On Reliability Functions for Single-Message Unequal Error Protection

Da Wang
EECS Dept., MIT
Cambridge, MA, USA
Email: dawang@mit.edu

Venkat Chandar
Lincoln Laboratory, MIT
Lexington, MA, USA
Email: vchandar@mit.edu

Sae-Young Chung
EE Dept., KAIST
Daejeon, Korea
Email: sychung@ee.kaist.ac.kr

Gregory W. Wornell
EECS Dept., MIT
Cambridge, MA, USA
Email: gww@mit.edu

***Abstract*—Single-message unequal error protection (UEP) is a channel coding scheme that protects one special message differently from other (regular) messages. This induces three different types of errors in the system: 1) *miss* (where we decode the special codeword as a regular codeword), 2) *false alarm* (where we decode a regular codeword as the special codeword), and 3) *decoding error* (where we decode a regular codeword to another regular codeword). In this paper, we investigate the fundamental limits of single-message UEP, in the context of discrete memoryless channels (DMCs) without feedback. Similar to Borade *et al.*, we use error exponents as the performance metric, and discuss maximizing the miss error exponent and the false alarm error exponent, respectively. We provide a new converse proof for the miss reliability function, i.e., the optimal miss error exponent as a function of communication rate, and extend the inner and outer bound results for the false alarm reliability function in Borade *et al.* from rates close to capacity to all rates up to capacity.**

I. Introduction

Classical information theory assumes all messages are equally important and protects them uniformly. However, in certain communication scenarios it may be advantageous to relax this uniformity assumption and to protect certain information more than others, which leads to the framework of unequal error protection (UEP). In this paper we investigate the fundamental limits of single-message unequal error protection, the problem of protecting a *special codeword* differently from other *regular codewords* in the codebook.

This codeword classification leads naturally to three different types of errors in the system: 1) *miss* (where we decode the special codeword as a regular codeword), 2) *false alarm* (where we decode a regular codeword as the special codeword), and 3) *decoding error* (where we decode a regular codeword to another regular codeword). We denote the probabilities of these three events by P_{m}, P_{f} and P_{d} respectively. In most applications, we are interested in the regime that P_{m}, P_{f} and P_{d} can be made arbitrarily small by increasing the codeword block length n. Given this reliability condition, we can define the error exponents e_{m}, e_{f} and e_{d} for these three probabilities, which tell us asymptotically how fast these error probabilities decay as the codeword block length increases. Specifically, we investigate the maximal attainable e_{m} without constraining e_{f} and e_{d}, and similarly, the maximal attainable e_{f} without constraining e_{m} and e_{d}. These two values correspond to the strongest protection for the special codeword and that for regular codewords respectively, and are denoted by E_{m} and E_{f}.

This work was supported in part by NSF under Grant No. CCF-1017772, by AFOSR under Grant No. FA9550-11-1-0183, and by MKE under Grant No. NIPA-2012-(C1090-1211-0005).

A. Related work

Some earlier information-theoretic analysis of single-message UEP is motivated by communication with delay requirements, such as distributed control or streaming media applications. In this setting, going beyond block coding, i.e., allowing non-blocking encoding schemes, improves performance. For example, one of the earliest work [1] uses a single-message UEP code in a variable-delay communication setting. In [2], the authors consider delay in streaming communication with noisy feedback, and use the special message in UEP as an NACK signal. This achieves an error exponent that is much higher than the traditional channel coding error exponent. In a related work [3], error probability of the special message is interpreted as "the best-case probability of (undetected) block error across messages" and the miss exponent (called *hallucination* exponent therein) "gives an upper bound on the maximum probability of bit error in a non-blocking streaming context where noiseless feedback is available and the destination is (occasionally) allowed to declare erasures." There the authors characterize the miss exponent E_{m} for the binary symmetric channel (BSC) at all rates up to capacity.

Another example application of UEP is frame synchronization. In [4], a slotted asynchronous channel model is proposed, and the channel output induced by noise can be viewed as that induced by a special codeword with repeated symbols. This corresponds to a UEP problem with a design constraint on the special message.

Finally, [5] provides a detailed introduction to the information theoretic perspectives of UEP and investigates the fundamental limits of UEP on the discrete memoryless channel (DMC). The authors characterize the miss exponent (*red-alert exponent* therein) at capacity and provide lower and upper bounds for the false alarm error exponent at capacity. In addition, the miss error exponent for DMC at all rates up to capacity is mentioned after [5, Lemma 1], and a proof is given in [6, Theorem 34]. This result also follows from [7, Lemma 1,Lemma 5]. Recently, [8] characterizes the miss error exponent for the AWGN channel with both peak and average power constraints at all rates up to capacity.

In this paper, we further investigate the problem of single-message unequal error protection for the general DMC in the absence of feedback. We provide a new converse proof for the miss error exponent at all rates up to capacity. Furthermore, we provide lower and upper bounds for the false alarm error exponent at all rates up to capacity, and both bounds match the results at capacity in [5]. All results are obtained via a few generalizations of standard results in the method of types [9].

II. Notations and Problem Formulation

A. Notations

This paper uses lower case letters (e.g., x) to denote a particular value of the corresponding random variable, which is denoted in capital letters (e.g., X). A vector is denoted by its length as a superscript (e.g., x^n). Calligraphic fonts (e.g., $\mathcal{X}$) represent a set and $\mathcal{P}(\mathcal{X})$ denotes all probability distributions on the alphabet $\mathcal{X}$. $\mathbb{Z}_+$ and $\mathbb{Z}^+$ denote the set of non-negative integers and the set of positive integers respectively. We define $\dot{\leq}$ in the exponential sense, i.e., for a sequence a_n, $a_n \dot{\leq} e^{nF} \Leftrightarrow F \geq \lim_{n\to\infty} \frac{1}{n}\log a_n$, and we define $\dot{\geq}$ and $\doteq$ similarly.

We define the index set of a certain element a in a sequence x^n by

$$\mathcal{I}_a(x^n) \triangleq \{i : x_i = a\},$$

and we may denote this simply by $\mathcal{I}_a$ when the sequence is clear from context. The subsequence of x^n formed from index set $\mathcal{I}_a$ is denoted by $x_{\mathcal{I}_a}$.

We define the *shell* of a probability distribution as

$$[P]_\delta \triangleq \{P' \in \mathcal{P}(\mathcal{X}) : \mathrm{Support}(P') \subset \mathrm{Support}(P) \\ \text{and } \|P - P'\|_\infty < \delta\},$$

where $\|P(\cdot)\|_\infty \triangleq \max_{x\in\mathcal{X}} |P(x)|$ is the infinity norm.

For distributions $P(\cdot) \in \mathcal{P}(\mathcal{X})$, $Q(\cdot) \in \mathcal{P}(\mathcal{Y})$ and conditional distributions $W(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, $V(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, define:

$$\begin{aligned}
[P \times W](x,y) &\triangleq W(y\,|\,x)\,P(x),\\
PW(y) &\triangleq \sum_{x\in\mathcal{X}} W(y\,|\,x)\,P(x),\\
H(W|P) &\triangleq \sum_{x\in\mathcal{X}} P(x) H(W(\cdot\,|\,x)),\\
I(P,W) &\triangleq \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} P(x)W(y|x)\log\frac{W(y|x)}{PW(y)},\\
D(P\,\|\,Q) &\triangleq \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)},\\
D(V\,\|\,W|P) &\triangleq \mathbb{E}_P[D(V(\cdot\,|\,P)\,\|\,W(\cdot\,|\,P))]\\
&\triangleq \sum_{x\in\mathcal{X}} P(x) D(V(\cdot\,|\,x)\,\|\,W(\cdot\,|\,x)),\\
D(Q\,\|\,W|P) &\triangleq \mathbb{E}_P[D(Q\,\|\,W(\cdot\,|\,P))]\\
&\triangleq \sum_{x\in\mathcal{X}} P(x) D(Q\,\|\,W(\cdot\,|\,x)).
\end{aligned}$$

where $D(V\,\|\,W|P)$ is the conditional information divergence between $V(\cdot|\cdot)$ and $W(\cdot|\cdot)$ under $P(\cdot)$, and $D(Q\,\|\,W|P)$ is the conditional information divergence between $Q(\cdot)$ and $W(\cdot|\cdot)$ under $P(\cdot)$.

Our proofs make use of the method of types [9] and follow the notations therein. Specifically, the *type* of a sequence x^n with length n is denoted by $\hat{P}_{x^n}$, where the type is the empirical distribution of this sequence, i.e., $\hat{P}_{x^n}(a) = N(a|x^n)/n, \forall a \in \mathcal{X}$, where $N(a|x^n)$ is the number of occurrences of symbol a in sequence x^n. A *type class* $\mathcal{T}_P^n$ is the set of sequences that have type P; a *typical set* $\mathcal{T}_{[P]_\delta}^n$ is the set of sequences that have type $P' \in [P]_\delta$; and a *typical shell* $\mathcal{T}_{[W]_\delta}^n(x^n)$ is the set of sequences y^n such that

$$\left|\hat{P}_{x^n,y^n}(a,b) - \hat{P}_{x^n}(a)W(b|a)\right| \leq \delta.$$

A *constant composition code* is a code whose codewords all have the same type, i.e., lie in the same type class.

Finally, we extend the notion of typical shell and define the *tightly typical shell* with respect to a sequence x^n as

$$\mathring{\mathcal{T}}_{[W]_\delta}^n(x^n) \triangleq \left\{y^n \in \mathcal{Y}^n : y_{\mathcal{I}_a} \in \mathcal{T}_{[W]_\delta}^{n_a}(x_{\mathcal{I}_a}), \quad \forall a \in \mathcal{X}\right\}, \tag{1}$$

where $\mathcal{I}_a = \mathcal{I}_a(x^n)$.

B. Problem formulation

We follow the standard definition of the discrete memoryless channel (DMC) $W : \mathcal{X} \to \mathcal{Y}$, which has 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 $x \in \mathcal{X}$ is denoted by $W_{Y|X}(\cdot|x)$:

$$W_{Y|X}(y|x) = \mathbb{P}[Y = y\,|\,X = x] \qquad \forall\, x \in \mathcal{X},\ \forall\, y \in \mathcal{Y}.$$

When the input and output alphabets are clear from context, W is used instead of $W_{Y|X}$. Also, for simplicity, we denote $W_x \triangleq W(\cdot\,|\,x)$.

In this paper, we only consider the case that the support of $W(\cdot\,|\,x)$ is $\mathcal{Y}$ for all $x \in \mathcal{X}$, i.e., for any $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, $W(y\,|\,x) > 0$, because the analysis for other cases is either simple or can be reduced to this case.

For a DMC, a length n block code with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and some finite message set $\mathcal{M}_{f_n} = \{0, 1, 2, \ldots, |\mathcal{M}_{f_n}|\}$ is composed of a pair of mappings, encoder mapping $f_n : \mathcal{M}_{f_n} \to \mathcal{X}^n$ and decoder mapping $g_n : \mathcal{Y}^n \to \mathcal{M}_{f_n}$. Given a message m, the encoder maps it to a sequence $x^n(m) \in \mathcal{X}^n$ and transmits this sequence through the channel, where we call $x^n(m)$ the *codeword* for message m and the entire set of codewords $\{x^n(m), m \in \mathcal{M}_{f_n}\}$ a *codebook*. The receiver receives a sequence $y^n \in \mathcal{Y}^n$, where $W^n(y^n|\,x^n(m)) \triangleq \prod_{i=1}^n W(y_i\,|\,x_i(m))$. We denote (f_n, g_n) by $\mathcal{C}^{(n)}$.

In the context of unequal error protection, we use message 0 to denote the special message and refer to the collection of regular codewords as the *regular codebook*. We use $\mathcal{A}_n \triangleq \cup_{m\neq 0} g^{-1}(m)$ to denote the decoding region for regular codewords and use $\mathcal{B}_n \triangleq g_n^{-1}(m = 0)$ to denote decoding region for the special codeword. If $y^n \in \mathcal{A}_n$, we consider the channel input to be a regular codeword $x^n(m), m \in \{1, 2, \cdots, |\mathcal{M}_{f_n}|\}$, otherwise we consider the channel input to be the special message $m = 0$.

A code $\hat{\mathcal{C}}^{(n)} = (\hat{f}_n, \hat{g}_n)$ is called a *subcode* of $\mathcal{C}^{(n)} = (f_n, g_n)$ if $\mathcal{M}_{\hat{f}_n} \subset \mathcal{M}_{f_n}$, $\hat{f}_n(0) = f_n(0)$ and for any $m \in \mathcal{M}_{\hat{f}_n}$ and $m \neq 0$, there is some $m' \in \mathcal{M}_{f_n}$ and $m' \neq 0$ such that $\hat{f}_n(m) = f_n(m')$.

The performance of a codebook over a channel W can be characterized by three types of error: 1) *miss*, where we decode a special codeword as a regular codeword; 2) *false alarm*, where we decode a regular codeword as a special codeword; 3) *decoding error*, where we decode a regular codeword incorrectly into another regular codeword. Formally, we define

$$\begin{aligned} P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) &\triangleq W^n\left(\mathcal{A}_n \,|\, x^n(0)\right), \\ P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) &\triangleq \max_{m\neq 0} P_{\mathrm{f}}(m) \triangleq \max_{m\neq 0} W^n\left(\mathcal{B}_n \,|\, x^n(m)\right), \\ P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right) &\triangleq \max_{m\neq 0} P_{\mathrm{d}}(m) \\ &\triangleq \max_{m\neq 0} \sum_{\hat{m}\neq m, \hat{m}\neq 0} W^n\left(g_n^{-1}(\hat{m}) \middle| f_n(m)\right). \end{aligned}$$

In addition, we define the rate of a code $\mathcal{C}^{(n)}$ as $R(\mathcal{C}^{(n)}) \triangleq \log|\mathcal{M}_{f_n}|/n$. For a sequence of codebooks $\mathcal{Q} = \{\mathcal{C}^{(n)}, n \in \mathbb{Z}_+\}$, we define its rate as $R_{\mathcal{Q}} = \liminf_{n\to\infty} R(\mathcal{C}^{(n)})$. When the codebook sequence is clear from context, we denote $P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right)$, $P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right)$ and $P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right)$ by $P_{\mathrm{d}}^{(n)}$, $P_{\mathrm{m}}^{(n)}$ and $P_{\mathrm{f}}^{(n)}$, and use R_n and R instead of $R(\mathcal{C}^{(n)})$ and $R_{\mathcal{Q}}$.

In this paper, unless otherwise mentioned, all rates are between 0 and the channel capacity $C(W) \triangleq \max_{P_X\in\mathcal{P}(\mathcal{X})} I(P_X, W)$.

Definition 1 (Reliable codebook sequence)**.** *A codebook sequence $\mathcal{Q} = \{\mathcal{C}^{(n)}, n \in \mathbb{Z}^+\}$ is called* reliable *if*

$$\limsup_{n\to\infty} P_{\mathrm{d}}^{(n)} = 0,$$
$$\limsup_{n\to\infty} P_{\mathrm{m}}^{(n)} = 0,$$
$$\limsup_{n\to\infty} P_{\mathrm{f}}^{(n)} = 0.$$

To investigate how the above error probabilities decay with the codeword block length n, we define their error exponents.

Definition 2 (Miss, false alarm, and decoding error exponents)**.** *Given a reliable codebook sequence $\mathcal{Q} = \{\mathcal{C}^{(n)}, n \in \mathbb{Z}^+\}$ with rate R and a DMC $(\mathcal{X}, \mathcal{Y}, W)$, define its* miss error exponent *as*

$$e_{\mathrm{m}}(\mathcal{Q}) \triangleq \liminf_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{m}}^{(n)}$$

Similarly, we define its false alarm error exponent e_{f} *and* decoding error exponent e_{d} *in terms of $P_{\mathrm{f}}^{(n)}$ and $P_{\mathrm{d}}^{(n)}$.*

A triplet of numbers $(e_{\mathrm{m}}(\mathcal{Q}), e_{\mathrm{f}}(\mathcal{Q}), e_{\mathrm{d}}(\mathcal{Q}))$ is called achievable *if they can be achieved simultaneously, and the set of achievable triples is denoted by $\mathcal{E}(\mathcal{Q}, R)$. In addition, we define $\mathcal{E}(R) \triangleq \cup_{\mathcal{Q}:R_{\mathcal{Q}}\geq R}\mathcal{E}(\mathcal{Q}, R)$.*

The most general problem—characterization of the achievable error exponent region $\mathcal{E}(R)$—is open and this paper focuses on the false alarm and miss errors by putting no constraints on e_{d}. In particular, this paper investigates the reliability functions for false alarm and miss errors:

Definition 3 (Reliability functions)**.** *For a DMC $(\mathcal{X}, \mathcal{Y}, W)$, given a rate R, we define the* miss reliability function *and the* false alarm reliability function *as*

$$E_{\mathrm{m}}(R) \triangleq \sup_{(e_{\mathrm{m}}, e_{\mathrm{f}}\geq 0, e_{\mathrm{d}}\geq 0)\in\mathcal{E}(R)} e_{\mathrm{m}},$$
$$E_{\mathrm{f}}(R) \triangleq \sup_{(e_{\mathrm{m}}\geq 0, e_{\mathrm{f}}, e_{\mathrm{d}}\geq 0)\in\mathcal{E}(R)} e_{\mathrm{f}}.$$

III. Miss reliability function

This section presents a characterization of the miss reliability function in Theorem 1. While this result is not new, we present it for completeness, and provide a new, conceptually simpler, converse proof based on a few generalizations of the method of types.

Theorem 1 (Miss reliability function, Theorem 34 in [6])**.** *A DMC $(\mathcal{X}, \mathcal{Y}, W)$ has miss reliability function*

$$\begin{aligned} &E_{\mathrm{m}}(R) \\ &= \max_{P_{X,S}:\mathbb{E}_{P_S}[I(P_{X|S}, W)]=R} \left[\sum_{b\in\mathcal{X}} P_S(b) D\left(P_{Y|S=b} \,\|\, W_b\right)\right], \end{aligned} \tag{2}$$

where $P_{Y|S=b} \triangleq P_{X|S=b}W$.

Specializing Theorem 1 to $R = C$, we have the miss error exponent at capacity.

Corollary 2 (Miss reliability function at capacity)**.** *Let P_Y^* be the (unique) capacity-achieving output distribution, then*

$$E_{\mathrm{m}}(R = C) = \arg\max_{x\in\mathcal{X}} D\left(P_Y^* \,\|\, W_x\right). \tag{3}$$

The converse argument to Theorem 1 is implied by a converse proof for UEP with feedback in [6]. Below we present the sketch of an alternative converse proof, based on a few generalizations of the method of types.

In the method of types, it is known that for every reliable channel code, there exists a constant composition subcode with essentially the same rate [9, Collrollary 6.3 and Exercise 6.19]. Lemma 3 below shows that a similar argument holds when we replace constant composition code by a code whose subsequences are constant composition.

Lemma 3. *For any $\lambda > 0$, if a code (f_n, g_n) satisfies*

$$|\mathcal{M}_{f_n}| \geq e^{n(R-\lambda)} \tag{4}$$

and $P_{\mathrm{d}} < \varepsilon$ for a given channel W, then for any given s^n with type P_S and $\mathcal{I}_b \triangleq \mathcal{I}_b(s^n)$, there exists a collection of types $\{P_b, b \in \mathcal{X}\}$ and a subcode $(\hat{f}_n, \hat{g}_n)$ with all regular codewords in the set

$$\hat{\mathcal{C}} \triangleq \left\{x^n : x_{\mathcal{I}_b} \in \mathcal{T}_{P_b}^{n_b}\right\},$$

where such that

$$\left|\mathcal{M}_{\hat{f}_n}\right| \geq \exp\{n[R-2\lambda]\} \tag{5}$$

and

$$\sum_{b\in\mathcal{X}} P_S(b) I(P_b, W) \geq R - 3\lambda. \tag{6}$$

when n sufficiently large.

The proof of Lemma 3 is omitted due to space constraint.

Lemma 3 enables us to focus the converse argument on the (more structured) subcode $\hat{\mathcal{C}}$ that satisfies

$$x_{\mathcal{I}_b} \in \mathcal{T}_{P_b}^{n_b}. \tag{7}$$

Define a collection of disjoint sets $\{\mathcal{F}_m\}$, where $\mathcal{F}_m \triangleq g_n^{-1}(m) \cap \mathring{\mathcal{T}}_{[W]_\delta}^n (x^n(m))$ (see (1)). It can be shown that for any $\varepsilon > 0$, when n sufficiently large,

$$W^n(\mathcal{F}_m | x^n(m)) \geq 1 - 2\varepsilon. \tag{8}$$

Now we generalize [9, Lemma 2.14] to the subcode in Lemma 3 and obtain the following lemma:

Lemma 4. *Given a sequence s^n with type P_S and $\mathcal{I}_b \triangleq \mathcal{I}_b(s^n)$, if a set $\mathcal{B} \subset \mathcal{Y}^n$ satisfies*

$$W^n(\mathcal{B}| x^n) \geq \eta,$$

then for any $\tau > 0$

$$|\mathcal{B}| \geq \exp\left\{\left[\sum_{b\in\mathcal{X}} n_b H\left(W|\hat{P}_{x_{\mathcal{I}_b}}\right) - \tau\right]\right\}.$$

The proof of Lemma 4 is omitted due to space constraint.

Noting (7) and (8), Lemma 4 indicates

$$|\mathcal{F}_m| \dot{\geq} \exp\left\{\left[\sum_{b\in\mathcal{X}} n_b H(W|P_b)\right]\right\}.$$

Note that for any $y^n \in \mathcal{F}_m \subseteq \mathcal{T}_{[W]_\delta}^n (x^n(m))$,

$$y_{\mathcal{I}_b} \in \mathcal{T}_{[P_b W]_{\delta'}}^n,$$

and

$$\begin{aligned} W^n(y^n | s^n) &= \prod_{b\in\mathcal{X}} W^{n_b}(y_{\mathcal{I}_b} | s_{\mathcal{I}_b}) = \prod_{b\in\mathcal{X}} W_b^{n_b}(y_{\mathcal{I}_b}) \\ &\dot{\geq} \prod_{b\in\mathcal{X}} e^{\{-n_b[D(P_b W \| W_b) + H(P_b W)]\}}. \end{aligned}$$

Therefore,

$$\begin{aligned} P_{\mathrm{m}}\left(\hat{\mathcal{C}}\right) &= W^n(\mathcal{A}_n | s^n) \geq W^n(\cup_m \mathcal{F}_m | s^n) \\ &\dot{\geq} e^{-n\left[\sum_b P_S(b) \tilde{E}_{\mathrm{m}}(P_b, R, W_b)\right]}, \end{aligned}$$

where

$$\tilde{E}_{\mathrm{m}}(P, R, W_x) \triangleq D(PW \| W_x) + I(P, W) - R. \tag{9}$$

Note that using a subcode reduces P_{m}, hence any lower bound to $P_{\mathrm{m}}\left(\hat{\mathcal{C}}\right)$ is a lower bound to $P_{\mathrm{m}}(\mathcal{C})$. Finally, we let $P_{X|S=b} \triangleq P_b$ and maximize over $P_{X,S}$ to complete the converse argument.

IV. False alarm reliability function

In this section we investigate the false alarm reliability function of a DMC $(\mathcal{X}, \mathcal{Y}, W)$. We start with the case of a special codeword consisting of a repeated symbol, then extend the results to an unconstrained special codeword. Unlike the case of miss detection, the exact characterization of the false alarm reliability function is open, but we are able to provide inner and outer bounds.

Theorem 5 (Bounds for the false alarm reliability function)**.** *The false alarm reliability function of an DMC $(\mathcal{X}, \mathcal{Y}, W)$ satisfies*

$$\underline{E_{\mathrm{f}}}(R) \leq E_{\mathrm{f}}(R) \leq \overline{E_{\mathrm{f}}}(R),$$

where

$$\begin{aligned} \underline{E_{\mathrm{f}}}(R) \triangleq & \max_{P_{X,S}: \mathbb{E}_{P_S}\left[I\left(P_{X|S}, W\right)\right] \geq R} \left[\sum_{b\in\mathcal{X}} P_S(b) \right. \\ & \left. \cdot \min_{V: P_{X|S=b} V = W_b} \sum_{a\in\mathcal{X}} P_{X|S=b}(a) D(V_a \| W_a)\right], \end{aligned} \tag{10}$$

$$\begin{aligned} \overline{E_{\mathrm{f}}}(R) \triangleq & \max_{P_{X,S}: \mathbb{E}_{P_S}\left[I\left(P_{X|S}, W\right)\right] \geq R} \\ & \sum_{a,b\in\mathcal{X}} P_{X,S}(a,b) D(W_b \| W_a). \end{aligned} \tag{11}$$

Specializing Theorem 5 to $R = C$, we obtain bounds for the false alarm error exponent at capacity.

Corollary 6 (Bounds for the false alarm reliability function at capacity)**.** *Let the set of capacity-achieving input distributions be $\Pi = \{P_X : I(P_X, W) = C\}$, then*

$$\underline{E_{\mathrm{f}}}(R = C) = \max_{P_X^* \in \Pi} \max_{b\in\mathcal{X}} \min_{V: P_X^* V = W_b} D(V \| W | P_X^*), \tag{12}$$

$$\overline{E_{\mathrm{f}}}(R = C) = \max_{P_X^* \in \Pi} \max_{b\in\mathcal{X}} \sum_{a\in\mathcal{X}} P_X^*(a) D(W_b \| W_a). \tag{13}$$

Remark 4. *Corollary 6 agrees with [5, Theorem 10], which specifies the false alarm error exponent at capacity. Therefore, Theorem 5 generalizes [5, Theorem 10] from at capacity to all rates up to capacity.*

A. Special codeword with a repeated symbol

This section assumes the special codeword to be $\star^n$ and investigates the corresponding performance, which is summarized in Theorem 7.

Theorem 7 (False alarm reliability function for special codeword with repeated symbol $\star$)**.** *Given a special codeword $\star^n$ where $\star \in \mathcal{X}$, the false alarm reliability function of an DMC $(\mathcal{X}, \mathcal{Y}, W)$ satisfies*

$$\underline{E_{\mathrm{f}}^{\mathrm{rep}}}(R) \leq E_{\mathrm{f}}^{\mathrm{rep}}(R) \leq \overline{E_{\mathrm{f}}^{\mathrm{rep}}}(R),$$

where

$$\begin{aligned} \mathcal{V}_\star &\triangleq \{V : P_X V = W_\star\}, \\ \underline{E_{\mathrm{f}}^{\mathrm{rep}}}(R) &\triangleq \max_{P_X: I(P_X, W) \geq R} \min_{V\in\mathcal{V}_\star} D(V \| W | P_X), \\ \overline{E_{\mathrm{f}}^{\mathrm{rep}}}(R) &\triangleq \min_{P_X: I(P_X, W) \geq R} D(W_\star \| W | P_X). \end{aligned}$$

The detailed proofs are omitted due to space limits and a sketch of the proof is provided in [4], where this corresponds to finding the miss error exponent in asynchronous communication.

B. Unconstrained special codeword

For the case of unconstrained special codeword, the inner and outer bounds of the false alarm reliability function are given in Theorem 5. Omitting the detailed proof due to space limits, here we provide proof sketches of the outer bound argument and the achievability scheme for the inner bound derivation, which is constructed via the achievability scheme for the repeated-symbol special message.

1) Achievability: To achieve (10), we choose a special codeword s^n with type P_S and let $\mathcal{I}_b = \mathcal{I}_b(s^n)$, and accordingly a set of distributions $\{P_b \in \mathcal{P}_{n_b}(\mathcal{X}), b \in \mathcal{X}\}$ such that

$$\sum_{b\in\mathcal{X}} P_S(b) I(P_b, W) \geq R.$$

Then for each $b \in \mathcal{X}$, we follow Theorem 7 to construct a codebook $\mathcal{C}_b = (f_b, g_b)$ that is constant composition with type P_b and rate $I(P_b, W)$. Using these codebooks as building blocks, we design the encoding and decoding schemes below.

Denote each message m by a $|\mathcal{X}|$-tuple $(i_1, i_2, \ldots, i_{|\mathcal{X}|})$, where $i_b \in \{0, 1, 2, \ldots, |\mathcal{C}_b|\}, b \in \mathcal{X}$. Then we define the following encoding and decoding functions for $\mathcal{C}^{(n)} = (f, g)$:

$$\begin{aligned} &f\left((i_1, i_2, \ldots, i_{|\mathcal{X}|})\right) \\ &= \begin{cases} x^n \text{ with } x_{\mathcal{I}_b} = f_b(i_b) & \text{when } i_b > 0 \;\forall\, b \in \mathcal{X} \\ s^n & \text{when } i_b = 0 \;\forall\, b \in \mathcal{X} \end{cases}, \\ &g(y^n) = \left(\hat{i}_1, \hat{i}_2, \ldots, \hat{i}_{|\mathcal{X}|}\right) \text{ with } \hat{i}_b = g_b(y_{\mathcal{I}_b}) \;\forall\, b \in \mathcal{X}. \end{aligned}$$

This leads to the following error events

$$\begin{aligned} \mathcal{E}_{\mathrm{e}} &= \{\text{any } g_b \text{ reports an error }\}, \\ \mathcal{E}_{\mathrm{m}} &= \left\{g_b(y_{\mathcal{I}_b}) \neq 0, \exists\, b \in \mathcal{X} \mid i_1 = i_2 = i_{|\mathcal{X}|} = 0\right\}, \\ \mathcal{E}_{\mathrm{f}} &= \left\{g_b(y_{\mathcal{I}_b}) = 0, \forall\, b \in \mathcal{X} \mid i_1 i_2 \cdots i_{|\mathcal{X}|} \neq 0\right\}, \end{aligned}$$

with the corresponding error probabilities

$$\begin{aligned} &P_{\mathrm{d}}\left(\mathcal{C}^{(n)}\right) \leq \sum_{b\in\mathcal{X}} P_{\mathrm{d}}(C_b), P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) \leq \sum_{b\in\mathcal{X}} P_{\mathrm{m}}(C_b), \\ &P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) = \prod_{b\in\mathcal{X}} P_{\mathrm{f}}(C_b) \\ &\qquad \dot{\leq} \exp\left\{-n \sum_{b\in\mathcal{X}} P_S(b) \left[\min_{V\in\mathcal{V}_b} D(V \,\|\, W|P_b)\right]\right\}. \end{aligned}$$

Therefore, we can choose P_S and $\{P_b\}$ to maximize this expression. Finally, we construct a joint distribution $P_{X,S}$ where $P_{X|S=b} \triangleq P_b$ and obtain (10).

2) Outer bound argument: Following Lemma 3 in Section III, for any given s^n with type P_S and $\mathcal{I}_b \triangleq \mathcal{I}_b(s^n)$, we only need to show the E_{f} outer bound for a subcode $\hat{\mathcal{C}}$ that has essentially the same rate and $x_{\mathcal{I}_b} \in \mathcal{T}^{n_b}_{P_b}$ for $b \in \mathcal{X}$. Given an arbitrary codeword x^n and the special codeword s^n, we define $\mathcal{J}_{a,b} \triangleq \{i : x_i = a, s_i = b\}$, and consider the following set

$$\mathcal{G}(x^n, s^n) \triangleq \left\{y^n : y_{\mathcal{J}_{a,b}} \in \mathcal{T}^n_{[W_b]_\delta}\right\}.$$

We can show that for each $y^n \in \mathcal{G}(x^n, s^n)$, $W^n(y^n| x^n)$ are not too small:

$$\begin{aligned} W^n(y^n|x^n) &= \prod_{a,b\in\mathcal{X}} W^{n_{a,b}}\left(y_{\mathcal{J}_{a,b}} \middle| x_{\mathcal{J}_{a,b}}\right) \\ &= \prod_{a,b\in\mathcal{X}} W^{n_{a,b}}\left(y_{\mathcal{J}_{a,b}} \middle| a^{n_{a,b}}\right) = \prod_{a,b\in\mathcal{X}} W_a^{n_{a,b}}(y_{\mathcal{J}_{a,b}}) \\ &\dot{\geq} \prod_{a,b\in\mathcal{X}} \exp\left\{-n_{a,b}\left[D(W_b \,\|\, W_a) + H(W_b)\right]\right\} \\ &\doteq e^{\left\{-n\left[\sum_{a,b\in\mathcal{X}} P_{X,S}(a,b) D(W_b \,\|\, W_a) + H(W|P_S)\right]\right\}}, \end{aligned}$$

where $P_{X,S}$ is the joint type of (x^n, s^n). Furthermore, we show that $|\mathcal{B} \cap \mathcal{G}(x^n, s^n)|$ is not too small either. First,

$$W^n(\mathcal{B} \cap \mathcal{G}(x^n, s^n)| s^n) \geq \varepsilon/2$$

because $W^n(\mathcal{B}| s^n) \geq 1 - \varepsilon$ and

$$\begin{aligned} W^n(\mathcal{G}(x^n, s^n)| s^n) &= \sum_{y^n\in\mathcal{G}(x^n,s^n)} W^n(y^n| s^n) \\ &= \sum_{y^n\in\mathcal{G}(x^n,s^n)} \prod_{a,b\in\mathcal{X}} W^{n_{a,b}}\left(y_{\mathcal{J}_{a,b}} \middle| s_{\mathcal{J}_{a,b}}\right) \\ &= \prod_{a,b\in\mathcal{X}} W_b^{n_{a,b}}(\mathcal{T}^{n_{a,b}}_{[W_b]_\delta}) \geq 1 - \varepsilon/2, \end{aligned}$$

when n sufficiently large. Then applying [9, Lemma 2.14],

$$|\mathcal{B} \cap \mathcal{G}(x^n, s^n)| \;\dot{\geq}\; e^{nH(W|P_S)}.$$

Therefore,

$$\begin{aligned} P_{\mathrm{f}}\left(\hat{\mathcal{C}}\right) &= W^n(\mathcal{B}_n| x^n) \geq |\mathcal{G}(x^n, s^n) \cap \mathcal{B}_n| \cdot W^n(y^n| x^n) \\ &\dot{\geq} \exp\left\{-n\left[\sum_{a,b\in\mathcal{X}} P_{X,S}(a,b) D(W_b \,\|\, W_a)\right]\right\}. \end{aligned}$$

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for helpful comments.

REFERENCES

[1] B. D. Kudryashov, "On message transmission over a discrete channel with noiseless feedback," *Problemy Peredachi Informatsii*, vol. 15, no. 1, pp. 3–13, 1979.

[2] S. Draper and A. Sahai, "Beating Burnashev in delay with noisy feedback," in *Allerton Conference in Communication, Control and Computing*, 2006.

[3] A. Sahai and S. C. Draper, "The "hallucination" bound for the BSC," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, Toronto, ON, Canada, 2008, pp. 717–721.

[4] D. Wang, V. Chandar, S. Chung, and G. W. Wornell, "Error exponents for asynchronous communication," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, St. Petersburg, Russia, 2011, pp. 1071–1075.

[5] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An Information-Theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5511–5539, 2009.

[6] S. P. Borade, "When all information is not created equal," Thesis, Massachusetts Institute of Technology, 2008.

[7] S. Gorantla, B. Nakiboğlu, T. Coleman, and L. Zheng, "Bit-wise unequal error protection for variable length blockcodes with feedback," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, Jun. 2010, pp. 241–245.

[8] B. Nazer, Y. Shkel, and S. C. Draper, "The AWGN red alert problem," *1102.4411*, Feb. 2011. [Online]. Available: http://arxiv.org/abs/1102.4411

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, Aug. 2011.