right] \\ &\quad + \Pr\left[\hat{M}' \neq M' \middle| M' = M''_3\right] \\ &\leq 2\Pr\left[\hat{M}' \neq M'\right].\end{aligned}$$

Furthermore, bit error probability of the new code is also at most twice the bit error probability of the length n code, i.e.,

$$\begin{aligned}\Pr\left[\hat{M}''_1 \neq M''_1 \middle| M''_1\right] &\leq \Pr\left[\hat{M}'_1 \neq M'_1 \middle| M'_1 = M''_1\right] \\ &\quad + \Pr\left[\hat{M}'_1 \neq M'_1 \middle| M'_1 = M''_1\right] \\ &\leq 2\Pr\left[\hat{M}'_1 \neq M'_1\right].\end{aligned}$$

Thus, using these codes one can obtain a capacity-achieving sequence $\mathcal{Q}'$ with $E_{\mathrm{b},\mathcal{Q}'} = \frac{E_{\mathrm{b},\mathcal{Q}}}{2}$ all members of which are fixed composition codes.

In the following discussion, we focus on capacity-achieving sequences, $\mathcal{Q}$'s which are composed of fixed composition codes only. We will show that $E_{\mathrm{b},\mathcal{Q}} = 0$ for all capacity-achieving $\mathcal{Q}$'s with fixed composition codes. Consequently, the discussion above implies that $E_{\mathrm{b}} = 0$.

We call the empirical distribution of a given output sequence, y^n, conditioned on the codeword $\bar{\boldsymbol{x}}(i)$, the conditional type of y^n given the message i and denote it by $\mathsf{V}(y^n, i)$. Furthermore, we call the set of y^n's whose conditional type with message i is V, the V-shell of i, and denote it by $\mathsf{T}_V(i)$. Similarly, we denote the set of output sequences y^n with the empirical distribution U_Y, by T_{U_Y}.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

We denote the empirical distribution of the codewords of the nth code of the sequence by $P_X^{(n)}$ and the corresponding output distribution by $P_Y^{(n)}$, i.e.,

$$P_Y^{(n)}(\cdot) = \sum_{i\in\mathcal{X}} W_{Y|X}(\cdot|i)P_X^{(n)}(i).$$

We simply use P_X and P_Y whenever the value of n is unambiguous from the context. Furthermore, $\mathbb{P}_Y^n(\cdot)$ stands for the probability measure on $\mathcal{Y}^n$ such that

$$\mathbb{P}_Y^n(y^n) = \prod_{k=1}^{n} P_Y(y_k).$$

$\mathcal{S}_{0,V}^{(n)}$ is the set of y^n's for which $\hat{M}_1 = 0$ and $\mathsf{V}(y^n, \hat{M}(y^n)) = V$.

$$\mathcal{S}_{0,V}^{(n)} \triangleq \left\{ y^n : \begin{array}{l} \mathsf{V}(y^n, \hat{M}(y^n)) = V \text{and} \\ \hat{M}(y^n) = (0,j) \text{ for some } j \in \mathcal{M}_2 \end{array} \right\}. \quad (3)$$

In other words, $\mathcal{S}_{0,V}^{(n)}$ is the set of y^n's such that $y^n \in \mathsf{T}_V(\hat{M}(y^n))$ and decoded value of the first bit is zero. Note that since for each $y^n \in \mathcal{Y}^n$ there is a unique $\hat{M}(y^n)$ and for each $y^n \in \mathcal{Y}^n$ and message $i \in \mathcal{M}$ there is unique $\mathsf{V}(y^n, i)$; each y^n belongs to a unique $\mathcal{S}_{0,V}^{(n)}$ or $\mathcal{S}_{1,V}^{(n)}$, i.e., $\mathcal{S}_{0,V}^{(n)}$'s and $\mathcal{S}_{1,V}^{(n)}$'s are disjoint sets that collectively cover the set $\mathcal{Y}^n$.

Let us define the typical neighborhood of $W_{Y|X}$ as $[W]$

$$[W] \triangleq \left\{ V_{Y|X} : \begin{array}{r} |V_{Y|X}(j|i)P_X^{(n)}(i) - W_{Y|X}(j|i)P_X^{(n)}(i)| \\ \leq \sqrt[4]{1/n} \quad \forall i,j \end{array} \right\}. \quad (4)$$

Let us denote the union of all $\mathcal{S}_{0,V}^{(n)}$'s for typical V's by $\mathcal{S}_0^{(n)} = \bigcup_{V\in[W]} \mathcal{S}_{0,V}^{(n)}$. We will establish the following inequality later. Let us assume for the moment that it holds.

$$\mathbb{P}_Y^n\left(\mathcal{S}_0^{(n)}\right) \geq e^{n(R^{(n)} - (C+\epsilon_n))}\left(\frac{1}{2} - \frac{|\mathcal{X}||\mathcal{Y}|}{8\sqrt{n}} - P_e\right) \quad (5)$$

where $\lim_{n\to\infty} \epsilon_n = 0$.

As a result of bound given in (5) and the blowing up lemma [13, Ch. 1, Lemma 5.4], we can conclude that for any capacity-achieving sequence $\mathcal{Q}$, there exists a sequence of (ℓ_n, η_n) pairs satisfying $\lim_{n\to\infty} \eta_n = 1$ and $\lim_{n\to\infty} \frac{\ell_n}{n} = 0$ such that

$$\mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_0^{(n)})\right) \geq \eta_n$$

where $\Gamma^{\ell_n}(A)$ is the set of all y^n's which differs from an element of A in at most ℓ_n places. Clearly, one can repeat the same argument for $\Gamma^{\ell_n}(\mathcal{S}_1^{(n)})$ to get

$$\mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_1^{(n)})\right) \geq \eta_n.$$

Consequently

$$\begin{aligned} &\mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_0^{(n)}) \bigcap \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})\right) \\ &= \mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_0^{(n)})\right) + \mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_1^{(n)})\right) \\ &\quad - \mathbb{P}_Y^n\left(\Gamma^{\ell_n}(\mathcal{S}_0^{(n)}) \bigcup \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})\right) \\ &\geq 2\eta_n - 1. \end{aligned}$$

Note that if $y^n \in \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})$, then there exist at least one element $\tilde{y}^n \in \mathsf{T}_{P_Y}$ which differs from y^n in at most $(|\mathcal{Y}||\mathcal{X}|n^{3/4} + \ell_n)$ places.[9] Thus, we can upper-bound its probability by

$$y^n \in \Gamma^{\ell_n}(\mathcal{S}_1^{(n)}) \Rightarrow \mathbb{P}_Y^n(y^n) \leq e^{-nH(P_Y) - (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}$$

where $\lambda = \min_{i,j} W_{Y|X}(j|i)$. Thus, we have

$$\begin{aligned} &|\Gamma^{\ell_n}(\mathcal{S}_0^{(n)}) \bigcap \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})| \\ &\qquad \geq (2\eta_n - 1)e^{nH(P_Y) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}. \end{aligned} \quad (6)$$

Note that for any $y^n \in \Gamma^{\ell_n}(\mathcal{S}_0^{(n)}) \bigcap \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})$, there exist a $\tilde{y}^n \in \mathsf{T}_W(i)$ for an i of the form $i = (0, M_2)$ which differs from y^n in at most $(|\mathcal{Y}||\mathcal{X}|n^{3/4} + \ell_n)$ places.[10] Consequently

$$\Pr[y^n | M = i] \geq e^{-nH(W_{Y|X}|P_X) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}. \quad (7)$$

Since $\mathcal{M}_2 = \frac{e^{nR^{(n)}}}{2}$, using (7) we can lower-bound the probability of y^n under the hypothesis $M_1 = 0$ as follows:

$$\begin{aligned} &\Pr[y^n | M_1 = 0] \\ &= \sum_{j\in\mathcal{M}_2} \Pr[y^n | M = (0,j)] \Pr[M = (0,j) | M_1 = 0] \\ &\geq 2e^{-n(H(W_{Y|X}|P_X) + R^{(n)}) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}. \end{aligned} \quad (8)$$

Clearly, the same holds for $M_1 = 1$ too, thus

$$\begin{aligned} &\Pr[y^n | M_1 = 1] \\ &\qquad \geq 2e^{-n(H(W_{Y|X}|P_X) + R^{(n)}) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}. \end{aligned} \quad (9)$$

Consequently, we get (10) at the bottom of the page, where (a) follows from (8) and (9) and (b) follows from (6).

Using Fano's inequality, we get

$$\mathcal{I}(M; Y^n) - nR^{(n)} \geq -\ln 2 - nR^{(n)}P_e^{(n)} \quad (11)$$

[9]Because of the integer constraints, T_{P_Y} might actually be an empty set. If so, we can make a similar argument for the U_Y^* which minimizes $\sum_j |U_Y(j) - P_Y(j)|$. However, this technicality is inconsequential.

[10]Integer constraints here are inconsequential as well.

$$\begin{aligned} \Pr\left[\hat{M}_1 \neq M_1\right] &\geq \sum_{y^n} \frac{1}{2}\min(\Pr[y^n | M_1 = 0], \Pr[y^n | M_1 = 1]) \\ &\overset{(a)}{\geq} \sum_{y^n \in \Gamma^{\ell_n}(\mathcal{S}_0^{(n)}) \bigcap \Gamma^{\ell_n}(\mathcal{S}_1^{(n)})} e^{-n(H(W_{Y|X}|P_X) + R^{(n)}) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda} \\ &\overset{(b)}{\geq} (2\eta_n - 1)e^{nH(P_Y) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda} e^{-n(H(W_{Y|X}|P_X) + R^{(n)}) + (|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda} \\ &= (2\eta_n - 1)e^{n(I(P_X,W) - R^{(n)}) + 2(|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda} \end{aligned} \quad (10)$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

where $\mathcal{I}(M;Y^n)$ is the mutual information between the message M and channel output Y^n. In addition, we can upper-bound $\mathcal{I}(M;Y^n)$ as follows:

$$\begin{aligned}
&\mathcal{I}(M;Y^n)\\
&= \sum_{i\in\mathcal{M},y^n\in\mathcal{Y}^n} \Pr[i,y^n]\ln\frac{\Pr[y^n|\,i]}{\Pr[y^n]}\\
&= \sum_{i\in\mathcal{M},y^n\in\mathcal{Y}^n} \Pr[i,y^n]\ln\frac{\Pr[y^n|\,i]}{\prod_{k=1}^n P_Y(y_k)}\\
&\quad - \sum_{y^n\in\mathcal{Y}^n}\Pr[y^n]\ln\frac{\Pr[y^n]}{\prod_{k=1}^n P_Y(y_k)}\\
&\overset{(a)}{\le} \sum_{i\in\mathcal{M}}\frac{1}{|\mathcal{M}|}\sum_{k=1}^n\sum_{y_k} W_{Y|X}(y_k|\bar{x}_k(i))\ln\frac{W_{Y|X}(y_k|\bar{x}_k(i))}{P_Y(y_k)}\\
&\overset{(b)}{=} nI(P_X,W) \qquad (12)
\end{aligned}$$

where $P_Y(\cdot)=\sum_{j\in\mathcal{X}}W_{Y|X}(\cdot)P_X(j)$. Step (a) follows the nonnegativity of KL divergence and step (b) follows from the fact that all the codewords are of type $P_X(\cdot)$.

Using (10), (11), and (12) we get

$$\Pr\left[\hat{M}_1\neq M_1\right] \ge (2\eta_n-1)e^{-\ln 2-nR^{(n)}P_e{}^{(n)}+2(|\mathcal{Y}||\mathcal{X}|n^{3/4}+\ell_n)\ln\lambda}.$$

Thus, using $\lim_{n\to\infty}P_e{}^{(n)}=0$, $\lim_{n\to\infty}\eta_n=1$, and $\lim_{n\to\infty}\frac{\ell_n}{n}=0$ we conclude that

$$\lim_{n\to\infty}\frac{-\ln\Pr{}^{(n)}\left[\hat{M}_1\neq M_1\right]}{n}=0.$$

Now the only thing left for proving $E_{\rm b}=0$ is to establish inequality (5). One can write the error probability of the nth code of $\mathcal{Q}$ as

$$\begin{aligned}
P_e^{(n)} &= \sum_{i\in\mathcal{M}^{(n)}}\frac{1}{\mathcal{M}}\sum_{y^n\in\mathcal{Y}^n}(1-\mathbb{I}_{\{\hat{M}(y^n)=i\}})\Pr[y^n|\,M=i]\\
&= \sum_{i\in\mathcal{M}^{(n)}} e^{-nR^{(n)}}\sum_V\sum_{y^n\in\mathsf{T}_V(i)}(1-\mathbb{I}_{\{\hat{M}(y^n)=i\}})\\
&\quad\cdot e^{-n(D(V_{Y|X}(\cdot|X)\|W_{Y|X}(\cdot|X)|P_X)+H(V_{Y|X}|P_X))}\\
&= \sum_V e^{-n(D(V_{Y|X}(\cdot|X)\|W_{Y|X}(\cdot|X)|P_X)+H(V_{Y|X}|P_X)+R^{(n)})}\\
&\quad\cdot\sum_{i\in\mathcal{M}^{(n)}}\sum_{y^n\in\mathsf{T}_V(i)}(1-\mathbb{I}_{\{\hat{M}(y^n)=i\}})\\
&= \sum_V e^{-n(D(V_{Y|X}(\cdot|X)\|W_{Y|X}(\cdot|X)|P_X)+H(V_{Y|X}|P_X)+R^{(n)})}\\
&\quad\cdot(Q_{0,V}+Q_{1,V}) \qquad (13)
\end{aligned}$$

where

$$Q_{k,V}=\sum_{\substack{i=(k,j)\\ j\in\mathcal{M}_2}}\sum_{y^n\in\mathsf{T}_V(i)}(1-\mathbb{I}_{\{\hat{M}(y^n)=i\}})$$

for $k=0,1$.

Note that $Q_{k,V}$ is the sum, over the messages i for which $M_1=k$, of the number of the elements in $\mathsf{T}_V(i)$ that are not decoded to message i. In a sense, it is a measure of the contribution of the V-shells of different codewords to the error probability. We will use (13) to establish lower bounds on $\mathbb{P}_Y^n(\mathcal{S}_{0,V}^{(n)})$'s.

Note that all elements of $\mathcal{S}_{0,V}^{(n)}$ have the same probability under $\mathbb{P}_Y^n(\cdot)$ and

$$\mathbb{P}_Y^n\left(\mathcal{S}_{0,V}^{(n)}\right)=|\mathcal{S}_{0,V}^{(n)}|e^{-\zeta n} \qquad (14)$$

where $\zeta=\sum_{x,y}P_X(x)V_{Y|X}(y|x)\ln\frac{1}{P_Y(y)}$.

Note that

$$\begin{aligned}
\zeta &= \sum_{x,y}P_X(x)V_{Y|X}(y|x)\ln\frac{1}{W_{Y|X}(y|x)}\\
&\quad+\sum_{x,y}P_X(x)V_{Y|X}(y|x)\ln\frac{W_{Y|X}(y|x)}{P_Y(y)}\\
&= D\left(V_{Y|X}(\cdot|X)\middle\|W_{Y|X}(\cdot|X)\middle|P_X\right)+H(V_{Y|X}|P_X)\\
&\quad+\sum_{x,y}P_X(x)(V_{Y|X}(y|x)-W_{Y|X}(y|x))\ln\frac{W_{Y|X}(y|x)}{P_Y(y)}\\
&\quad+I(P_X,W_{Y|X}).
\end{aligned}$$

Recall that $I(P_X,W_{Y|X})\le C$ and $\min_{i,j}W_{Y|X}(i|j)=\lambda$. Thus, using the definition of $[W_{Y|X}]$ given in (4) we get

$$\zeta\le C+\epsilon_n+D\left(V_{Y|X}(\cdot|X)\middle\|W_{Y|X}(\cdot|X)\middle|P_X\right)+H(V_{Y|X}|P_X) \qquad (15)$$

for all $V_{Y|X}\in[W_{Y|X}]$ where $\epsilon_n=\frac{|\mathcal{X}||\mathcal{Y}|}{\sqrt[4]{n}}\ln\frac{1}{\lambda}$.

Note that

$$\begin{aligned}
|\mathcal{S}_{0,V}^{(n)}| &= |\mathcal{M}_2^{(n)}|\cdot|\mathsf{T}_V(i)|-Q_{0,V}\\
&= \frac{1}{2}|\mathsf{T}_V(i)|e^{nR^{(n)}}-Q_{0,V}. \qquad (16)
\end{aligned}$$

Recalling that $\mathcal{S}_{0,V}^{(n)}$'s are disjoint and using (14), (15), and (16) we get

$$\begin{aligned}
&\mathbb{P}_Y^n\left(\mathcal{S}_0^{(n)}\right)\\
&\ge\sum_{V\in[W]}\mathbb{P}_Y^n\left(\mathcal{S}_{0,V}^{(n)}\right)\\
&\ge\sum_{V\in[W]}e^{-n(C+\epsilon_n)}\left(\frac{1}{2}|\mathsf{T}_V(i)|e^{nR^{(n)}}-Q_{0,V}\right)\\
&\quad e^{-n(D(V_{Y|X}(\cdot|X)\|W_{Y|X}(\cdot|X)|P_X)+H(V_{Y|X}|P_X))}\\
&\overset{(a)}{\ge} e^{n(R^{(n)}-(C+\epsilon_n))}\left(\frac{1}{2}\sum_{V\in[W]}\sum_{y^n\in\mathsf{T}_V(i)}\Pr[y^n|\,M=i]-P_e\right)\\
&\overset{(b)}{\ge} e^{n(R^{(n)}-(C+\epsilon_n))}\left(\frac{1}{2}-\frac{|\mathcal{X}||\mathcal{Y}|}{8\sqrt{n}}-P_e\right)
\end{aligned}$$

where (a) follows (13) and (b) follows from the Chebyshev inequality.[11] •

B. Proof of Theorem 2

1) Achievability: $E_{\rm md}\ge\tilde{\rm C}$:

Proof: For each block length n, the special message is sent with the length n repetition sequence $\bar{x}^n(1)=(x_r,x_r\ldots,x_r)$

[11]The claim in (b) is identical to the one in [13, Remark on page 34]

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

where x_r is the input letter satisfying

$$D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|x_r)\right) = \max_i D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|i)\right).$$

The remaining $|\mathcal{M}| - 1$ ordinary codewords are generated randomly and independently of each other using capacity-achieving input distribution P_X^* independent and identically distributed (i.i.d.) over time.

Let us denote the empirical distribution of a particular output sequence y^n by $\mathsf{Q}_{(y^n)}$. The receiver decodes to the special message only when the output distribution is not close to P_Y^*. Being more precise, the set of output sequences close to P_Y^*, $[P_Y^*]$, and decoding region of the special message, $\mathcal{G}(1)$, is given as follows:

$$[P_Y^*] = \{P_Y(\cdot) : ||P_Y(i) - P_Y^*(i)|| \le \sqrt[4]{1/n} \quad \forall i \in \mathcal{Y}\}$$
$$\mathcal{G}(1) = \{y^n : \mathsf{Q}_{(y^n)} \notin [P_Y^*]\}.$$

Since there are at most $(n+1)^{|\mathcal{Y}|}$ different empirical output distribution for elements of $\mathcal{Y}^n$ we get

$$\Pr{}^{(n)}\left[y^n \notin \mathcal{G}(1)|\, M = 1\right] \le (n+1)^{|\mathcal{Y}|} e^{-n \min_{Q_Y \in [P_Y^*]} D\left(Q_Y(\cdot)\|W_{Y|X}(\cdot|x_r)\right)}.$$

Thus

$$\lim_{n\to\infty} \frac{-\ln \Pr{}^{(n)}\left[y^n \notin \mathcal{G}(1)|\, M = 1\right]}{n} = D\left(P_Y^*(\cdot)\|\, W_{Y|X}(\cdot|x_r)\right) = \tilde{C}.$$

Now the only thing we are left with to prove is that we can have low enough probability for the remaining messages. For doing that we will first calculate the average error probability of the following random code ensemble.

Entries of the codebook, other than the ones corresponding to the special message, are generated independently using a capacity-achieving input distribution P_X^*. Because of the symmetry average error probability is the same for all $i \neq 1$ in $\mathcal{M}$. Let us calculate the error probability of the message $M = 2$.

Assuming that the second message was transmitted, $\Pr\left[y^n \in \mathcal{G}(1)|\, M = 2\right]$ is vanishingly small. It is because the output distribution for the random ensemble for ordinary codewords is i.i.d. P_Y^*. Chebyshev inequality guarantees that probability of the output type being outside a $\sqrt[4]{1/n}$ ball around P_Y^*, i.e., $[P_Y^*]$, is of the order $\sqrt{1/n}$.

Assuming that the second message was transmitted, $\Pr\left[y^n \in \cup_{i>2}\mathcal{G}(i)|\, M = 2\right]$ is vanishingly small due to the standard random coding argument for achieving capacity [35].

Thus, for any $P_e > 0$ for all large enough n, the average error probability of the code ensemble is smaller than P_e thus, we have at least one code with that P_e. For that code, at least half of the codewords have an error probability less then $2P_e$. •

2) Converse: $E_{\mathrm{md}} \le \tilde{C}$: In the Section VIII-D.2, we will prove that even with feedback and variable decoding time, the missed-detection exponent of a single special message is at most $\tilde{C}$. Thus, $E_{\mathrm{md}} \le \tilde{C}$.

C. Proof of Theorem 3

1) Achievability: $E_{\mathrm{md}} \ge E(r)$:

Proof:

Special Codewords: At any given block length n, we start with an optimum codebook (say $\mathcal{C}_{special}$) for $\lceil e^{nr} \rceil$ messages. Such optimum codebook achieves error exponent $E(r)$ for every message in it.

$$\Pr\left[\hat{M} \neq i | M = i\right] \doteq e^{-nE(r)}, \quad \forall i \in \mathcal{M}_s \equiv \{1, 2, \ldots, \lceil e^{nr} \rceil\}.$$

Since there are at most $(n+1)^{|\mathcal{X}|}$ different types, there is at least one type T_{P_X} which has $\frac{\lceil e^{nr} \rceil}{(1+n)^{|\mathcal{X}|}}$ or more codewords. Let us throw away all other codewords from $\mathcal{C}_{\mathrm{special}}$ and let us call the remaining fixed-composition codebook as $\mathcal{C}'_{\mathrm{special}}$. Codebook $\mathcal{C}'_{\mathrm{special}}$ is used for transmitting the special messages.

As shown in Fig. 3(a), let the noise ball around the codeword for the special message i be $\mathcal{B}_i$. These balls need not be disjoint. Let $\mathcal{B}$ denote the union of these balls of all special messages.

$$\mathcal{B} = \bigcup_{i \in \mathcal{M}_s} \mathcal{B}_i.$$

If the output sequence $y^n \in \mathcal{B}$, the first stage of the decoder decides a special message was transmitted. The second stage then chooses the ML candidate among the messages in $\mathcal{M}_s$.

Let us define $\mathcal{B}_i$ precisely now.

$$\mathcal{B}_i = \{y^n : \mathsf{V}(y^n, i) \in \mathcal{W}(r + \epsilon, P_X)\}$$

where

$$\mathcal{W}(r, P_X) = \left\{V_{Y|X} : \begin{array}{r} D\left(V_{Y|X}(\cdot|X)\middle\|\, W_{Y|X}(\cdot|X)\middle|\, P_X\right) \\ \le E_{\mathrm{sp}}(r; P_X) \end{array}\right\}.$$

Recall that the sphere-packing exponent for input type P_X at rate r, $E_{\mathrm{sp}}(r; P_X)$, is given by

$$E_{\mathrm{sp}}(r; P_X) = \min_{\substack{V_{Y|X}: \\ I(P_X, V_{Y|X}) \le r}} D\left(V_{Y|X}(\cdot|X)\middle\|\, W_{Y|X}(\cdot|X)\middle|\, P_X\right).$$

Ordinary Codewords: The ordinary codewords are generated randomly using a capacity-achieving input distribution P_X^*. This is the same as Shannon's construction for achieving capacity. The random coding construction provides a simple way to show that in the cavity $\mathcal{B}^c$ (complement of $\mathcal{B}$), we can essentially fit enough typical noise-balls to achieve capacity. This avoids the complicated task of carefully choosing the ordinary codewords and their decoding regions in the cavity, $\mathcal{B}^c$.

If the output sequence $y^n \in \mathcal{B}^c$, the first stage of the decoder decides an ordinary message was transmitted. The second stage then chooses the ML candidate from ordinary codewords.

Error Analysis: First, consider the case when a special codeword $\bar{x}^n(i)$ is transmitted. By Stein's lemma and definition of $\mathcal{B}_i$, the probability of $y^n \notin \mathcal{B}_i$ has exponent $E_{\mathrm{sp}}(r + \epsilon; P_X)$. Hence, the first-stage error exponent is at least $E_{\mathrm{sp}}(r + \epsilon; P_X)$.

Assuming correct first-stage decoding, the second-stage error exponent for special messages equals $E(r)$. Hence, the effective error exponent for special messages is

$$\min\{E(r), E_{\mathrm{sp}}(r + \epsilon; P_X)\}.$$

Since $E(r)$ is at most the sphere-packing exponent $E_{\mathrm{sp}}(r; P_X)$, [19], choosing arbitrarily small ϵ ensures that missed-detection exponent of each special message equals $E(r)$.

Now consider the situation of a uniformly chosen ordinary codeword being transmitted. We have to make sure that the error

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

probability is vanishingly small now. In this case, the output sequence distribution is i.i.d. P_Y^* for the random coding ensemble. The first-stage decoding error happens when $y^n \in \bigcup \mathcal{B}_i$. Again, by Stein's lemma, this exponent for any particular $\mathcal{B}_i$ equals $E_{\rm o}$

$$\begin{aligned} E_{\rm o} &= \min_{V_{Y|X} \in \mathcal{W}(r+\epsilon, P_X)} D\left(V_{Y|X}(\cdot|X) \middle\| P_Y^*(\cdot) \middle| P_X\right) \\ &\overset{(a)}{=} \min_{V_{Y|X} \in \mathcal{W}(r+\epsilon, P_X)} I(P_X, V_{Y|X}) + D\left((PV)_Y(\cdot) \| P_Y^*(\cdot)\right) \\ &\overset{(b)}{\geq} \min_{V_{Y|X} \in \mathcal{W}(r+\epsilon, P_X)} I(P_X, V_{Y|X}) \\ &\overset{(c)}{\geq} r + \epsilon \end{aligned}$$

where in $(PV)_Y$ in (a) is given by

$$(PV)_Y(j) = \sum_i P_X(i) V_{Y|X}(j|i),$$

(b) follows from the nonnegativity of the KL divergence, and (c) follows from the definition of sphere-packing exponent and $\mathcal{W}(r+\epsilon, P_X)$.

Applying union bound over the special messages, the probability of first-stage decoding error after sending an ordinary message is at most $\doteq \exp(nr - nE_{\rm o})$. We have already shown that $E_{\rm o} \geq r + \epsilon$, which ensures that probability of a first-stage decoding error for ordinary messages is at most $\doteq e^{-n\epsilon}$ for the random coding ensemble. Recall that for the random coding ensemble, the average error probability of the second-stage decoding also vanishes below capacity. To summarize, we have shown these two properties of the random coding ensemble.

1) Error probability of first-stage decoding vanishes as $a^{(n)} \doteq \exp(-n\epsilon)$ with n when a uniformly chosen ordinary message is transmitted.
2) Error probability of second-stage decoding (say $b^{(n)}$) vanishes with n when a uniformly chosen ordinary message is transmitted.

Since the first error probability is at most $4a^{(n)}$ for some $3/4$ fraction of codes in the random ensemble, and the second error probability is at most $4b^{(n)}$ for some $3/4$ fraction, there exists a particular code which satisfies both these properties. The overall error probability for ordinary messages is at most $4(a^{(n)}+b^{(n)})$, which vanishes with n. We will use this particular code for the ordinary codewords. This de-randomization completes our construction of a reliable code for ordinary messages to be combined with the code $\mathcal{C}_{\rm special}$ for special messages. •

2) Converse: $E_{\rm md} \leq E(r)$*:* The converse argument for this result is obvious. Removing the ordinary messages from the code can only improve the error probability of the special messages. Even then (by definition) the best missed detection exponent for the special messages equals $E(r)$.

D. Proof of Theorem 4

Let us now address the case with erasures. In this achievability result, the first stage of decoding remains unchanged from the no-erasure case.

Proof: We use essentially the same strategy as before. Let us start with a good code for $\lceil e^{nr} \rceil$ messages allowing erasure decoding. Forney had shown in [18] that, for symmetric channels, an error exponent equal to $E_{\rm sp}(r) + C - r$ is achievable while ensuring that erasure probability vanishes with n. We can use that code for these $\lceil e^{nr} \rceil$ codewords. As before, for $y^n \in \bigcup_i \mathcal{B}_i$, the first stage decides a special codeword was sent. Then the second stage applies the erasure decoding method in [18] amongst the special codewords.

With this decoding rule, when a special message is transmitted, error probability of the two-stage decoding is bottlenecked by the first stage: its error exponent $E_{\rm sp}(r+\epsilon)$ is smaller than that of the second stage $(E_{\rm sp}(r)+C-r)$. By choosing arbitrarily small ϵ, the special messages can achieve $E_{\rm sp}(r)$ as their missed-detection exponent.

The ordinary codewords are again generated i.i.d. P_X^*. If the first stage decides in favor of the ordinary messages, ML decoding is implemented among ordinary codewords. If an ordinary message was transmitted, we can ensure a vanishing error probability as before by repeating earlier arguments for no-erasure case. •

VIII. Variable Length Block Codes With Feedback: Proofs

In this section, we will present a more detailed discussion of bit-wise and message-wise UEP for variable-length block codes with feedback by proving Theorems 5–9. In the proofs of converse results, we need to discuss issues related to the conditional entropy of the messages given the observation of the receiver. In those discussions, we use the following notation for conditional entropy and conditional mutual information:

$$\mathcal{H}(M|Y^n) = \sum_{i \in \mathcal{M}} \Pr[M = i | Y^n] \ln \frac{1}{\Pr[M = i | Y^n]}$$

$$\mathcal{I}(M; Y_{n+1} | Y^n) = \mathcal{H}(M|Y^n) - E\left[\mathcal{H}(M|Y^{n+1}) \middle| Y^n\right].$$

It is worth noting that this notation is different from the one widely used, which includes a further expectation over the conditioned variable. "$H(M|Y^n)$" in the conventional notation, stands for the $E[\mathcal{H}(M|Y^n)]$ and "$H(M|Y^n = y^n)$" stands for $\mathcal{H}(M|Y^n)$.

A. Proof of Theorem 5

1) Achievability: $E_b^f \geq \tilde{\rm C}$*:* This single special bit exponent is achieved using the missed detection exponent of a single special message, indicating a decoding error for the special bit. The decoding error for the bit goes unnoticed when this special message is not detected. This shows how feedback connects bit-wise UEP to message-wise UEP in a fundamental manner.

Proof: We will prove that $E_{\rm b}^f \geq \tilde{C}$ by constructing a capacity-achieving sequence with feedback $\mathcal{Q}$, such that $E_{{\rm b},\mathcal{Q}}^f = \tilde{C}$. For that, let $\mathcal{Q}'$ be a capacity-achieving sequence such that $E_{{\rm md},\mathcal{Q}'} = \tilde{C}$. Note that existence of such a $\mathcal{Q}'$ is guaranteed as a result of Theorem 2. We first construct a two-phase fixed-length block code with feedback and erasures. Then using this we obtain the kth element of $\mathcal{Q}$.

In the first phase, one of the two input symbols, x_0 and x_1, with distinct output distributions[12] is send for $\lceil \sqrt{k} \rceil$ time units

[12]Two input symbols x_0 and x_1 are such that $W(\cdot|x_1) \neq W(\cdot|x_0)$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

depending on M_1. At time $\lceil\sqrt{k}\rceil$ receiver makes tentative decision $\tilde{M}_1$ on message M_1. Using Chernoff bound it can easily be shown that, [36, Theorem 5]

$$\Pr\left[\tilde{M}_1 \neq M_1\right] \leq e^{-\mu\sqrt{k}}, \quad \text{where } \mu > 0.$$

Actual value of μ, however, is immaterial to us; we are merely interested in finding an upper bound on $\Pr\left[\tilde{M}_1 \neq M_1\right]$ which goes to zero as k increases.

In the second phase, the transmitter uses the kth member of $\mathcal{Q}'$. The message in the second phase, M', is determined by M_2 depending on whether M_1 is decoded correctly or not at the end of the first phase.

$$\tilde{M}_1 \neq M_1 \Rightarrow \mathrm{M}' = 1$$
$$\tilde{M}_1 = M_1 \text{ and } M_2 = i \Rightarrow \mathrm{M}' = i+1, \quad \forall i.$$

At the end of the second phase, the decoder decodes M' using the decoder of $\mathcal{Q}'$. If the decoded message is one, i.e., $\hat{\mathrm{M}}' = 1$, then the receiver declares an erasure, else, $\hat{M}_1 = \tilde{M}_1$ and $\hat{M}_2 = \hat{\mathrm{M}}' - 1$.

Note that erasure probability of the two-phase fixed-length block code is upper-bounded as

$$\begin{aligned}\Pr\left[\hat{\mathrm{M}}' = 1\right] &\leq \Pr\left[\tilde{M}_1 \neq M_1\right] + \Pr\left[\mathrm{M}' = 1 | \, \mathrm{M}' \neq 1\right] \\ &\leq e^{-\mu\sqrt{k}} + \frac{\mathcal{M}'^{(k)}}{\mathcal{M}'^{(k)} - 1} P_e{}'^{(k)} \end{aligned} \tag{17}$$

where $P_e{}'^{(k)}$ is the error probability of the kth member of $\mathcal{Q}'$.

Similarly, we can upper-bound the probabilities of two error events associated with the two-phase fixed-length block code as follows:

$$\Pr\left[\hat{M}_1 \neq M_1 \, , \, \hat{\mathrm{M}}' \neq 1\right] \leq P_e{}'^{(k)}(1) \tag{18}$$

$$\Pr\left[\hat{M} \neq M \, , \, \hat{\mathrm{M}}' \neq 1\right] \leq \frac{\mathcal{M}'^{(k)}}{\mathcal{M}'^{(k)} - 1} P_e{}'^{(k)} + P_e{}'^{(k)}(1) \tag{19}$$

where $P_e{}'^{(k)}(1)$ is the conditional error probability of the first message in the kth element of $\mathcal{Q}'$.

If there is an erasure, the transmitter and the receiver will repeat what they have done again, until they get $\hat{\mathrm{M}}' \neq 1$. If we sum the probabilities of all the error events, including error events in the possible repetitions we get

$$\Pr\left[\hat{M}_1 \neq M_1\right] = \frac{\Pr\left[\hat{M}_1 \neq M_1 \, , \, \hat{\mathrm{M}}' \neq 1\right]}{1 - \Pr\left[\hat{\mathrm{M}}' = 1\right]} \tag{20}$$

$$\Pr\left[\hat{M} \neq M\right] = \frac{\Pr\left[\hat{M} \neq M \, , \, \hat{\mathrm{M}}' \neq 1\right]}{1 - \Pr\left[\hat{\mathrm{M}}' = 1\right]}. \tag{21}$$

Note that the expected decoding time of the code is

$$E[\tau] = \frac{k + \lceil\sqrt{k}\rceil}{1 - \Pr\left[\hat{\mathrm{M}}' = 1\right]}. \tag{22}$$

Using (17), (18), (19), (20), (21), and (22) one can conclude that the resulting sequence of variable-length block codes with feedback, $\mathcal{Q}$, is reliable. Furthermore, $R_{\mathcal{Q}} = C$ and $E^f_{\mathrm{b},\mathcal{Q}} = \tilde{C}$. •

2) Converse: $E_b^{\mathrm{F}} \leq \tilde{\mathrm{C}}$: We will use a converse result we have not proved yet, namely, the converse part of Theorem 8, i.e., $E^f_{\mathrm{md}} \leq \tilde{C}$.

Proof: Consider a capacity-achieving sequence, $\mathcal{Q}$, with message set sequence $\mathcal{M}^{(k)} = \{0,1\} \times \mathcal{M}_2^{(k)}$. Using $\mathcal{Q}$, we construct another capacity-achieving sequence $\mathcal{Q}'$ with a special message 0, with message set sequence $\mathcal{M}'^{(k)} = \{0\} \cup \mathcal{M}_2^{(k)}$ such that $E^f_{\mathrm{md},\mathcal{Q}'} = E^f_{\mathrm{b},\mathcal{Q}}$. This implies $E^f_{\mathrm{b}} \leq E^f_{\mathrm{md}}$, which, together with Theorem 8, $E^f_{\mathrm{md}} \leq \tilde{C}$, gives us $E^f_{\mathrm{b}} \leq \tilde{C}$.

Let us denote the message of $\mathcal{Q}$ by M and that of $\mathcal{Q}'$ by M'. The kth code of $\mathcal{Q}'$ is as follows. At time 0, the receiver chooses randomly an M_1 for the kth element of $\mathcal{Q}$ and sends its choice through a feedback channel to the transmitter. If the message of $\mathcal{Q}'$ is not 0, i.e., $\mathrm{M}' \neq 0$, then the transmitter uses the codeword for $M = (M_1, \mathrm{M}')$ to convey M'. If $\mathrm{M}' = 0$, the receiver picks an M_2 with uniform distribution on $\mathcal{M}_2$ and uses the codeword for $M = (1 - M_1, M_2)$ to convey that $\mathrm{M}' = 0$.

The receiver makes decoding using the decoder of $\mathcal{Q}$: if $\hat{M} = (M_1, i)$ then $\hat{\mathrm{M}}' = i$, if $\hat{M} = (1 - M_1, i)$ then $\hat{\mathrm{M}}' = 0$. One can easily show that the expected decoding time and error probability of both of the codes are same. Furthermore, the error probability of M_1 in $\mathcal{Q}$ is equal to the conditional error probability of message $\mathrm{M}' = 0$ in $\mathcal{Q}'$ thus, $E^f_{\mathrm{md},\mathcal{Q}'} = E^f_{\mathrm{b},\mathcal{Q}}$. •

B. Proof of Theorem 6

1) Achievability: $E^f_{\mathrm{bits}}(r) \geq \left(1 - \frac{r}{\mathrm{C}}\right)\tilde{\mathrm{C}}$:

Proof: We will construct the capacity-achieving sequence with feedback $\mathcal{Q}$ using a capacity-achieving sequence $\mathcal{Q}'$ satisfying $E_{\mathrm{md},\mathcal{Q}'} = \tilde{C}$, as we did in the proof of Theorem 5. We know that such a sequence exists, because of Theorem 8.

For the kth member of $\mathcal{Q}$, consider the following two-phase errors and erasures code. In the first phase, the transmitter uses the $\lfloor rk \rfloor$th element of $\mathcal{Q}'$ to convey M_1. The receiver makes a tentative decision $\tilde{M}_1$. In the second phase, the transmitter uses the $\lfloor (C - r)k \rfloor$th element of $\mathcal{Q}'$ to convey M_2 and whether $\tilde{M}_1 = M_1$ or not, with a mapping similar to the one we had in the proof of Theorem 5.

$$\tilde{M}_1 \neq M_1 \Rightarrow \mathrm{M}' = 1$$
$$\tilde{M}_1 = M_1 \text{ and } M_2 = i \Rightarrow \mathrm{M}' = i + 1 \quad \forall i.$$

Thus, $\mathcal{M}_1^{(k)} = \mathcal{M}'^{(\lfloor rk \rfloor)}$ and $\mathcal{M}_2^{(k)} \cup \{|\mathcal{M}_2^{(k)}| + 1\} = \mathcal{M}'^{(\lfloor (C-r)k \rfloor)}$. If we apply a decoding algorithm, like the one we had in the proof of Theorem 5 ; going through essentially the same analysis with proof of Theorem 5, we can conclude that $\mathcal{Q}$ is a capacity-achieving sequence and $E^f_{\mathrm{bits},\mathcal{Q}} = \left(1 - \frac{r}{C}\right)\tilde{C}$ and $r_{\mathcal{Q}} = r$. •

2) Converse: $E^f_{\mathrm{bits}}(r) \leq \left(1 - \frac{r}{\mathrm{C}}\right)\tilde{\mathrm{C}}$: In establishing the converse we will use a technique that was used previously in [4], together with Lemma 1 which we will prove in the converse part Theorem 8.

Proof: Consider any variable-length block code with feedback whose message set $\mathcal{M}$ is of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

Let t_δ be the first time instance that an $i \in \mathcal{M}_1$ becomes more likely than $(1-\delta)$ and let $\tau_\delta = t_\delta \wedge \tau$.

Recall that $\min_{i,j} W_{Y|X}(j|i) = \lambda$, as a result, the definition of τ_δ implies that $\min_{i\in\mathcal{M}_1}(1 - \Pr[M_1 = i|\, y^{\tau_\delta}]) \geq \lambda\delta$. Thus, using Markov inequality for P_e we get

$$\Pr[\tau_\delta = \tau] \leq \frac{P_e}{\lambda\delta}. \tag{23}$$

We use (23) to bound the expected value of the entropy of the first part of the message at time τ_δ as follows:

$$\begin{aligned} &E[\mathcal{H}(M_1|Y^{\tau_\delta})] \\ &\quad = E\left[\mathcal{H}(M_1|Y^{\tau_\delta})\mathbb{I}_{\{\tau_\delta=\tau\}}\right] + E\left[\mathcal{H}(M_1|Y^{\tau_\delta})\mathbb{I}_{\{\tau_\delta<\tau\}}\right] \\ &\quad \leq \frac{P_e}{\lambda\delta}\ln|\mathcal{M}_1| + (\ln 2 + \delta\ln|\mathcal{M}_1|) \\ &\quad = \ln 2 + (\frac{P_e}{\lambda\delta} + \delta)\ln|\mathcal{M}_1|). \end{aligned}$$

It has already been established in [4] that

$$\frac{E[\mathcal{H}(M) - \mathcal{H}(M|Y^{\tau_\delta})]}{E[\tau_\delta]} \leq C. \tag{24}$$

Thus

$$\begin{aligned} E[\tau_\delta] &\geq \frac{1}{C}(E[\mathcal{H}(M) - \mathcal{H}(M_1|Y^{\tau_\delta}) - \mathcal{H}(M_2|M_1, Y^{\tau_\delta})]) \\ &\geq \frac{1}{C}(-\ln 2 + (1 - \frac{P_e}{\lambda\delta} - \delta)\ln|\mathcal{M}_1|). \end{aligned} \tag{25}$$

The bound given in inequality (25) specifies the time needed for getting a likely candidate, $\tilde{M}_1$. As in the case in [4], the remaining time is the time spent waiting for confirmation. But unlike in [4], the transmitter needs to convey also M_2 during that time.

For each realization of Y^{τ_δ}, divide the message set into disjoint subsets $\Theta_0, \Theta_1, \ldots, \Theta_{|\mathcal{M}_2|}$ as follows:

$$\begin{aligned} \Theta_0 &= \{l : l \in \mathcal{M}, l = (i,j) \ \text{ where } i \neq \tilde{\mathcal{M}}_1(Y^{\tau_\delta})\} \\ \Theta_j &= \{l : l \in \mathcal{M}, l = (\tilde{\mathcal{M}}_1(Y^{\tau_\delta}), j)\} \quad \forall j \in \{1, 2, \ldots |\mathcal{M}_2|\} \end{aligned}$$

where $\tilde{M}_1(Y^{\tau_\delta})$ is the most likely message given Y^{τ_δ}. Furthermore, let the auxiliary message, M', be the index of the set that M belongs to, i.e., $M \in \Theta_{\mathrm{M}'}$.

The decoder for the auxiliary message decodes the index of the decoded message at the decoding time τ, i.e.,

$$\hat{\mathrm{M}}'(Y^\tau) = j \Leftrightarrow \hat{M}(Y^\tau) \in \Theta_j.$$

With these definition we have

$$\Pr\left[\hat{M}(Y^\tau) \neq M \middle| Y^{\tau_\delta}\right] \geq \Pr\left[\hat{\mathrm{M}}'(Y^\tau) \neq \mathrm{M}' \middle| Y^{\tau_\delta}\right]$$

$$\Pr\left[\hat{M}_1(Y^\tau) \neq M_1 \middle| Y^{\tau_\delta}\right] \geq \Pr\left[\hat{\mathrm{M}}'(Y^\tau) \neq 0 \middle| Y^{\tau_\delta}, \mathrm{M}' = 0\right] \Pr[\mathrm{M}' = 0|\, Y^{\tau_\delta}].$$

Now, we apply Lemma 1, which will be proved in Section VIII-D.2. To ease the notation we use following shorthand notation:

$$\begin{aligned} P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\} &= \Pr\left[\hat{\mathrm{M}}'(Y^\tau) \neq \mathrm{M}' \middle| Y^{\tau_\delta}\right] \\ P_e^{\mathrm{M}'}\{0, Y^{\tau_\delta}\} &= \Pr\left[\hat{\mathrm{M}}'(Y^\tau) \neq 0 \middle| Y^{\tau_\delta}, \mathrm{M}' = 0\right] \\ \xi(Y^{\tau_\delta}) &= \Pr[\mathrm{M}'(Y^{\tau_\delta}) = 0|\, Y^{\tau_\delta}]. \end{aligned}$$

As a result of Lemma 1, for each realization of $y^{\tau_\delta} \in \mathcal{Y}^{\tau_\delta}$ such that $\tau_\delta < \tau$, we have the expression at the bottom of the page. By multiplying both sides of the inequality with $\mathbb{I}_{\{\tau_\delta<\tau\}}$, we get an expression that holds for all Y^{τ_δ}

$$\begin{aligned} &\mathbb{I}_{\{\tau_\delta<\tau\}}(1 - \xi(Y^{\tau_\delta}) - P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\ln\frac{1}{P_e^{\mathrm{M}'}\{0, Y^{\tau_\delta}\}} \\ &\quad \leq \mathbb{I}_{\{\tau_\delta<\tau\}}\ln 2 + \mathbb{I}_{\{\tau_\delta<\tau\}}E[\tau - \tau_\delta|\, Y^{\tau_\delta}] \\ &\qquad \mathcal{J}\left(\frac{\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta}) - \ln 2 - P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\ln|\mathcal{M}_2|}{E[\tau - \tau_\delta|\, Y^{\tau_\delta}]}\right). \end{aligned} \tag{26}$$

Now we take the expectation of both sides over Y^{τ_δ}. For the right-hand side, we get the identity in (27) given at the top of the following page, where (a) follows the concavity of $\mathcal{J}(\cdot)$ and Jensen's inequality when we interpret $\frac{E[\tau-\tau_\delta|Y^{\tau_\delta}]\mathbb{I}_{\{\tau_\delta<\tau\}}}{E[\tau-\tau_\delta]}$ as probability distribution over $\mathcal{Y}^{\tau_\delta}$ and (b) follows the fact that $\mathcal{J}(\cdot)$ is a decreasing function.

Now we lower-bound $E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})\right]$ in terms of $E[\mathcal{H}(M|Y^{\tau_\delta})]$. Note that

$$\begin{aligned} &\mathcal{H}(M|Y^{\tau_\delta}) \\ &\quad = \Pr\left[M_1 \neq \tilde{M}_1(Y^{\tau_\delta}) \middle| Y^{\tau_\delta}\right]\mathcal{H}(M|M_1 \neq \tilde{M}_1(Y^{\tau_\delta}), Y^{\tau_\delta}) \\ &\qquad + \mathcal{H}(\mathrm{M}'|Y^{\tau_\delta}) \\ &\quad \leq \Pr\left[M_1 \neq \tilde{M}_1(Y^{\tau_\delta}) \middle| Y^{\tau_\delta}\right]\ln|\mathcal{M}_1||\mathcal{M}_2| + \mathcal{H}(\mathrm{M}'|Y^{\tau_\delta}). \end{aligned}$$

Furthermore, for all Y^{τ_δ} such that $\tau > \tau_\delta$

$$\Pr\left[\tilde{M}_1(Y^{\tau_\delta}) \neq M_1 \middle| Y^{\tau_\delta}\right] \leq \delta.$$

Thus

$$\begin{aligned} &E[\mathbb{I}_{\{\tau_\delta<\tau\}}\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})] \\ &\geq E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(\mathcal{H}(M|Y^{\tau_\delta}) - \delta\ln|\mathcal{M}_1||\mathcal{M}_2|)\right] \\ &= E\left[(1 - \mathbb{I}_{\{\tau_\delta=\tau\}})\mathcal{H}(M|Y^{\tau_\delta})\right] - \delta\ln|\mathcal{M}_1||\mathcal{M}_2| \\ &\geq E[\mathcal{H}(M|Y^{\tau_\delta})] - (\Pr[\tau_\delta = \tau] + \delta)\ln|\mathcal{M}_1||\mathcal{M}_2| \\ &\overset{(a)}{\geq} E[\mathcal{H}(M|Y^{\tau_\delta})] - \left(\frac{P_e}{\lambda\delta} + \delta\right)\ln|\mathcal{M}_1||\mathcal{M}_2| \end{aligned}$$

$$(1 - \xi(Y^{\tau_\delta}) - P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\ln\frac{1}{P_e^{\mathrm{M}'}\{0, Y^{\tau_\delta}\}} \leq \ln 2 + E[\tau - \tau_\delta|\, Y^{\tau_\delta}]\,\mathcal{J}\left(\frac{\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta}) - \ln 2 - P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\ln|\mathcal{M}_2|}{E[\tau - \tau_\delta|\, Y^{\tau_\delta}]}\right).$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

$$\begin{aligned} RHS &= E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\ln 2\right] + E\left[\mathbb{I}_{\{\tau_\delta<\tau\}} E\left[\tau-\tau_\delta|Y^{\tau_\delta}\right]\mathcal{J}\left(\frac{\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})-\ln 2-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\ln|\mathcal{M}_2|}{E\left[\tau-\tau_\delta|Y^{\tau_\delta}\right]}\right)\right] \\ &\le \ln 2 + E\left[E\left[\tau-\tau_\delta|Y^{\tau_\delta}\right]\mathcal{J}\left(\frac{\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})-\ln 2-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\ln|\mathcal{M}_2|}{E\left[\tau-\tau_\delta|Y^{\tau_\delta}\right]}\right)\mathbb{I}_{\{\tau_\delta<\tau\}}\right] \\ &\overset{(a)}{\le} \ln 2 + E\left[\tau-\tau_\delta\right]\mathcal{J}\left(E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\frac{\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})-\ln 2-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\ln|\mathcal{M}_2|}{E\left[\tau-\tau_\delta\right]}\right]\right) \\ &\overset{(b)}{\le} \ln 2 + E\left[\tau-\tau_\delta\right]\mathcal{J}\left(\frac{E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\mathcal{H}(\mathrm{M}'|Y^{\tau_\delta})\right]-\ln 2-P_e\ln|\mathcal{M}_2|}{E\left[\tau-\tau_\delta\right]}\right) \end{aligned} \tag{27}$$

$$\overset{(b)}{\ge} (1-\frac{P_e}{\lambda\delta}-\delta)\ln|\mathcal{M}_1||\mathcal{M}_2| - CE\left[\tau_\delta\right] \tag{28}$$

where (a) follows from the inequality (23), (b) follows from the inequality (24). Since $\mathcal{J}(\cdot)$ is decreasing in its argument, inserting (28) in (27) we get

$$\text{RHS} \le \ln 2 + E\left[\tau-\tau_\delta\right] \mathcal{J}\left(\frac{\ln|\mathcal{M}_1||\mathcal{M}_2|\left(1-\frac{P_e}{\lambda\delta}-\delta-P_e\right)-E\left[\tau_\delta\right]C-\ln 2}{E\left[\tau-\tau_\delta\right]}\right). \tag{29}$$

Note that $\forall a>0, b>0, C>0$

$$\begin{aligned} &\frac{d}{dx}(b-x)\mathcal{J}\left(\frac{a-Cx}{b-x}\right)|_{x=x_0} \\ &= -\mathcal{J}\left(\frac{a-Cx_0}{b-x_0}\right) - \left(C-\frac{a-Cx_0}{b-x_0}\right)\frac{d}{dx}\mathcal{J}(x)|_{x=\frac{a-Cx_0}{b-x_0}} \\ &\overset{(a)}{\le} -\mathcal{J}(C) \end{aligned}$$

where (a) follows the concavity of $\mathcal{J}(\cdot)$. Thus, the upper bound given in (29) is decreasing in $E\left[\tau_\delta\right]$. Hence, using the lower bound on $E\left[\tau_\delta\right]$, given in (23), we get (30) at the bottom of the page. Now we consider the left-hand side (LHS) and take the expectation of the inequality given in (26) to get (31) also shown at the bottom of the page, where (a) follows log-sum inequality and (b) follows from the fact that $x\ln x \ge -e^{-1}$.

Note that

$$\begin{aligned} &E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\right] \\ &\quad\ge E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta}))\right] - E\left[P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\}\right] \\ &\quad\ge E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\right](1-\delta) - P_e \\ &\quad\ge 1-\frac{P_e}{\lambda\delta}-\delta \end{aligned} \tag{32}$$

where in last step we have used (23). Furthermore, we get (33) at the top of the following page. Thus, using (31), (32), and (33) we get

$$LHS \ge -e^{-1} - (1-\frac{P_e}{\lambda\delta}-\delta)\ln\frac{{P_e}^{M_1}}{\lambda\delta}. \tag{34}$$

Using the inequalities (30), (34) and choosing $\delta=\sqrt{P_e}$ we get $E^f_{\text{bits},\mathcal{Q}} \le \left(1-\frac{r_{\mathcal{Q}}}{C}\right)\mathcal{J}(C)$. Since $\mathcal{J}(C)=\tilde{C}$, this implies $E^f_{\text{bits}}(r) \le \left(1-\frac{r}{C}\right)\tilde{C}$. •

C. Proof of Theorem 7

1) Achievability:

Proof: The proof is very similar to the achievability proof for Theorem 6. Choose a capacity-achieving sequence $\mathcal{Q}'$ such that $E^f_{\text{b},\mathcal{Q}'}=\tilde{C}$. The capacity-achieving sequence with feedback $\mathcal{Q}$ uses L elements of $\mathcal{Q}'$ as follows.

For the kth element of code $\mathcal{Q}$, the transmitter uses the $\lfloor k\cdot r_1\rfloor$th element of $\mathcal{Q}'$ to send the first part of the message, M_1.

$$\text{RHS} \le \ln 2 + \left(E\left[\tau\right] - (1-\delta-\frac{P_e}{\lambda\delta})\frac{\ln|\mathcal{M}_1|}{C} + \frac{\ln 2}{C}\right)\mathcal{J}\left(\frac{\left(1-\frac{P_e}{\lambda\delta}-\delta-P_e\right)\ln|\mathcal{M}_2| - P_e\ln|\mathcal{M}_1| - 2\ln 2}{E\left[\tau\right] - (1-\delta-\frac{P_e}{\lambda\delta})\frac{\ln|\mathcal{M}_1|}{C} + \frac{\ln 2}{C}}\right). \tag{30}$$

$$\begin{aligned} \text{LHS} &= E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\ln\frac{1}{P_e^{\mathrm{M}'}\{0,Y^{\tau_\delta}\}}\right] \\ &\overset{(a)}{\ge} E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\right]\ln\frac{E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\right]}{E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})P_e^{\mathrm{M}'}\{0,Y^{\tau_\delta}\}\right]} \\ &\overset{(b)}{\ge} -e^{-1} + E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\right]\ln\frac{1}{E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\}P_e^{\mathrm{M}'}\{0,Y^{\tau_\delta}\}\right]} \\ &\ge -e^{-1} + E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}(1-\xi(Y^{\tau_\delta})-P_e^{\mathrm{M}'}\{Y^{\tau_\delta}\})\right]\ln\frac{1}{E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}P_e^{\mathrm{M}'}\{0,Y^{\tau_\delta}\}\right]} \end{aligned} \tag{31}$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

$$
\begin{aligned}
&E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}P_e^{\mathrm{M}'}\{0,Y^{\tau_\delta}\}\right]\\
&=E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\Pr\left[\hat{M}_1=\tilde{M}_1\middle|Y^{\tau_\delta},\tilde{M}_1\neq M_1\right]\right]\\
&\le E\left[\mathbb{I}_{\{\tau_\delta<\tau\}}\Pr\left[\hat{M}_1=\tilde{M}_1\middle|Y^{\tau_\delta},\tilde{M}_1\neq M_1\right]\frac{\Pr\left[\tilde{M}_1\neq M_1\middle|Y^{\tau_\delta}\right]}{\delta\lambda}\right]\\
&\le\frac{{P_e}^{M_1}}{\delta\lambda}.
\end{aligned}
\tag{33}
$$

In the remaining phases, $l\ge 2$, the transmitter uses $\lfloor k\cdot r_l\rfloor$th element of $\mathcal{Q}'$. The special message of the code for phase l is allocated to the error event in previous phases.

$$
\begin{aligned}
&(\tilde{M}_1,\ldots,\tilde{M}_{(l-1)})\neq(M_1,\ldots,M_{(l-1)})\Rightarrow \mathrm{M}'_l=1,\quad\forall l\\
&(\tilde{M}_1,\ldots,\tilde{M}_{(l-1)})=(M_1,\ldots,M_{(l-1)})\Rightarrow \mathrm{M}'_l=M_l+1,\quad\forall l.
\end{aligned}
$$

Thus, $\mathcal{M}_1^{(k)}=\mathcal{M}'^{(\lfloor rk\rfloor)}$ and for all $l\ge 1$ $\mathcal{M}_l^{(k)}\cup\{|\mathcal{M}_l^{(k)}|+1\}=\mathcal{M}'^{(\lfloor r_l k\rfloor)}$. If for all $l\in\{2,3,\ldots,L\}$, $\hat{\mathrm{M}}'_l\neq 1$, the receiver decodes all parts of the information, else, it declares an erasure. We skip the error analysis because it is essentially the same as that for Theorem 6. •

2) Converse:

Proof: We prove the converse of Theorem 7 by contradiction. Evidently, for all $j\in\{1,2,\ldots L\}$

$$
\begin{aligned}
&{P_e}^{M_1,M_2,\ldots,M_j}\ge\max\{{P_e}^{M_1},{P_e}^{M_2},\ldots,{P_e}^{M_j}\}\\
&{P_e}^{M_1,M_2,\ldots,M_j}\le{P_e}^{M_1}+{P_e}^{M_2}+\cdots+{P_e}^{M_j}.
\end{aligned}
$$

Thus, if there exists a scheme that can reach an error exponent vector outside the region given in Theorem 7, there is at least one E_i such that $E_i\ge(1-\frac{\sum_{j=1}^{i}r_j}{C})\tilde{C}$. Then we can have two super messages as follows:

$$
\mathrm{M}'_1=(M_1,M_2,\ldots,M_i)\ \text{ and }\ \mathrm{M}'_2=(M_{i+1},M_{i+2},\ldots,M_l).
$$

Recall that ${P_e}^{M_1}\le{P_e}^{M_2}\le\cdots\le{P_e}^{M_l}$. Thus, this new code is a capacity-achieving code, whose special bits have rate $r_{\mathcal{Q}'}$ and $E^f_{\mathrm{bits},\mathcal{Q}'}>E^f_{\mathrm{bits}}(r_{\mathcal{Q}'})$. This contradicts Theorem 6 which we have already proved. Thus, all the achievable error exponent regions should lie in the region given in Theorem 7. •

D. Proof of Theorem 8

1) Achievability: $E^f_{\mathrm{md}}\ge\tilde{\mathrm{C}}$: Note that any fixed-length block code without feedback is also a variable-length block code with feedback, thus, $E^f_{\mathrm{md}}\ge E_{\mathrm{md}}$. Using the capacity-achieving sequence, we have used in the achievability proof of Theorem 2, we get $E^f_{\mathrm{md}}\ge\tilde{C}$.

2) Converse: $E^f_{\mathrm{md}}\le\tilde{\mathrm{C}}$: Now we prove that even with feedback and variable decoding time, the best missed detection exponent of a single special message is less then or equal to $\tilde{C}$, i.e., $E^f_{\mathrm{md}}\le\tilde{C}$. Since the set of capacity-achieving sequences is a subset of capacity-achieving sequences with feedback and variable decoding time, this also implies that $E_{\mathrm{md}}\le\tilde{C}$.

Instead of directly proving the converse part of Theorem 8 we first prove Lemma 1 as follows.

Lemma 1: For any variable-length block code with feedback, a message set $\mathcal{M}$, initial entropy $\mathcal{H}(M)$, and average error probability P_e, the conditional error probability of each message is lower-bounded as follows:

$$
\begin{aligned}
&\Pr\left[\hat{M}\neq i\middle|M=i\right]\\
&\ge e^{-\frac{1}{1-\Pr[M=i]-P_e}\left(\mathcal{J}\left(\frac{\mathcal{H}(M)-h(P_e)-P_e\ln|\mathcal{M}|}{E[\tau]}\right)E[\tau]+\ln 2\right)},\ \forall i
\end{aligned}
\tag{35}
$$

where $\mathcal{J}(R)$ is given by the following optimization over probability distributions on $\mathcal{X}$ given in (36) at the bottom of the page. It is worthwhile remembering the notation we introduced previously that

$$
\begin{aligned}
(P^iW)_Y(\cdot)&=\sum_{j\in\mathcal{X}}P_X^i(j)W_{Y|X}(\cdot|j)\\
I(P_X^i,W_{Y|X})&=\sum_{j\in\mathcal{X},k\in\mathcal{Y}}P_X^i(j)W_{Y|X}(k|i)\ln\frac{W_{Y|X}(k|i)}{(P^iW)_Y(k)}.
\end{aligned}
$$

The first thing to note about Lemma 1 is that it is not necessarily for the case of uniform probability distribution on the message set $\mathcal{M}$. Furthermore, as long as $\Pr[M=i]\ll 1$, the lower bound on $\Pr\left[\hat{M}\neq i\middle|M=i\right]$ depends on the *a priori* probability distribution of the messages only through the entropy of it, $\mathcal{H}(M)$.

In (36), α is simply a time-sharing variable, which allows us to use an (x_i,P_X^i) pair with low mutual information and high divergence together with another (x_i,P_X^i) pair with high mutual information and low divergence. As a result of Carathéodory's theorem, we see that time sharing between two points of the form (x_i,P_X^i) is sufficient for obtaining optimal performance, i.e., allowing time sharing between more than two points of the form (x_i,P_X^i) will not improve the value of $\mathcal{J}(R)$.

Indeed, for any $R\in[0,C]$, one can use the optimizing values of α, x_1, x_2, P_X^1, and P_X^2 in a scheme like the one in The-

$$
\mathcal{J}(R)=\max_{\substack{\alpha,x_1,x_2,P_X^1,P_X^2:\\ \alpha I(P_X^1,W_{Y|X})+(1-\alpha)I(P_X^2,W_{Y|X})\ge R}}\alpha D\left((P^1W)_Y(\cdot)\middle\|W(\cdot|x_1)\right)+(1-\alpha)D\left((P^2W)_Y(\cdot)\middle\|W(\cdot|x_2)\right).
\tag{36}
$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

orem 2 with time sharing and prove that the missed detection exponent of $\mathcal{J}(R)$ is achievable for a reliable sequence of rate R. In that α determines how long the input letter $x_1 \in \mathcal{X}$ is used for the special message while P_X^1 is being used for the ordinary codewords. Furthermore, arguments very similar to those of Theorem 8 can be used to prove no missed detection exponent higher than $\mathcal{J}(R)$ is achievable for reliable sequences of rate R. Thus, $\mathcal{J}(R)$ is the best exponent a message can get in a rate R reliable sequence.

One can show that $\mathcal{J}(R)$ is a concave function of R over its support $[0, C]$. Furthermore, $\mathcal{J}(0) = D_{\max}$ and $\mathcal{J}(C) = \tilde{C}$. Thus, $\mathcal{J}(R)$ is a concave strictly decreasing function of R for $0 \le R \le C$.

Proof (of Lemma 1): Recall that $\mathcal{G}(i)$ is the decoding region for $M = i$, i.e., $\mathcal{G}(i) = \{y^\tau : \hat{M}(y^\tau) = i\}$. Then as a result of data processing inequality for KL divergence we have

$$\begin{aligned} &E\left[\ln\frac{\Pr[Y^\tau]}{\Pr[Y^\tau|\,M=i]}\right] \\ &\ge \Pr[\mathcal{G}(i)]\ln\frac{\Pr[\mathcal{G}(i)]}{\Pr[\mathcal{G}(i)|\,M=i]} \\ &\quad + \Pr\left[\overline{\mathcal{G}(i)}\right]\ln\frac{\Pr\left[\overline{\mathcal{G}(i)}\right]}{\Pr\left[\overline{\mathcal{G}(i)}\middle|\,M=i\right]} \\ &\ge -h(\Pr[\mathcal{G}(i)]) + \Pr\left[\overline{\mathcal{G}(i)}\right]\ln\frac{1}{\Pr\left[\overline{\mathcal{G}(i)}\middle|\,M=i\right]} \\ &\ge -\ln 2 + \Pr\left[\overline{\mathcal{G}(i)}\right]\ln\frac{1}{\Pr\left[\overline{\mathcal{G}(i)}\middle|\,M=i\right]} \end{aligned} \tag{37}$$

where in the last step we have used, the fact that $h(\Pr[\mathcal{G}(i)]) \le \ln 2$. In addition

$$\begin{aligned} \Pr\left[\overline{\mathcal{G}(i)}\right] &\ge \Pr\left[\overline{\mathcal{G}(i)}\middle|\,M\ne i\right]\Pr[M\ne i] \\ &\ge \sum_{j\ne i}\Pr[\mathcal{G}(j)|\,M=j]\Pr[M=j] \\ &\ge (1-P_e-\Pr[M=i]). \end{aligned} \tag{38}$$

Thus, using (37) and (38) we get

$$\Pr\left[\overline{\mathcal{G}(i)}\middle|M=i\right] \ge e^{-\frac{1}{1-P_e-\Pr[M=i]}\left(\ln 2+E\left[\ln\frac{\Pr[Y^\tau]}{\Pr[Y^\tau|M=i]}\right]\right)}. \tag{39}$$

Now we lower-bound the error probability of the special message by upper-bounding $E\left[\ln\frac{\Pr[Y^\tau]}{\Pr[Y^\tau|M=i]}\right]$. For that let us consider the following stochastic sequence:

$$S_n = \ln\frac{\Pr[Y^n]}{\Pr[Y^n|\,M=i]} - \sum_{t=1}^{n} E\left[\ln\frac{\Pr\left[Y_t|\,Y^{t-1}\right]}{\Pr[Y_t|\,M=i,Y^{t-1}]}\middle|\,Y^{t-1}\right].$$

Note that $E[S_{n+1}|\,Y^n] = S_n$ and since $\min W_{i,j} = \lambda$ we have $E[|S_{n+1}-S_n||\,Y^n] \le 2\ln\frac{1}{\lambda}$. Thus, S_n is a martingale, furthermore, since $E[\tau] < \infty$ we can use [37, Theorem 2, p. 487], to get

$$E[S_\tau] = S_0 = 0.$$

Thus

$$E\left[\ln\frac{\Pr[Y^\tau]}{\Pr[Y^\tau|\,M=1]}\right] = E\left[\sum_{t=1}^{\tau}E\left[\ln\frac{\Pr\left[Y_t|\,Y^{t-1}\right]}{\Pr[Y_t|\,M=1,Y^{t-1}]}\middle|\,Y^{t-1}\right]\right]. \tag{40}$$

Note that

$$E\left[\ln\frac{\Pr\left[Y_t|\,Y^{t-1}\right]}{\Pr[Y_t|\,M=1,Y^{t-1}]}\middle|\,Y^{t-1}\right] = E\left[\ln\frac{\Pr\left[Y_t|\,Y^{t-1}\right]}{W_{Y|X}(Y_t|\bar{x}_t(1))}\middle|\,Y^{t-1}\right].$$

As a result of the definition of $\mathcal{J}(\cdot)$ given in (36), we have

$$E\left[\ln\frac{\Pr\left[Y_t|\,Y^{t-1}\right]}{\Pr[Y_t|\,M=1,Y^{t-1}]}\middle|\,Y^{t-1}\right] \le \mathcal{J}\left(\mathcal{I}\left(X_t;Y_t\middle|Y^{t-1}\right)\right) \tag{41}$$

where $\mathcal{I}\left(X_t;Y_t\middle|Y^{t-1}\right)$ is given by[13]

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

Given Y^{t-1} random variables $M - X_t - Y_t$ forms a Markov chain. Thus

$$\mathcal{I}\left(X_t;Y_t\middle|Y^{t-1}\right) \ge \mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right). \tag{42}$$

Since $\mathcal{J}(\cdot)$ is a decreasing function, (40), (41), and (42) lead to

$$E\left[\ln\frac{\Pr[Y^\tau]}{\Pr[Y^\tau|\,M=1]}\right] \le E\left[\sum_{t=1}^{\tau}\mathcal{J}\left(\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right)\right]. \tag{43}$$

Note that

$$\begin{aligned} &E\left[\sum_{t=1}^{\tau}\mathcal{J}\left(\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right)\right] \\ &= E\left[\tau\sum_{t=1}^{\tau}\frac{1}{\tau}\mathcal{J}\left(\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right)\right] \\ &\overset{(a)}{\le} E\left[\tau\mathcal{J}\left(\sum_{t=1}^{\tau}\frac{1}{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right)\right] \\ &= E[\tau]\,E\left[\frac{\tau}{E[\tau]}\mathcal{J}\left(\sum_{t=1}^{\tau}\frac{1}{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right)\right] \\ &\overset{(b)}{\le} E[\tau]\,\mathcal{J}\left(E\left[\frac{\tau}{E[\tau]}\sum_{t=1}^{\tau}\frac{1}{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right]\right) \\ &= E[\tau]\,\mathcal{J}\left(E\left[\frac{\sum_{t=i}^{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)}{E[\tau]}\right]\right) \end{aligned} \tag{44}$$

where in both (a) and (b) we use the concavity of the $\mathcal{J}(\cdot)$ function together with Jensen's inequality. Thus, using (39), (43),

[13]Note that unlike the conventional definition of conditional mutual information, $\mathcal{I}(X_t;Y_t|\,Y^{t-1})$ is not averaged over the conditioned random variable Y^{t-1}.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

and (44) we get

$$\Pr\left[\hat{M}\neq i\middle|\, M=i\right] \geq e^{-\frac{1}{1-P_e-\Pr[M=i]}\left(\mathcal{J}\left(\frac{E\left[\sum_{t=i}^{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right]}{E[\tau]}\right)E[\tau]+\ln 2\right)}.$$

Since $\mathcal{J}(R)$ is decreasing in R, the only thing we are left to show is that

$$E\left[\sum_{t=i}^{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right]\geq \mathcal{H}(M)-h(P_e)-P_e\ln(|\mathcal{M}|-1). \tag{45}$$

For that, consider the stochastic sequence

$$V_n=\mathcal{H}(M|Y^n)+\sum_{t=1}^{n}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right).$$

Clearly, $E[V_{n+1}|Y^n]=V_n$ and $E[|V_n|]<\infty$, thus $\{V_n\}$ is a martingale. Furthermore, $E[|V_{n+1}-V_n||Y^n]\leq K$ and $E[\tau]<\infty$ thus, using a version of Doob's optional stopping theorem, [37, Theorem 2, p. 487], we get

$$\begin{aligned}V_0&=E[V_\tau]\\&=E[\mathcal{H}(M|Y^\tau)]+E\left[\sum_{t=1}^{\tau}\mathcal{I}\left(M;Y_t\middle|Y^{t-1}\right)\right].\end{aligned} \tag{46}$$

One can write Fano's inequality as follows:

$$\mathcal{H}(M|Y^\tau)\leq h\left(\Pr\left[\hat{M}(Y^\tau)\neq M\middle|Y^\tau\right]\right)+\Pr\left[\hat{M}(Y^\tau)\neq M\middle|Y^\tau\right]\ln|\mathcal{M}|.$$

Consequently

$$E[\mathcal{H}(M|Y^\tau)]\leq E\left[h\left(\Pr\left[\hat{M}(Y^\tau)\neq M\middle|Y^\tau\right]\right)\right]+E\left[\Pr\left[\hat{M}(Y^\tau)\neq M\middle|Y^\tau\right]\right]\ln|\mathcal{M}|.$$

Using the concavity of binary entropy

$$E[\mathcal{H}(M|Y^\tau)]\leq h(P_e)+P_e\ln|\mathcal{M}|. \tag{47}$$

Using (46) together with (47) we get the desired condition given in (45). •

The preceding proof is for encoding schemes which do not have any randomization (time sharing), but same ideas can be used to establish the exact same result for general variable-length block codes with randomization. Now we are ready to prove the converse part of the Theorem 8.

Proof (of Converse Part of Theorem 8): In order to prove $E_{\mathrm{md}}^{f}\leq\tilde{C}$, first note that for capacity-achieving sequences we consider $\Pr[M=i]=\frac{1}{|\mathcal{M}^{(k)}|}$. Thus

$$\frac{-\ln({P_e}^{M}(i))^{(k)}}{E\left[\tau^{(k)}\right]}\leq\frac{\mathcal{J}\left(\frac{\ln|\mathcal{M}^{(k)}|-h(P_e(k))-{P_e}^{(k)}\ln|\mathcal{M}(k)|}{E\left[\tau^{(k)}\right]}\right)+\frac{\ln 2}{E\left[\tau^{(k)}\right]}}{1-{P_e}^{(k)}-\frac{1}{|\mathcal{M}^{(k)}|}}.$$

Thus, for any capacity-achieving sequence with feedback

$$\lim_{k\to\infty}-\frac{\ln({P_e}^{M}(i))^{(k)}}{E\left[\tau^{(k)}\right]}\leq\mathcal{J}(C)=\tilde{C}. \tag{48}$$

•

E. Proof of Theorem 9

In this subsection, we will show how the strategy for sending a special bit can be combined with the Yamamoto–Itoh strategy when many special messages demand a missed-detection exponent. However, unlike previous results about capacity-achieving sequences, Theorems 5–8, we will have an additional uniform delay assumption.

We will restrict ourself to uniform delay capacity-achieving sequences.[14] Clearly, capacity-achieving sequences in general need not to be uniformly delayed. Indeed, many messages, $i\in\mathcal{M}$, can get an expected delay, $E[\tau|\,M=i]$ much larger than the average delay $E[\tau]$. This in turn can decrease the error probability of these messages. The potential drawback of such codes, is that their average delay is sensitive to assumption of messages being chosen according to a uniform probability distribution. Expected decoding time, $E[\tau]$, can increase a lot if the code is used in a system in which the messages are not chosen uniformly.

It is worth emphasizing that all previously discussed exponents (single-message exponent E_{md}^{f}, single-bit exponent E_{b}^{f}, many-bits exponent $E_{\mathrm{b}}^{f}(r)$, and achievable multilayer exponent regions) remain unchanged whether or not this uniform delay constraint is imposed. Thus, the flexibility to provide different expected delays to different messages does not improve those exponents.

However, this is not true for the message-wise UEP with exponentially many messages. Removing the uniform delay constraint can considerably enhance the protection of special messages at rates higher than $(1-\frac{\tilde{C}}{D_{\max}})C$. Indeed, one can make the exponent of all special messages C. The flexibility of providing more resources (decoding delay) to special messages achieves this enhancement. However, we will not discuss those cases in this paper and stick to uniform delay codes.

1) Achievability: $E_{\mathrm{md}}^{f}(r)\geq\min\{\tilde{\mathrm{C}},(1-\frac{r}{\mathrm{C}})\mathrm{D}_{\max}\}$*:* The optimal scheme here reverses the trick for achieving E_{b}^{f}: first, a special bit tells the receiver whether the message being transmitted is special one or not. After the decoding of this bit, the message itself is transmitted. This further emphasizes how feedback connects bit-wise and message-wise UEP, when used with variable decoding time.

Proof: Like all the previous achievability results, we construct a capacity-achieving sequence $\mathcal{Q}$, with the desired asymptotic behavior. A sequence of multiphase fixed-length errors and erasures codes, $\mathcal{Q}'$, is used as the building block of $\mathcal{Q}$. Let us consider the kth member of $\mathcal{Q}'$. In the first phase, the transmitter

[14]Recall that for any reliable variable-length block code with feedback Γ is defined as

$$\Gamma=\frac{\max_{i\in\mathcal{M}}E[\tau|\,M=i]}{E[\tau]}$$

and uniform delay reliable sequences are the ones that satisfy $\lim_{k\to\infty}\Gamma_{\mathcal{Q}}^{(k)}=1$.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

sends one of the two input symbols with distinct output distributions for $\lfloor\sqrt{k}\rfloor$ time units in order to tell whether $M \in \mathcal{M}_s^{(k)}$ or not. Let b be $b = \mathbb{I}_{\left\{M\in\mathcal{M}_s^{(k)}\right\}}$. Then, as mentioned in Section VIII-A.1, with a threshold decoding we can achieve

$$\Pr\left[\hat{b}\neq 1\middle| b=1\right] = \Pr\left[\hat{b}\neq 0\middle| b=0\right] \leq e^{-\sqrt{k}\mu} \qquad (49)$$

where $\mu > 0$. However, the actual value of μ is not important here, we are merely interested in an upper bound vanishing with increasing k.

In the second phase, one of two length k codes is used depending on $\hat{b}$.

- If $\hat{b} = 0$, in the second phase, the transmitter uses the kth member of a capacity-achieving sequence, $\mathcal{Q}''$ such that $E_{\mathrm{b},\mathcal{Q}''} = \tilde{C}$. We know that such a sequence exists because of Theorem 2. The message M' of the $\mathcal{Q}''$ is determined using the following mapping:

$$M \in \mathcal{M}_s \Rightarrow \mathrm{M}' = 1$$
$$M \notin \mathcal{M}_s \Rightarrow \mathrm{M}' = M - |\mathcal{M}_s| + 1.$$

  At the end of the second phase, the receiver decodes M'. If $\hat{\mathrm{M}}' = 1$, then the receivers declares an erasure, $\tilde{M} =$ erasure. If $\hat{\mathrm{M}}' \neq 1$, then $\hat{M} = \tilde{M} = \hat{\mathrm{M}}' + |\mathcal{M}_s| - 1$.
- If $\hat{b} = 1$, the transmitter uses a two-phase code with errors and erasures in the second phase, like the one described by Yamamoto and Itoh in [40]. The two phases of this code are called communication and control phases, respectively.
  In the communication phase, the transmitter uses the $\lceil rk\rceil$th member of a capacity-achieving sequence $\mathcal{Q}''$ with $E_{\mathrm{b},\mathcal{Q}''} = \tilde{C}$, to convey its message M'. The auxiliary message M' is determined as follows:

$$M \notin \mathcal{M}_s \Rightarrow \mathrm{M}' = 1$$
$$M \in \mathcal{M}_s \Rightarrow \mathrm{M}' = M + 1.$$

  The decoded message of the $\lceil rk\rceil$th member of $\mathcal{Q}''$ is called the tentative decision of communication phase and denoted by $\tilde{\mathrm{M}}'$.
  In the control phase,
  — if $\tilde{\mathrm{M}}' = \mathrm{M}'$, the tentative decision is confirmed by sending the accept symbol x_a for $\ell(k) = k - \lceil \frac{r}{C}k\rceil$ time units.
  — if $\tilde{\mathrm{M}}' \neq \mathrm{M}'$, the tentative decision is rejected by sending the reject symbol x_d for $\ell(k) = k - \lceil \frac{r}{C}k\rceil$ time units

  where x_a and x_d are the maximizers in the following optimization problem:

$$D_{\max} = \max_{i,j} D\left(W_{Y|x}(\cdot|i)\middle\| W_{Y|X}(\cdot|j)\right)$$
$$= D\left(W_{Y|x}(\cdot|x_a)\middle\| W_{Y|X}(\cdot|x_d)\right).$$

  If the output sequence in the last $k - \lceil \frac{r}{C}k\rceil$ time steps is typical with $W_{Y|X}(\cdot|x_a)$ then $\hat{\mathrm{M}}' = \tilde{\mathrm{M}}'$; else, an erasure is declared for M'. Note that the total probability of $W_{Y|X}(\cdot|x_a)$ typical sequences are less than $e^{-\ell(k)(D_{\max}-\delta_{\ell(k)})}$ when $\tilde{\mathrm{M}}' \neq \mathrm{M}'$ and more than $1-\delta_{\ell(k)}$ when $\tilde{\mathrm{M}}' = \mathrm{M}'$ where $\lim_{\ell(k)\to\infty}\delta_{\ell(k)} = 0$, [13, Corrollary 1.2, p. 19].
  If $\hat{\mathrm{M}}' =$ erasure or if $\hat{\mathrm{M}}' = 1$, then the receiver declares an erasure for M, $\tilde{M} =$ erasure. If $\hat{\mathrm{M}}' \in \{2, 3, \ldots, |\mathcal{M}_s|+1\}$, then $\hat{M} = \tilde{M} = \hat{M} - 1$.

Now we can calculate the error and erasure probabilities of the two-phase fixed-length block code. Let us denote the erasures by $\tilde{M} =$ erasure for each k.

For $i \in \mathcal{M}_s$ using (49) and Bayes rule we get

$$\Pr\left[\tilde{M} = \text{erasure}\middle| M = i\right] \leq e^{-\mu\sqrt{k}} + \left(P_{e,\mathcal{Q}'}^{(k-\ell(k))} + \delta_{\ell(k)}\right) \qquad (50)$$

$$\Pr\left[\tilde{M}\neq i, \tilde{M}\neq \text{erasure}\middle| M = i\right] \leq e^{-\mu\sqrt{k}} P_{e\mathcal{Q}'}^{k}(1) + P_{e,\mathcal{Q}'}^{(k-\ell(k))} e^{-\ell(k)(D_{\max}-\delta_{\ell(k)})}. \qquad (51)$$

For $i \notin \mathcal{M}_s$ using (49) and the Bayes rule we get

$$\Pr\left[\tilde{M}=\text{erasure}\middle| M=i\right] \leq e^{-\mu\sqrt{k}} + P_{e,\mathcal{Q}'}^{(k)} \qquad (52)$$

$$\Pr\left[\tilde{M}\neq i, \tilde{M}\neq\text{erasure}\middle| M=i\right] \leq e^{-\mu\sqrt{k}} + P_{e,\mathcal{Q}'}^{(k)}. \qquad (53)$$

Whenever $\tilde{M} =$ erasure, than the transmitter and the receiver try to send the message once again from scratch using the same strategy. Then, for any $i \in \mathcal{M}$

$$\Pr\left[\hat{M}\neq i\middle| M=i\right] = \frac{\Pr\left[\tilde{M}\neq i, \tilde{M}\neq\text{erasure}\middle| M=i\right]}{1-\Pr\left[\tilde{M}=\text{erasure}\middle| M=i\right]} \qquad (54)$$

$$E\left[\tau\middle| M=i\right] = \frac{k+\sqrt{k}}{1-\Pr\left[\tilde{M}=\text{erasure}\middle| M=i\right]}. \qquad (55)$$

Using (50), (51), (52), (53), (54), and (55) we conclude that $\mathcal{Q}$ is a capacity-achieving sequence such that

$$\lim_{k\to\infty} -\frac{\ln\max_{i\in\mathcal{M}_s}\Pr\left[\tilde{M}\neq i, \hat{M}\neq\text{erasure}\middle| M=i\right]}{E[\tau]} = \min\{\tilde{C}, (1-\frac{r}{C})D_{\max}\}$$
$$\lim_{k\to\infty}\frac{\ln|\mathcal{M}_s^{(k)}|}{E[\tau]} = r. \qquad \bullet$$

2) Converse: $E_{\mathrm{md}}^{f}(r) \leq \min\{\tilde{\mathrm{C}}, (1-\frac{r}{\mathrm{C}})\mathrm{D}_{\max}\}$:

Proof: Consider any uniform delay capacity-achieving sequence $\mathcal{Q}$. Note that by excluding all $i \notin \mathcal{M}_s^{(k)}$ we get a reliable sequence $\mathcal{Q}'$ such that

$$P_e'^{(k)} \leq \Pr^{(k)}\left[\hat{M}\neq M\middle| M\in\mathcal{M}_s\right]$$
$$E\left[\tau'^{(k)}\right] \leq \Gamma^{(k)} E\left[\tau^{(k)}\right].$$

Thus

$$\frac{-\ln\Pr\left[\hat{M}\neq M\middle| M\in\mathcal{M}_s\right]^{(k)}}{E\left[\tau^{(k)}\right]} \leq -\frac{\ln P_e'^{(k)}}{E\left[\tau'^{(k)}\right]}\Gamma^{(k)}.$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

$$\sum_{y^n\in\mathcal{Y}^n}\Pr[y^n|\,M=1]\ln\frac{\Pr[y^n|\,M=1]}{\Pr[y^n|\,M\neq 1]}\le\sum_{i=2}^{|\mathcal{M}|}\frac{1}{|\mathcal{M}|-1}\sum_{y^n\in\mathcal{Y}^n}\Pr[y^n|\,M=1]\ln\frac{\Pr[y^n|\,M=1]}{\Pr[y^n|\,M=i]}$$
$$=\sum_{i=2}^{|\mathcal{M}|}\frac{1}{|\mathcal{M}|-1}\sum_{y^n\in\mathcal{Y}^n}\Pr[y^n|\,M=1]\sum_{k=1}^{n}\ln\frac{\Pr[y_k|\,M=1,y^{k-1}]}{\Pr[y_k|\,M=i,y^{k-1}]}$$
$$=\sum_{k=1}^{n}\sum_{i=2}^{|\mathcal{M}|}\frac{D\left(W_{Y|X}(\cdot|\bar{x}_k(1))\big\|\,W_{Y|X}(\cdot|\bar{x}_k(i))\right)}{|\mathcal{M}|-1}\tag{57}$$

Consequently, $E_{\mathrm{md}}^{f}(r)\le(1-\frac{r}{C})D_{\max}$. Similarly, by excluding all but one of the elements of $\mathcal{M}_s$, we can prove that $E_{\mathrm{md}}^{f}(r)\le\tilde{C}$, using Theorem 8 and uniform delay condition. •

IX. Avoiding False Alarms: Proofs

A. Block Codes Without Feedback: Proof of Theorem 10

1) Lower Bound: $E_{fa}\ge E_{fa}^{l}$:

Proof: As a result of the coding theorem [13, Ch. 2, Corollary 1.3, p. 102] we know that there exits a reliable sequence $\mathcal{Q}'$ of fixed composition codes whose rate is C and whose nth elements composition $P_X^{(n)}$ satisfies

$$\sum_{i\in\mathcal{X}}|P_X^{(n)}(i)-P_X^{*}(i)|\le\sqrt[4]{\frac{1}{n}}.$$

We use the codewords of the nth element of $\mathcal{Q}'$ as the codewords of the ordinary messages in the nth code in $\mathcal{Q}$. For the special message we use a length-n repetition sequence $\bar{x}^n(1)=(x_{f_l},x_{f_l},\ldots,x_{f_l})$.

The decoding region for the special message is essentially the bare minimum. We include typical channel outputs within the decoding region of the special message to ensure small missed detection probability for the special message, but we exclude all other output sequence y^n.

$$\mathcal{G}(1)=\left\{y^n:\sum_{i\in\mathcal{Y}}|\mathsf{Q}_{(y^n)}(i)-W_{Y|X}(i|x_{f_l})|\le\sqrt[4]{1/n}\right\}.$$

Note that this definition of $\mathcal{G}(1)$ itself ensures that special message is transmitted reliably whenever it is sent,

$$\lim_{n\to\infty}\Pr^{(n)}\left[\hat{M}\neq 1\middle|\,M=1\right]=0.$$

The decoding region of the ordinary messages, $j=\{2,3,\ldots\mathcal{M}^{(n)}\}$, is the intersection of the corresponding decoding region in $\mathcal{Q}'$ with the complement of $\mathcal{G}(1)$. Thus, the fact that $\mathcal{Q}'$ is a reliable sequence implies that

$$\lim_{n\to\infty}\Pr^{(n)}\left[y^n\in\bigcup_{j\notin\{1,i\}}\mathcal{G}(j)\middle|\,M=i\right]=0.$$

Consequently, we have reliable communication for ordinary messages as long as $\lim_{n\to\infty}\Pr^{(n)}[\mathcal{G}(1)|\,M=j]=0$, $\forall j\neq 1$. But we prove a much stronger result to ensure that $\Pr^{(n)}\left[\hat{M}=1\middle|\,M\neq 1\right]$ is decaying fast enough. Before doing that let us note that in the second stage of the decoding, when we are choosing a message among the ordinary ones, ML decoder can be used instead of the decoding rule of the original code. Doing that will only decrease the average error probability.

Note the probability of a V-shell of a message i is equal to

$$\Pr^{(n)}[\mathsf{T}_V(i)|\,M=i]=e^{-nD\left(V_{Y|X}(\cdot|X)\|W_{Y|X}(\cdot|X)|P_X^{(n)}\right)}.$$

Note that also that $\mathcal{G}(1)$ can be written as the union of V-shells of a message i as follows:

$$\mathcal{G}(1)=\bigcup_{V_{Y|X}\in\mathcal{V}^{(n)}}\mathsf{T}_V(i),\quad\forall i\neq 1$$

where

$$\mathcal{V}^{(n)}=\{V_{Y|X}:\sum_j|(P^nV)_Y(j)-W_{Y|X}(j|x_{f_l})|\le\sqrt[4]{1/n}\}.$$

Note that since there are at most $(1+n)^{|\mathcal{X}||\mathcal{Y}|}$ different conditional types

$$\Pr^{(n)}[\mathcal{G}(1)|\,M=i]\le(1+n)^{|\mathcal{X}||\mathcal{Y}|}\max_{V_{Y|X}\in\mathcal{V}^{(n)}}\Pr[\mathsf{T}_V(i)|\,M=i].$$

Thus, for all $i>1$

$$\lim_{n\to\infty}\frac{-\ln\Pr^{(n)}[\mathcal{G}(1)|\,M=i]}{n}=\min_{\substack{V_{Y|X}:\\(P^*V)_Y(\cdot)=W_{Y|X}(\cdot|x_{f_l})}}D\left(V_{Y|X}(\cdot|X)\big\|\,W_{Y|X}(\cdot|X)\big|\,P_X^{*}\right).$$

•

2) Upper Bound: $E_{fa}\le E_{fa}^{u}$:

Proof: As a result of data processing inequality for KL divergence we have

$$\sum_{y^n\in\mathcal{Y}^n}\Pr[y^n|\,M=1]\ln\frac{\Pr[y^n|\,M=1]}{\Pr[y^n|\,M\neq 1]}$$
$$\ge\Pr[\mathcal{G}(1)|\,M=1]\ln\frac{\Pr[\mathcal{G}(1)|\,M=1]}{\Pr[\mathcal{G}(1)|\,M\neq 1]}+\Pr\left[\overline{\mathcal{G}(1)}\middle|\,M=1\right]\ln\frac{\Pr\left[\overline{\mathcal{G}(1)}\middle|\,M=1\right]}{\Pr\left[\overline{\mathcal{G}(1)}\middle|\,M\neq 1\right]}$$
$$\ge-\ln 2-\Pr[\mathcal{G}(1)|\,M=1]\ln\Pr[\mathcal{G}(1)|\,M\neq 1].\tag{56}$$

Using the convexity of the KL divergence we get (57) at the top of the page, where $\bar{x}_k(i)$ denotes the input letter for codeword of message i, at time k.

Let us denote the empirical distribution of the $\bar{x}_k(i)$ for time k, by P_{X_k}.

$$P_{X_k}(i)=\frac{\sum_{j\in\mathcal{M}}\mathbb{1}_{\{\bar{x}_k(j)=i\}}}{|\mathcal{M}|},\quad\forall i\in\mathcal{X}.$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

Using (56) and (57) we get

$$\Pr\left[\mathcal{G}(1)|M\neq 1\right] \geq e^{-\frac{|\mathcal{M}|\sum_k D\left(W_{Y|X}(\cdot|x_k(1))\middle\|W_{Y|X}(\cdot|X_k)\middle|P_{X_k}\right)}{(|\mathcal{M}|-1)\Pr[\mathcal{G}(1)|M=1]}\frac{\ln 2}{\Pr[\mathcal{G}(1)|M=1]}}. \quad (58)$$

We show below that for all capacity-achieving codes, almost all of the k's have a P_{X_k} which is essentially equal to P_X^*. To do that, let us first define the set $\mathcal{P}(\epsilon)$ and $\delta(\epsilon)$

$$\mathcal{P}(\epsilon) \triangleq \{P_X : I(P_X, W_{Y|X}) \geq C-\epsilon\}$$
$$\delta(\epsilon) \triangleq \max_{P_X\in\mathcal{P}(\epsilon)} \sum_i |P_X(i) - P_X^*(i)|.$$

Note that $\lim_{\epsilon\to 0}\delta(\epsilon)=0$. As a result of Fano's inequality, we have

$$\mathcal{I}(M;Y^n) \geq nR^{(n)}(1-P_e) - \ln 2. \quad (59)$$

On the other hand, using standard manipulations on mutual information we get

$$\mathcal{I}(M;Y^n) = \sum_{k=1}^n I(P_{X_k}, W_{Y|X}) \leq Cn - \epsilon\sum_{k=1}^n \mathbb{I}_{\{P_{X_k}\notin\mathcal{P}(\epsilon)\}}. \quad (60)$$

Using (60) in (59) we get

$$\sum_{k=1}^n \mathbb{I}_{\{P_{X_k}\notin\mathcal{P}(\epsilon)\}} \leq n\frac{(C-R^{(n)}(1-P_e)-\ln 2/n)}{\epsilon}.$$

Let $\epsilon(n)$ be

$$\epsilon(n) = \sqrt{C - R^{(n)}(1-P_e) - \frac{\ln 2}{n}}$$

then $\lim_{n\to\infty}\epsilon(n)=0$ and

$$\sum_{k=1}^n \mathbb{I}_{\{P_{X_k}\notin\mathcal{P}(\epsilon^{(n)})\}} \leq n\epsilon^{(n)}. \quad (61)$$

Note for any $P_X\in\mathcal{P}(\epsilon^{(n)})$, we have

$$\begin{aligned} &D\left(W_{Y|X}(\cdot|x_k(1))\middle\|W_{Y|X}(\cdot|X_k)\middle|P_X\right) \\ &\leq D\left(W_{Y|X}(\cdot|x_k(1))\middle\|W_{Y|X}(\cdot|X)\middle|P_X^*\right) + \delta(\epsilon^{(n)})D_{\max} \\ &\leq E_{\mathrm{fa}}^{\mathrm{u}} + \delta(\epsilon^{(n)})D_{\max} \end{aligned} \quad (62)$$

where $E_{\mathrm{fa}}^{\mathrm{u}} = \max_{i\in\mathcal{X}} D\left(W_{Y|X}(\cdot|i)\middle\|W_{Y|X}(\cdot|X)\middle|P_X^*\right)$.

Using (61) and (62)

$$\sum_k D\left(W_{Y|X}(\cdot|x_k(1))\middle\|W_{Y|X}(\cdot|X_k)\middle|P_{X_k}\right) \leq n(E_{\mathrm{fa}}^{\mathrm{u}} + \delta(\epsilon^{(n)})D_{\max} + \epsilon^{(n)}D_{\max}).$$

Inserting this in (58) we get

$$\lim_{n\to\infty}\left(\frac{-\ln\Pr^{(n)}[\mathcal{G}(1)|M\neq 1]}{n}\right) \leq E_{\mathrm{fa}}^{\mathrm{u}}. \quad \bullet$$

B. Variable-Length Block Codes With Feedback: Proof of Theorem 11

1) Achievability: $E_{fa}^f \geq \mathrm{D}_{\max}$:

Proof: We construct a capacity-achieving sequence with feedback $\mathcal{Q}$ by using a construction like the one we have for $E_{\mathrm{md}}^f(r)$. In fact, this scheme achieves the false-alarm exponent simultaneously with the best missed detection exponent $\tilde{C}$ for the special message.

We use a fixed-length multiphase errors and erasure code as the building block for the kth member of $\mathcal{Q}$. In the first phase, $b=\mathbb{I}_{\{M=1\}}$ is conveyed using a length $\lceil\sqrt{k}\rceil$ repetition code, like we did in Sections VIII-A.1 and VIII-E.1. Recall that

$$\Pr\left[\hat{b}\neq 1\middle|\, b=1\right] = \Pr\left[\hat{b}\neq 0\middle|\, b=0\right] \leq e^{-\mu\sqrt{k}}, \quad \mu>0. \quad (63)$$

In the second phase, one of the two length-k codes is used depending on $\hat{b}$.

- If $\hat{b}=0$, the transmitter uses the kth member of a capacity-achieving sequence, $\mathcal{Q}'$, such that $E_{\mathrm{md},\mathcal{Q}'}=\tilde{C}$ to convey the message. We know that such a sequence exists because of Theorem 2. Let the message of $\mathcal{Q}$ be the message of $\mathcal{Q}'$, i.e., the auxiliary message
$$\mathrm{M}' = M.$$
If at the end of the second phase $\hat{\mathrm{M}}'=1$, the receiver declares an erasure, $\hat{M}=$erasure, else, M is decoded $\hat{M}=\tilde{M}=\hat{\mathrm{M}}'$.
- If $\hat{b}=1$, the transmitter uses a length-k repetition code to convey whether $M=1$ or not.
  - If $M=1$, $\mathrm{M}'=1$ and the transmitter sends the codeword $(x_a,x_a,\ldots,x_a)$;
  - If $M\neq 1$, $\mathrm{M}'=0$ and the transmitter sends the codeword $(x_d,x_d,\ldots,x_d)$;

  where x_a and x_d are the maximizers achieving $D_{\max}$
$$\begin{aligned} D_{\max} &= \max_{i,j} D\left(W_{Y|x}(\cdot|i)\middle\|W_{Y|X}(\cdot|j)\right) \\ &= D\left(W_{Y|x}(\cdot|x_a)\middle\|W_{Y|X}(\cdot|x_d)\right). \end{aligned}$$
The receiver decodes $\hat{\mathrm{M}}'=1$ only when the output sequence is typical with $W_{Y|X}(\cdot|x_a)$. Evidently as before we have, [13, Corollary 1.2, p. 19]
$$\Pr\left[\hat{\mathrm{M}}'=0\middle|\, M=1\right] \leq \delta_k \quad (64)$$
$$\Pr\left[\hat{\mathrm{M}}'=1\middle|\, M=0\right] \leq e^{-k(D_{\max}-\delta_k)} \quad (65)$$
where $\lim_{k\to\infty}\delta_k=0$.
If $\hat{\mathrm{M}}'=1$ then $\hat{M}=1$, else the receiver declares an erasure for the whole block, i.e., $\hat{M}=$ erasure.

Now we can calculate the error and erasure probabilities for $(\lceil k\rceil+k)$ long block code. Using (63), (64), (65), and the Bayes' rule we get

$$\Pr\left[\hat{M}=\text{erasure}\,\middle|\, M=1\right] \leq e^{-\mu\sqrt{k}} + \delta_k \quad (66)$$

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

$$\Pr\left[\tilde{M}=\text{erasure}\,\middle|\,M=i\neq 1\right]\leq e^{-\mu\sqrt{k}}+P_{e\mathcal{Q}'}^{(k)} \quad (67)$$

$$\Pr\left[\tilde{M}\in\mathcal{M}\setminus\{1\}\,\middle|\,M=1\right]\leq e^{-\mu\sqrt{k}}P_{e\mathcal{Q}'}^{(k)}(1) \quad (68)$$

$$\Pr\left[\tilde{M}\in\mathcal{M}\setminus\{1,i\}\,\middle|\,M=i\neq 1\right]\leq P_{e\mathcal{Q}'}^{(k)} \quad (69)$$

$$\Pr\left[\tilde{M}=1\,\middle|\,M=i\neq 1\right]\leq e^{-\mu\sqrt{k}}e^{-k(D_{\max}-\delta_k)}. \quad (70)$$

Whenever $\tilde{M}=$ erasure than the transmitter tries to send the message again from scratch, using the same strategy. Consequently, all of the above error probabilities are scaled by a factor of $\frac{1}{1-\Pr[\tilde{M}=\text{erasure}|M=i]}$ when we consider the corresponding error probabilities for the variable decoding time code. Furthermore

$$E[\tau|\,M=i]=\frac{k+\sqrt{k}}{1-\Pr\left[\tilde{M}=\text{erasure}\,\middle|\,M=i\right]}. \quad (71)$$

Using (66), (67), (68), (69), (70), and (71) we conclude that $\mathcal{Q}$ is a capacity-achieving code with $E_{\text{md},\mathcal{Q}}^{f}=\tilde{C}$ and $E_{\text{fa},\mathcal{Q}}^{f}=D_{\max}$. •

2) Converse: $E_{fa}^{f}\leq \mathrm{D}_{\max}$:

Proof: Note that as a result of convexity of KL divergence we have

$$\begin{aligned}
&E\left[\ln\frac{\Pr[Y^{\tau}|\,M=1]}{\Pr[Y^{\tau}|\,M\neq 1]}\,\middle|\,M=1\right]\\
&\geq \Pr[\mathcal{G}(1)|\,M=1]\ln\frac{\Pr[\mathcal{G}(1)|\,M=1]}{\Pr[\mathcal{G}(1)|\,M\neq 1]}\\
&\quad+\Pr\left[\overline{\mathcal{G}(1)}\,\middle|\,M=1\right]\ln\frac{\Pr\left[\overline{\mathcal{G}(1)}\,\middle|\,M=1\right]}{\Pr\left[\overline{\mathcal{G}(1)}\,\middle|\,M\neq 1\right]}\\
&\geq -\ln 2+\Pr[\mathcal{G}(1)|\,M=1]\ln\frac{1}{\Pr[\mathcal{G}(1)|\,M\neq 1]}. \quad (72)
\end{aligned}$$

It has already been proven in [4] that

$$E\left[\ln\frac{\Pr[Y^{\tau}|\,M=1]}{\Pr[Y^{\tau}|\,M\neq 1]}\,\middle|\,M=1\right]\leq D_{\max}E[\tau|\,M=1]. \quad (73)$$

Note that as a result of the definition of Γ, we have $E[\tau|\,M=1]\leq E[\tau]\,\Gamma$. Using this together with (72) and (73) the we get

$$\Pr[\mathcal{G}(1)|\,M\neq 1]\geq e^{-\frac{\ln 2+\Gamma D_{\max}E[\tau]}{\Pr[\mathcal{G}(1)|M=1]}}.$$

Thus. for any uniform delay reliable sequence $\mathcal{Q}$, we have $E_{\text{fa},\mathcal{Q}}^{f}\leq D_{\max}$. •

APPENDIX

A. Equivalent Definitions of UEP Exponents

We could have defined all the UEP exponents in this paper without using the notion of capacity-achieving sequences. As an example, in this appendix we define the single-bit exponent in this alternate manner and show that both definitions lead to identical results. In this alternative, first $\bar{E}_{\mathrm{b}}(R)$ is defined as the best exponent for the special bit at a given data rate R and then it is minimized over all $R<C$ to obtain $\bar{E}_{\mathrm{b}}$.

Definition 14: For any $R\geq 0$, $\mathcal{Z}(R)$ is the set of sequence of codes $\mathcal{Q}$, with message sets $\mathcal{M}^{(n)}$ such that

$$|\mathcal{M}^{(n)}|\geq e^{Rn} \quad \text{and} \quad \mathcal{M}^{(n)}=\mathcal{M}_1\times\mathcal{M}_2^{(n)}$$

where $\mathcal{M}_1=\{0,1\}$.

Definition 15: For a sequence of codes, $\mathcal{Q}$, such that $\lim_{n\to\infty}\Pr^{(n)}\left[\hat{M}\neq M\right]=0$, a single-bit exponent $E_{\mathrm{b},\mathcal{Q}}$ equals

$$E_{\mathrm{b},\mathcal{Q}}\triangleq\liminf_{n\to\infty}\frac{-\ln\Pr^{(n)}\left[\hat{M}_1\neq M_1\right]}{n}. \quad (74)$$

Definition 16: $\bar{E}_{\mathrm{b}}(R)$ and the single-bit exponents $\bar{E}_{\mathrm{b}}$ are defined as

$$\bar{E}_{\mathrm{b}}(R)\triangleq\sup_{\mathcal{Q}\in\mathcal{Z}(R)}E_{\mathrm{b},\mathcal{Q}}$$

$$\bar{E}_{\mathrm{b}}\triangleq\inf_{R<C}\bar{E}_{\mathrm{b}}(R).$$

Note that according to this definition, the special bit can achieve the exponent $\bar{E}_{\mathrm{b}}$, no matter how close the rate is to capacity. We now show why this definition is equivalent to the earlier definition in terms of capacity-achieving sequences given in Section III.

Lemma 2: $\bar{E}_{\mathrm{b}}=E_{\mathrm{b}}$

Proof: $E_{\mathrm{b}}\leq\bar{E}_{\mathrm{b}}$:

By definition of E_{b}, for any given $\delta>0$, there exists a capacity-achieving sequence $\mathcal{Q}$ such that $E_{\mathrm{b}\mathcal{Q}}=E_{\mathrm{b}}$ and for large enough n, $R^{(n)}\geq C-\delta$. If we replace first n members of $\mathcal{Q}$ with codes whose rates are $(C-\delta)$ or higher we get another sequence $\mathcal{Q}'$ such that $\mathcal{Q}'\in\mathcal{Z}(C-\delta)$ where $E_{\mathrm{b}\mathcal{Q}'}=E_{\mathrm{b}}$. Thus, $\bar{E}_{\mathrm{b}}(C-\delta)\geq E_{\mathrm{b}}$ for all $\delta>0$. Consequently

$$\bar{E}_{\mathrm{b}}\geq E_{\mathrm{b}}.$$

$E_{\mathrm{b}}\geq\bar{E}_{\mathrm{b}}$:

Let us first fix an arbitrarily small $\delta>0$. In the table in Fig. 6, row k represents a code-sequence $\bar{\mathcal{Q}}_k\in\mathcal{Z}(C-1/k)$, whose single-bit exponent

$$E_{\mathrm{b},\bar{\mathcal{Q}}_k}\geq\bar{E}_{\mathrm{b}}(R)-\delta.$$

Let $\bar{\mathcal{Q}}_k(l)$ represent length-l code in this sequence. We construct a capacity-achieving sequence $\mathcal{Q}$ from this table by sequentially choosing elements of $\mathcal{Q}$ from rows $1,2,\ldots$ as follows.

- For each sequence $\bar{\mathcal{Q}}_i$, let n_i denote the smallest block length n at which
  1) the single bit error probability satisfies
  $$\Pr^{(n)}\left[\hat{M}_1\neq M_1\right]\leq e^{-n(\bar{E}_{\mathrm{b}}(R)-2\delta))};$$
  2) the overall error probability satisfies
  $$\Pr^{(n)}\left[\hat{M}\neq M\right]\leq 1/i;$$
  3) $n_i\geq n_{i-1}$.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

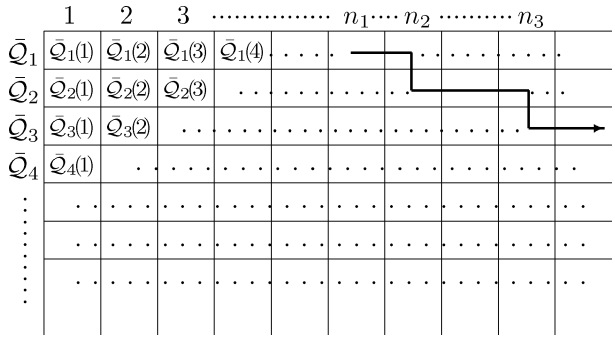


Fig. 6. Row k denotes a reliable code sequence at rate $C - 1/k$. Bold path shows capacity-achieving sequence $\mathcal{Q}$.

- Given the sequence $n_1, n_2, \ldots$, we choose the members of our capacity-achieving code from the code table shown in Fig. 6 as follows.
  - — **Initialize:** We use first $n_2 - 1$ members of $\bar{\mathcal{Q}}_1$ as the first $n_2 - 1$ members of the new code.
  - — **Iterate:** We choose codes of length n_i to $n_{i+1} - 1$ from the code sequence $\bar{\mathcal{Q}}_{i+1}$, i.e.,

$$\left(\bar{\mathcal{Q}}_i(n_i), \bar{\mathcal{Q}}_i(n_i + 1) \ldots, \bar{\mathcal{Q}}_i(n_{i+1} - 1)\right).$$

Thus, $\mathcal{Q}$ is a sampling of the code table as shown by the bold path in Fig. 6. Note that this choice of $\mathcal{Q}$ is a capacity-achieving sequence; moreover, it will also achieve a single-bit exponent

$$E_{\mathrm{b},\mathcal{Q}} = \inf_{R<C} \{\bar{E}_{\mathrm{b}}(R) - 2\delta\} = \bar{E}_{\mathrm{b}} - 2\delta.$$

Choosing arbitrarily small δ proves $E_{\mathrm{b}} \geq \bar{E}_{\mathrm{b}}$. •

ACKNOWLEDGMENT

The authors are indebted to Robert Gallager for his insights and encouragement for this work in general. In particular, Theorem 3 was mainly inspired from his remarks. Helpful discussions with David Forney and Emre Telatar are also gratefully acknowledged.

REFERENCES

[1] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, Jan. 1989.

[2] A. Albanese, J. Blomer, J. Edmonds, M. Luby, and M. Sudan, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.

[3] L. A. Bassalygo, V. A. Zinoviev, V. V. Zyablov, M. S. Pinsker, and G. S. Poltyrev, "Bounds for codes with unequal protection of two sets of messages," *Probl. Perd. Inform.*, vol. 15, no. 3, pp. 44–49, Jul.-Sep. 1979.

[4] P. Berlin, B. Nakibouglu, B. Rimoldi, and E. Telatar, "A simple converse of burnashev's reliability function," *IEEE Trans. Inf. Theory,* , vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

[5] S. Borade and S. Sanghavi, "Some fundamental coding theoretic limits of unequal error protection," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Seoul, Korea, Jul. 2009, pp. 2231–2235.

[6] S. Borade and L. Zheng, "Euclidean information theory," in *Proc. 45th Annu. Allerton Conf. Communications, Control and Computers*, Monticello, IL, Sep. 2007, pp. 633–640.

[7] I. Boyarinov and G. Katsman, "Linear unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 2, pp. 168–175, Mar. 1981.

[8] S. I. Bross and S. Litsyn, "Improved upper bounds for codes with unequal error protection," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3329–3333, Jul. 2006.

[9] M. V. Burnashev, "Data transmission over a discrete channel with feedback, random transmission time," *Probl. Perd. Inform.*, vol. 12, no. 4, pp. 10–30, 1976.

[10] A. R. Calderbank and N. Seshadri, "Multilevel codes for unequal error protection," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1234–1248, Jul. 1968.

[11] M. C. Thomas and A. T. Joy, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[12] I. Csiszár, "Joint source-channel error exponent," *Probl. Contr. Inf. Theory*, vol. 9, no. 5, pp. 315–328, 1980.

[13] C. Imre and K. János, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL: Academic, 1982.

[14] P. Cuff, "Communication requirements for generating correlated random variables," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 1393–1397.

[15] S. Diggavi and D. Tse, "On successive refinement of diversity," in *Proc. 42nd Annu. Allerton Conf. Communications, Control and Computing*, Monticello, IL, Sep./Oct. 2004, pp. 1641–1650.

[16] V. N. Dynkin and V. A. Togonidze, "Cyclic codes with unequal protection of symbols," *Probl. Perd. Inform.*, vol. 12, no. 1, pp. 24–28, Jan.-Mar. 1976.

[17] G. D. Forney, Jr., "On Exponential Error Bounds for Random Codes on the BSC,", unpublished.

[18] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 206–220, Mar. 1968.

[19] R. G. Gallager, "Fixed Composition Arguments and Lower Bounds to Error Probability," [Online]. Available: http://web.mit.edu/gallager/www/notes/notes5.pdf, unpublished

[20] G. G. Robert, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[21] W. J. van Gils, "Two topics on linear unequal error protection codes: Bounds and their lengths and cyclic code classes," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 866–876, Sep. 1983.

[22] C. Kilgus and W. Gore, "A class of cyclic unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 5, pp. 687–690, Sep. 1972.

[23] B. D. Kudryashov, "On message transmission over a discrete channel with noiseless feedback," *Probl. Perd. Inform.*, vol. 15, no. 1, pp. 3–13, 1973.

[24] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 3, pp. 600–607, Oct. 1967.

[25] A. Montanari and G. D. Forney, Jr., "On Exponential Error Bounds for Random Codes on the DMC," [Online]. Available: http://www.stanford.edu/~montanar/PAPERS/FILEPAP/dmc.ps, unpublished

[26] R. H. Morelos-Zaragoza and S. Lin, "On a class of optimal nonbinary linear unequal error protection codes for two sets of messages," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 196–200, Jan. 1994.

[27] F. Özbudak and H. Stichtenoth, "Constructing linear unequal error protection codes from algebraic curves," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1523–1527, Jun. 2003.

[28] H. Pishro-Nik, N. Rahnavard, and F. Fekri, "Nonuniform error correction using low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2702–2714, Jul. 2005.

[29] C. Poulliat, D. Declercq, and I. Fijalkow, "Optimization of ldpc codes for uep channels," in *Proc. IEEE Int. Symp. Information Theory, (ISIT)*, Chicago, IL, Jun./Jul. 2004, p. 450.

[30] C. Poulliat, D. Declercq, and I. Fijalkow, "Enhancement of unequal error protection properties of ldpc codes," *EURASIP J. Wirel. Commun. Netw.*, vol. 2007, no. 3, pp. 1–13, 2007.

[31] N. Rahnavard and F. Fekri, "Unequal error protection using low-density parity-check codes," in *Proc. IEEE Int. Symp. Information Theory (2004.)* , Chicago, IL, Jun./Jul. 2004, p. 449.

[32] N. Rahnavard, H. Pishro-Nik, and F. Fekri, "Unequal error protection using partially regular ldpc codes," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 387–391, Mar. 2007.

[33] A. Sahai and S. Mitter, "The Necessity and Sufficiency of Anytime Capacity for Control Over a Noisy Communication Link: Part II: Vector Systems," vol. arXiv:cs/0610146v2 [cs.IT] [Online]. Available: http://arxiv.org/abs/cs/0610146

[34] A. Sahai and S. Mitter, "Source Coding and Channel Requirements for Unstable Processes," vol. arXiv:cs/0610143v2 [cs.IT] [Online]. Available: http://arxiv.org/abs/cs/0610143

[35] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. Oct. 1948.

[36] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inf. Contr.*, vol. 10, no. 1, pp. 65–103, 1967.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.

[37] N. S. Albert, *Probability*. New York: Springer-Verlag, 1996.
[38] M. Trott, "Unequal error protection codes: Theory and practice," in *Proc. IEEE Information Theory Workshop*, Haifa, Israel, Jun. 1996.
[39] B. Vasić, A. Cvetković, S. Sankaranarayanan, and M. Marcellin, "Adaptive error protection low-density parity-check codes for joint source-channel coding schemes," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Lausanne, Switzerland, Jun./Jul. 2003, p. 267.
[40] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 729–733, Nov. 1979.

Shashibhushan Borade (S'01) received the B.Tech. degree in electrical engineering from Indian Institute of Technology, Bombay, India, in 2002. He received the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2004 and 2008.

His research interests include statistics, information theory, and finance theory. He is currently working as a quantitative researcher at D. E. Shaw and Co.

Mr. Borade received the Institute Silver Medal from IIT Bombay in 2002, the Presidential Fellowship from MIT in 2002, and the Hewlett-Packard Graduate Fellowship in 2005.

Barış Nakiboğlu received the B.S. degrees in electrical and electronics engineering and physics from Middle East Technical University (METU), Ankara, Turkey, in 2002 and the M.S. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2005.

He is currently working toward the Ph.D. degree at the Research Laboratory of Electronics (RLE) at MIT.

Lizhong Zheng (S'00–M'02) received the B.S and M.S. degrees, in 1994 and 1997, respectively, from the Department of Electronic Engineering, Tsinghua University, China, and the Ph.D. degree, in 2002, from the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley.

Since 2002, he has been working in the Department of Electrical Engineering and Computer Sciences at Massachusetts Institute of Technology, Cambridge, where he is currently the KDD Associate Professor. His research interests include information theory, wireless communications , and wireless networks.

Prof, Zheng received the Eli Jury award from the University of CCalifornia, Berkeley in 2002, the IEEE Information Theory Society Paper Award in 2003, and the NSF CAREER award in 2004. He is currently an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY.

Authorized licensed use limited to: MIT Libraries. Downloaded on February 11, 2010 at 14:44 from IEEE Xplore. Restrictions apply.