Bit-Wise Unequal Error Protection for Variable-Length Block Codes With Feedback

Barış Nakiboğlu, Siva K. Gorantla, Lizhong Zheng, and Todd P. Coleman, *Senior Member, IEEE*

***Abstract*—The *bit-wise* unequal error protection problem, for the case when the number of groups of bits ℓ is fixed, is considered for variable-length block codes with feedback. An encoding scheme based on fixed-length block codes with erasures is used to establish inner bounds to the achievable performance for finite expected decoding time. A new technique for bounding the performance of variable-length block codes is used to establish outer bounds to the performance for a given expected decoding time. The inner and the outer bounds match one another asymptotically and characterize the achievable region of rate-exponent vectors, completely. The single-message *message-wise* unequal error protection problem for variable-length block codes with feedback is also solved as a necessary step on the way.**

***Index Terms*—Block codes, Burnashev's exponent, discrete memoryless channels (DMCs), error exponents, errors-and-erasures decoding variable-length block coding, feedback, Kudryashov's signaling, unequal error protection (UEP), variable-length communication, Yamamoto–Itoh scheme.**

I. Introduction

In the conventional formulation of digital communication problem, the primary concern is the correct transmission of the message; hence, there is no distinction between different error events. In other words, there is a tacit assumption that all error events are equally undesirable; incorrectly decoding to a message $\bar{\tilde{m}}$ when a message $\tilde{m}$ is transmitted is as undesirable as incorrectly decoding to a message $\bar{\tilde{m}}$ when a message $\tilde{\tilde{m}}$ is transmitted, for any $\bar{m}$ other than $\tilde{m}$ and $\bar{\tilde{m}}$ other than $\tilde{\tilde{m}}$. Therefore, the performance criteria used in the conventional formulation (minimum distance between codewords, maximum conditional error probability among messages, average error probability, etc.) are oblivious to any precedence order that might exist among the error events.

In many applications, however, there is a clear order of precedence among the error events. For example in Internet communication, packet headers are more important than the actual payload data. Hence, a code used for Internet communication can enhance the protection against the erroneous transmission of the packet headers at the expense of the protection against the erroneous transmission of payload data. In order to appreciate such a coding scheme, one may analyze error probability of the packet headers and error probability of payload data separately, instead of analyzing the error probability of the overall message composed of packet header and payload data. Such a formulation for Internet communication is an unequal error protection (UEP) problem, because of the separate calculation of the error probabilities of the parts of the messages.

Problems capturing the disparity of undesirability among various classes of error events, by assigning and analyzing distinct performance criteria for different classes of error events, are called UEP problems. UEP problems have already been studied widely by researchers in communication theory, coding theory, and computer networks from the perspectives of their respective fields. In this paper, we enhance the information theoretic perspective on UEP problems [2], [5] for variable-length block codes by generalizing the results of [2] to the rates below capacity.

In information theoretic UEP, error events are grouped into different classes and the probabilities associated with these different classes of error events are analyzed separately. In order to prioritize protection against one or the other class of error events, corresponding error exponent is increased at the expense of the other error exponents. There are various ways to choose the error event classes but two specific choices of error event classes stand out because of their intuitive familiarity and practical relevance; they correspond to the message-wise UEP and the bit-wise UEP. In the following, we first describe these two types of UEP and then specify the UEP problems we are interested in this paper.

In the *message-wise* UEP, the message set $\mathcal{M}$ is assumed to be the union of ℓ disjoint sets for some fixed ℓ, i.e., $\mathcal{M} = \cup_{j=1}^{\ell} \mathcal{M}_j$ where $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ for all $i \neq j$. For

Manuscript received January 10, 2011; revised April 23, 2012; accepted October 08, 2012. Date of publication November 16, 2012; date of current version February 12, 2013. This work was supported in part by the National Science Foundation (NSF) through the NSF Cyberphysical Systems Program NSF0932410, in part by the NSF Science and Technology Center under Grant CCF-0939370, and in part by the Defense Advanced Research Projects Agency's Information Theory of Mobile Ad-Hoc Networks program under Grant W911NF-07-1-0029.

B. Nakiboğlu is with the Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA 94720-1770 USA (e-mail: nakib@eecs.berkeley.edu).

S. K. Gorantla was with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with Adchemy Inc., Foster City, CA 94402 USA (e-mail: sgorant2@illinois.edu).

L. Zheng is with the Department of Electrical Engineering and Computer Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: lizhong@mit.edu).

T. P. Coleman was with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Department of Bioengineering, University of California at San Diego, La Jolla, CA 92093 USA (e-mail: tpcoleman@ucsd.edu).

Communicated by D. Guo, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2227671

each set $\mathcal{M}_j$, the maximum error probability[1] $P_{\mathrm{e}}\{j\}$, the rate $\mathrm{R}_{\{j\}}$, and the error exponent $\mathrm{E}_{\{j\}}$ are defined as the corresponding quantities defined in the conventional problem, i.e., $P_{\mathrm{e}}\{j\} = \max_{\mathsf{m}\in\mathcal{M}_j} \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{m} \middle| \mathsf{M} = \mathsf{m}\right]$, $\mathrm{R}_{\{j\}} = \frac{\ln|\mathcal{M}_j|}{n}$, $\mathrm{E}_{\{j\}} = \frac{-\ln P_{\mathrm{e}}\{j\}}{n}$, for all j in $\{1,2,\ldots,\ell\}$ where n is the length of the code. The ultimate aim is calculating the achievable region of rate vector error exponent vector pairs, $(\mathrm{R}_{\{\cdot\}},\mathrm{E}_{\{\cdot\}})$'s where[2] $\mathrm{R}_{\{\cdot\}} = (\mathrm{R}_{\{1\}},\mathrm{R}_{\{2\}},\ldots,\mathrm{R}_{\{\ell\}})$ and $\mathrm{E}_{\{\cdot\}} = (\mathrm{E}_{\{1\}},\mathrm{E}_{\{2\}},\ldots,\mathrm{E}_{\{\ell\}})$. The *message-wise* UEP problem was the first information theoretic UEP problem to be considered; it was considered by Csiszár in his work on joint source channel coding [5]. Csiszár showed that for any integer ℓ, block length n, and ℓ-dimensional rate vector $\mathrm{R}_{\{\cdot\}}$ such that $0 \leq \mathrm{R}_{\{j\}} \leq C$ for $j = 1,2,\ldots,\ell$, there exists a length n block code with message set $\mathcal{M} = \cup_{j=1}^{\ell}\mathcal{M}_j$ where $|\mathcal{M}_j| = e^{n(\mathrm{R}_{\{j\}}-\varepsilon_n)}$ such that the conditional error probability of each message in each $\mathcal{M}_j$ is less then $e^{-n(E_r(\mathrm{R}_{\{j\}})-\varepsilon_n)}$ where $E_r(\cdot)$ is the random coding exponent and ε_n converges to zero as n diverges.[3]

The *bit-wise* UEP problem is the other canonical form of the information theoretic UEP problems. In the *bit-wise* UEP problem, the message set $\mathcal{M}$ is assumed to be the Cartesian product of $\mathcal{M}_1,\mathcal{M}_2,\ldots,\mathcal{M}_\ell$ for some fixed ℓ, i.e., $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_\ell$. Thus, the transmitted message M and the decoded message $\widehat{\mathsf{M}}$ are given by $\mathsf{M} = (\mathsf{M}_1,\mathsf{M}_2,\ldots,\mathsf{M}_\ell)$ and $\widehat{\mathsf{M}} = (\widehat{\mathsf{M}}_1,\widehat{\mathsf{M}}_2,\ldots,\widehat{\mathsf{M}}_\ell)$, receptively. Furthermore, M_j's and $\widehat{\mathsf{M}}_j$'s are called the transmitted and decoded submessages, respectively. The error events of interest in the *bit-wise* UEP problem are the ones corresponding to the erroneous transmission of the submessages. The error probability $P_{\mathrm{e}}(j)$, rate R_j, and the error exponent E_j of submessages are given by $P_{\mathrm{e}}(j) = \mathbf{P}\left[\widehat{\mathsf{M}}_j \neq \mathsf{M}_j\right]$, $\mathrm{R}_j = \frac{\ln|\mathcal{M}_j|}{n}$, $\mathrm{E}_j = \frac{-\ln P_{\mathrm{e}}(j)}{n}$ for all j in $\{1,2,\ldots,\ell\}$ where n is the block length. As was the case in the *message-wise* UEP problem, the ultimate aim in the *bit-wise* UEP problem is determining the achievable region of the rate vector error exponent vector pairs[4] $(\vec{\mathrm{R}},\vec{\mathrm{E}})$. The formulation of Internet communication problem we have considered above, with packet header and payload data, is a *bit-wise* UEP problem with two submessages, i.e., with $\ell = 2$.

There is some resemblance in the definitions of *message-wise* and *bit-wise* UEP problems, but they have very different behavior in many problems. For example, consider the *message-wise* UEP problem and the *bit-wise* UEP problem with $\ell = 2$, $\mathcal{M}_1 = \{1,2\}$, and $\mathcal{M}_2 = \{3,4,\ldots,e^{n(C-\varepsilon_n)}\}$ for some ε_n that goes to zero as n diverges. It is shown in [2, Th. 1] that if $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ and $\mathbf{P}\left[\mathsf{M}_2 \neq \widehat{\mathsf{M}}_2\right] \leq \widetilde{\varepsilon}_n$ for some $\widetilde{\varepsilon}_n$ that goes to zero as n diverges, then[5] $\mathrm{E}_1 = 0$. Thus, in the *bit-wise* UEP problem, even a bit cannot have a positive error exponent. As result of [5, Th. 5], on the other hand, if $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$, we know that $\mathcal{M}_1$ can have an error exponent $\mathrm{E}_{\{1\}}$ as high as $E_r(0) > 0$, while having a small error probability for $\mathcal{M}_2$, i.e., $\max_{\mathsf{m}\in\mathcal{M}_2} \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{m} \middle| \mathsf{M} = \mathsf{m}\right] \leq \widetilde{\varepsilon}_n$ for some $\widetilde{\varepsilon}_n$ that goes to zero as n diverges. Thus, in the *message-wise* UEP problem, it is possible to give an error exponent as high as $E_r(0)$ to $\mathcal{M}_1$.

The *message-wise* and the *bit-wise* UEP problems cover a wide range of problems of practical interest. Yet, as noted in [2], there are many UEP problems of practical importance that are neither *message-wise* nor *bit-wise* UEP problems. One of our aims in studying the *message-wise* and the *bit-wise* UEP problems is gaining insights and devising tools for the analysis of those more complicated problems.

In the aforementioned discussion, the UEP problems are described for fixed-length block codes for the sake of simplicity. One can, however, easily define the corresponding problems for various families of codes: with or without feedback, fixed or variable length, by modifying the definitions of the error probability, the rate, and the error exponent appropriately. Furthermore, parameter ℓ representing the number of groups of bits or messages is assumed to be fixed in the aforementioned discussion for simplicity. However, both the *message-wise* and the *bit-wise* UEP problems can be defined for ℓ's that are increasing with block length n in fixed-length block codes and for ℓ's that are increasing with expected block length $\mathbf{E}[\mathsf{T}]$ in variable-length block codes. In fact, Csiszár's result discussed previously [5, Th. 5] is proved not only for constant ℓ but also for any ℓ_n sequence satisfying $\lim_{n\to\infty}\frac{\ln \ell_n}{n} = 0$.

In this paper, we consider two closely related UEP problems for variable-length block codes over a discrete memoryless channels (DMCs) with noiseless feedback: the *bit-wise* UEP problem and the single-message *message-wise* UEP problem.

1) In the *bit-wise* UEP problem, there are ℓ submessages each with different priority and rate. For all fixed values of ℓ, we characterize the tradeoff between the rates and the error exponents of these submessages by revealing the region of achievable rate vector, exponent vector pairs. For fixed ℓ, this problem is simply the variable-length code version of the previously described *bit-wise* UEP problem.
2) In the single-message *message-wise* UEP problem, we characterize the tradeoff between the exponents of the minimum and the average conditional error probability. Thus, this problem is similar to the previously described *message-wise* UEP problem for the case $\ell = 2$ and $\mathcal{M}_1 = \{1\}$. But unlike that problem, we work with variable-length codes and average conditional error prob-

[1]This formulation is called the missed detection formulation of the *message-wise* UEP problem in [2]. If $\mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{m} \middle| \mathsf{M} = \mathsf{m}\right]$ is replaced with $\mathbf{P}\left[\widehat{\mathsf{M}} = \mathsf{m} \middle| \mathsf{M} \neq \mathsf{m}\right]$, we get the false alarm formulation of the *message-wise* UEP problem. In this paper, we restrict our discussion to the missed detection problem and use *message-wise* UEP without any qualifications to refer to the missed detection formulation of the *message-wise* UEP problem.

[2]Here, ℓ is assumed to be a fixed integer. All rate-exponent vectors, achievable or not, are in the region of $\mathbf{R}^{2\ell}$ in which $\mathrm{R}_{\{j\}} \geq 0$ and $\mathrm{E}_{\{j\}} \geq 0$ for all $1 \leq j \leq \ell$. $\mathbf{R}^{2\ell}$ is the 2ℓ-dimensional real vector space with the norm $\left\|\vec{x}\right\| = \sup_j |x_j|$

[3]Csiszár proved the aforementioned result not only for the case when ℓ is constant for all n but also for the case when ℓ_n is a sequence such that $\lim_{n\to\infty}\frac{\ln \ell_n}{n} = 0$. See [5, Th. 5].

[4]Similar to the *message-wise* UEP problem discussed previously, in the current formulation of *bit-wise* UEP problem, we assume ℓ to be fixed. Thus, all rate-exponent vectors, achievable or not, are in region of $\mathbf{R}^{2\ell}$ in which $\mathrm{R}_j \geq 0$ and $\mathrm{E}_j \geq 0$ for all $1 \leq j \leq \ell$, by definition.

[5]The channel is assumed to have no zero probability transition.

ability rather than fixed-length codes and the maximum error probability.

The *bit-wise* UEP problem for fixed number of groups of bits, i.e., fixed ℓ, and the single-message *message-wise* UEP problem were first considered in [2], for the case when the rate is (very close to) the channel capacity; we solve both of these problems for all achievable rates.

In fact, in [2], single-message *message-wise* UEP problem is solved not only at capacity, but also for all the rates below capacity both for fixed-length block codes without feedback and for variable-length block codes with feedback, but only for case when overall error exponent is zero (see [2, Appendix D]). Recently, Wang *et al*. [11] put forward a new proof based on method of types for the same problem.[6] Nazer *et al*. [9], on the other hand, investigated the problem for fixed-length block codes on additive white Gaussian noise channels at zero error exponent and derived the exact analytical expression in terms of rate and power constraints.

Before starting our presentation, let us give a brief outline of this paper. In Section II, we specify the channel model and make a brief overview of stopping times and variable-length block codes. In Section III, we first present the single-message *message-wise* UEP problem and fixed ℓ version of the *bit-wise* UEP problem for variable-length block codes; then, we state the solutions of these two UEP problems. In Section IV, we present inner bounds for both the single-message *message-wise* UEP problem and the *bit-wise* UEP problem. In Section V, we introduce a new technique, Lemma 5, for deriving outer bounds for variable-length block codes and apply it to the two UEP problems we are interested in. Finally, in Section VI, we discuss the qualitative ramifications of our results in terms of the design of communication systems with UEP and the limitations of our analysis. The proofs of the propositions in Sections III–V are deferred to the Appendices.

II. Preliminaries

As it is customary, we use upper case letters, e.g., M, X, Y, T for random variables and lower case letters, e.g., m, x, y, t for their sample values.

We denote discrete sets by capital letters with calligraphic fonts, e.g., $\mathcal{M}, \mathcal{X}, \mathcal{Y}$ and power sets of discrete sets by $\wp(\cdot)$, e.g., $\wp(\mathcal{M})$, $\wp(\mathcal{X})$, $\wp(\mathcal{Y})$. In order to denote the set of all probability distributions on a discrete set, we use $P(\cdot)$, e.g., $P(\mathcal{M})$, $P(\mathcal{X})$, $P(\mathcal{Y})$.

Definition 1 (Total Variation): For any discrete set $\mathcal{Z}$ and for any $\mu_1, \mu_2 \in P(\mathcal{Z})$, the total variation $\Delta(\mu_1, \mu_2)$ is defined as

$$\Delta(\mu_1, \mu_2) = \tfrac{1}{2} \sum\nolimits_{\mathsf{z} \in \mathcal{Z}} |\mu_1(\mathsf{z}) - \mu_2(\mathsf{z})|. \tag{1}$$

We denote the indicator function by $\mathbb{1}_{\{\cdot\}}$, i.e., $\mathbb{1}_{\{\Gamma\}} = 1$ when event Γ happens $\mathbb{1}_{\{\Gamma\}} = 0$ otherwise. We denote the binary entropy function by $h(\cdot)$, i.e.,

$$h(s) \stackrel{\triangle}{=} -s \ln s - (1-s)\ln(1-s) \qquad \forall s \in [0,1]. \tag{2}$$

[6]In addition to their new proof in missed-detection problem [11, Th. 1], Wang *et al*. present a completely new result on the false-alarm formulation of the problem [11, Th. 5].

A. Channel Model

We consider a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and $|\mathcal{X}| - \text{by} - |\mathcal{Y}|$ transition probability matrix W. Each row of W corresponds to a probability distribution on $\mathcal{Y}$, i.e., $W_{\mathsf{x}} \in P(\mathcal{Y})$ for all $\mathsf{x} \in \mathcal{X}$. For the reasons that will become clear shortly, in Section II-D, we assume that $W_{\mathsf{x}}(\mathsf{y}) > 0$ for all $\mathsf{x} \in \mathcal{X}$ and $\mathsf{y} \in \mathcal{Y}$ and denote the smallest transitions probability by λ

$$\lambda \stackrel{\triangle}{=} \min_{\mathsf{x},\mathsf{y}} W_{\mathsf{x}}(\mathsf{y}) > 0. \tag{3}$$

The input and output letters at time τ, up to time τ, and between time τ_1 and τ_2 are denoted by X_τ, Y_τ, X^τ, Y^τ, $\mathsf{X}_{\tau_1}^{\tau_2}$, and $\mathsf{Y}_{\tau_1}^{\tau_2}$ respectively. DMCs are both memoryless and stationary; hence, the conditional probability of $\mathsf{Y}_\tau = \mathsf{y}$ given $(\mathsf{X}^\tau, \mathsf{Y}^{\tau-1})$ is given by

$$\mathbf{P}\left[\mathsf{Y}_\tau = \mathsf{y} \middle| \mathsf{X}^\tau, \mathsf{Y}^{\tau-1}\right] = W_{\mathsf{X}_\tau}(\mathsf{y}).$$

Definition 2 (Empirical Distribution): For any $\tau_2 \geq \tau_1$ and any sequence $\mathsf{z}_{\tau_1}^{\tau_2}$ such that $\mathsf{z}_j \in \mathcal{Z}$ for all $j \in [\tau_1, \tau_2]$, the empirical distribution $\mathrm{Q}_{\{\mathsf{z}_{\tau_1}^{\tau_2}\}}$ is given by

$$\mathrm{Q}_{\{\mathsf{z}_{\tau_1}^{\tau_2}\}}(\mathsf{z}) = \frac{1}{\tau_2 - \tau_1 + 1} \sum\nolimits_{\tau=\tau_1}^{\tau_2} \mathbb{1}_{\{\mathsf{z}_\tau = \mathsf{z}\}} \qquad \forall \mathsf{z} \in \mathcal{Z}. \tag{4}$$

Note that we replace $\mathsf{z}_{\tau_1}^{\tau_2}$ by $\mathsf{Z}_{\tau_1}^{\tau_2}$ when the empirical distribution $\mathrm{Q}_{\{\mathsf{Z}_{\tau_1}^{\tau_2}\}}(\mathsf{z})$ becomes a random variable for each $\mathsf{z} \in \mathcal{Z}$.

B. Stopping Times

Stopping times are central in the formal treatment of variable-length codes; it is not possible to define or comprehend variable-length codes without a solid understanding of stopping times. For those readers who are not already familiar with the concept of the stopping times, we present a brief overview in this section.

In order to make our presentation more accessible, we use the concept of power sets, rather than sigma fields in the definitions. We can do that only because the random variables we use to define stopping times are discrete random variables. In the general case, when the underlying variables are not necessarily discrete, one needs to use the concept of sigma fields instead of power set.

Let us start with introducing the concept of Markov times. For an infinite sequence of random variables $\mathsf{Z}_1, \mathsf{Z}_2, \ldots$, a positive, $integer^*$-valued[7] function T defined on $\mathcal{Z}^\infty$ is a Markov time, if for all positive integers τ, it is possible to determine whether $\mathsf{T} = \tau$ or not by considering Z^τ only, i.e., if $\mathbb{1}_{\{\mathsf{T}=\tau\}}$ is not only a function of Z^∞ but also a function of Z^τ for all positive integers τ. The formal definition is given in the following.

Definition 3 (Markov Time): Let Z_1^∞ be an infinite sequence of $\mathcal{Z}$-valued random variables Z_τ for $\tau \in \{1, 2, \ldots\}$ and T be a function of Z^∞ which takes values from the set $\{1, 2, \ldots, \infty\}$. Then, the random variable T is a Markov time with respect to Z^τ if

$$\{\mathsf{z}^\infty : \mathsf{T} = \tau \text{ if } \mathsf{Z}^\infty = \mathsf{z}^\infty\} \in \wp(\mathcal{Z}^\tau) \times \{\mathcal{Z}_{\tau+1}^\infty\} \qquad \forall \tau \in \{1, 2, \ldots\} \tag{5}$$

[7]$Integer^*$ is the set of all integers together with two infinities, i.e., $\{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$.

where $\wp(\mathcal{Z}^\tau) \times \{\mathcal{Z}_{\tau+1}^\infty\}$ is the Cartesian product of the power set of $\mathcal{Z}^\tau$ and the one element set $\{\mathcal{Z}_{\tau+1}^\infty\}$.

We denote Z_τ's from $\tau = 1$ to $\tau = \mathsf{T}$ by Z^T and their sample values by z^t. The set of all sample values of Z^T such that $\mathsf{T} = \tau$, on the other hand, is denoted by $\mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau$. We denote union of all $\mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau$'s for finite τ's by $\mathcal{Z}^{\mathsf{T}*}$ and the union of all $\mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau$'s by $\mathcal{Z}^\mathsf{T}$, i.e.,

$$\mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau \stackrel{\triangle}{=} \{\mathsf{z}^\tau : \mathsf{T} = \tau \text{ if } \mathsf{Z}^\tau = \mathsf{z}^\tau\} \quad \tau \in \{1, \ldots, \infty\} \tag{6a}$$

$$\mathcal{Z}^{\mathsf{T}*} \stackrel{\triangle}{=} \bigcup_{1 \le \tau < \infty} \mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau \tag{6b}$$

$$\mathcal{Z}^\mathsf{T} \stackrel{\triangle}{=} \mathcal{Z}^{\mathsf{T}*} \bigcup \mathcal{Z}_{\{\mathsf{T}=\infty\}}^\infty. \tag{6c}$$

For an arbitrary, positive, $integer^*$-valued function T of Z^∞, however, one cannot talk about Z^T, because the value of T can in principle depend on $\mathsf{Z}_{\mathsf{T}+1}^\infty$. For a Markov time T, however, the value of T does not depend on $\mathsf{Z}_{\mathsf{T}+1}^\infty$. That is why we can define Z^T, $\mathcal{Z}_{\{\mathsf{T}=\tau\}}^\tau$, $\mathcal{Z}^{\mathsf{T}*}$, and $\mathcal{Z}^\mathsf{T}$ for any Markov time T.

Given an infinite sequence of z_τ's, i.e., z^∞, either $\mathsf{z}^\infty \in \mathcal{Z}_{\{\mathsf{T}=\infty\}}^\infty$ or z^∞ has a unique subsequence z^τ that is in $\mathcal{Z}^{\mathsf{T}*}$.

In most practical situations, one is interested in Markov times that are guaranteed to have a finite value; those Markov times are called stopping times.

Definition 4 (Stopping Time): A Markov time T with respect to Z^τ is a stopping time iff $\mathbf{P}[\mathsf{T} < \infty]$.

Note that if T is a stopping time, then $\mathbf{P}\left[\mathsf{Z}^\mathsf{T} \in \mathcal{Z}^{\mathsf{T}*}\right] = 1$. Furthermore, unlike $\mathcal{Z}^\mathsf{T}$, $\mathcal{Z}^{\mathsf{T}*}$ is a countable set for all stopping times T because $|\mathcal{Z}|$ is finite.[8]

C. Variable-Length Block Codes

A variable-length block code on a DMC is given by a random decoding time T, an encoding scheme Φ, and a decoding rule Ψ satisfying $\mathbf{P}[\mathsf{T} < \infty] = 1$.

1) *Decoding time* T is a Markov time with respect to the receiver's observation Y^τ, i.e., given Y^τ, receiver knows whether $\mathsf{T} = \tau$ or not. Hence, T is a random quantity rather than a constant; thus, neither the decoder nor the receiver knows the value of T *a priori*. But as time passes, both the decoder and the encoder (because of feedback link) will be able to decide whether T has been reached or not, just by considering the current and past channel outputs.
2) *Encoding scheme* Φ is a collection of mappings that determines the input letter at time $(\tau + 1)$ for each message in the finite message set $\mathcal{M}$, for each $\mathsf{y}^\tau \in \mathcal{Y}^\tau$ such that $\mathsf{T} > \tau$
$$\Phi(\cdot, \mathsf{y}^\tau) : \mathcal{M} \to \mathcal{X} \qquad \forall \mathsf{y}^\tau : \mathsf{T} > \tau.$$
3) *Decoding rule* is a mapping from the set of output sequences y^τ such that $\mathsf{T} = \tau$ to the finite message set $\mathcal{M}$ which determines the decoded message, $\widehat{\mathsf{M}}$. With a slight abuse of notation, we denote the set of all, possibly infinite, output sequences y^τ such that $\{\mathsf{T} = \tau \text{ if } \mathsf{Y}^\tau = \mathsf{y}^\tau\}$ by[9] $\mathcal{Y}^\mathsf{T}$ and write the decoding rule Ψ as
$$\Psi(\cdot) : \mathcal{Y}^\mathsf{T} \to \mathcal{M}.$$
4) Note that because of the condition $\mathbf{P}[\mathsf{T} < \infty] = 1$, decoding time is not only a Markov time, but also a stopping time.[10]

At time zero, the message M chosen uniformly at random from $\mathcal{M}$ is given to the transmitter; the transmitter uses the codeword associated M, i.e., $\Phi(\mathsf{M}, \cdot)$, to convey the message M until the decoding time T. Then, the receiver chooses the decoded message $\widehat{\mathsf{M}}$ using its observation Y^T and the decoding rule Ψ, i.e., $\widehat{\mathsf{M}} = \Psi(\mathsf{Y}^\mathsf{T})$. The error probability, the rate, and the error exponent of a variable-length block code are given by

$$P_\mathrm{e} = \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M}\right] \tag{7a}$$

$$\mathrm{R} = \frac{\ln |\mathcal{M}|}{\mathbf{E}[\mathsf{T}]} \tag{7b}$$

$$\mathrm{E} = \frac{-\ln P_\mathrm{e}}{\mathbf{E}[\mathsf{T}]}. \tag{7c}$$

Indeed, one can interpret the variable-length block codes on DMCs as trees; for a more detailed discussion of this interpretation, readers may go over[1, Sec. II].

D. Reliable Sequences for Variable-Length Block Codes

In order to suppress the secondary terms while discussing the main results, we use the concept of reliable sequences. In a sequence of codes, we denote the error probability and the message set of the κth code of the sequence by $P_\mathrm{e}^{(\kappa)}$ and $\mathcal{M}^{(\kappa)}$, respectively.

Definition 5 (Reliable Sequence): A sequence of variable-length block codes $\mathbb{Q}$ is reliable if the error probabilities of the codes vanish and the size of the message sets of the codes diverge[11]
$$\lim_{\kappa \to \infty} \left(P_\mathrm{e}^{(\kappa)} + \tfrac{1}{|\mathcal{M}^{(\kappa)}|}\right) = 0$$
where $P_\mathrm{e}^{(\kappa)}$ and $\mathcal{M}^{(\kappa)}$ are the error probability and the message set for the κth code of the reliable sequence, respectively.

Note that in a sequence of codes, each code has an associated probability space. We denote the random variables in these probability spaces together with a superscript corresponding to the code. For example, the decoding time of the κth code in the sequence is denoted by $\mathsf{T}^{(\kappa)}$. The expected value of random variables in the probability space associated with the κth code in the sequence is denoted[12] by $\mathbf{E}^{(\kappa)}[\cdot]$.

[8] $\mathcal{Z}^{\mathsf{T}*}$ is a countable set even when $|\mathcal{Z}|$ is countably infinite.

[9] See (6).

[10] Having a finite decoding time with probability one, i.e., $\mathbf{P}[\mathsf{T} < \infty] = 1$, does not imply having a finite expected value for the decoding time, i.e., $\mathbf{E}[\mathsf{T}] < \infty$. Thus, a variable-length code can, in principle, have an infinite expected decoding time.

[11] Recall that the decoding time of a variable-length block code is finite with probability one. Thus, $\mathbf{P}^{(\kappa)}\left[\mathsf{T}^{(\kappa)} < \infty\right] = 1$ for all κ for a reliable sequence.

[12] Evidently, it is possible to come up with a probability space that includes all of the codes in a reliable sequence and invoke independence between random quantities associated with different codes. We choose the current convention to emphasize independence explicitly in the notation we use.

Definition 6 (Rate of a Reliable Sequence): The rate of a reliable sequence $\mathbb{Q}$ is the limit infimum of the rates of the individual codes

$$\mathrm{R}_{\mathbb{Q}} \stackrel{\triangle}{=} \liminf_{\kappa\to\infty} \frac{\ln |\mathcal{M}^{(\kappa)}|}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}.$$

Definition 7 (Capacity): The capacity of a channel for variable-length block codes is the supremum of the rates of the all reliable sequences

$$C \stackrel{\triangle}{=} \sup_{\mathbb{Q}} \mathrm{R}_{\mathbb{Q}}.$$

The capacity of a DMC for variable-length block codes is identical to the usual channel capacity [3]. Hence

$$C = \max_{\mu\in P(\mathcal{X})} \sum\nolimits_{\mathsf{x},\mathsf{y}} \mu(\mathsf{x}) W_{\mathsf{x}}(\mathsf{y}) \ln \frac{W_{\mathsf{x}}(\mathsf{y})}{\bar{\mu}(\mathsf{y})} \tag{8}$$

where $\bar{\mu}(\mathsf{y}) = \sum_{\mathsf{x}} \mu(\mathsf{x}) W_{\mathsf{x}}(\mathsf{y})$.

Definition 8 (Error Exponent of a Reliable Sequence): The error exponent of a reliable sequence $\mathbb{Q}$ is the limit infimum of the error exponents of the individual codes

$$\mathrm{E}_{\mathbb{Q}} \stackrel{\triangle}{=} \liminf_{\kappa\to\infty} \frac{-\ln {P_{\mathbf{e}}}^{(\kappa)}}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}.$$

Definition 9 (Reliability Function): The reliability function of a channel for variable-length block codes at rate $\mathrm{R} \in [0, C]$ is the supremum of the exponents of all reliable sequences whose rate is R or higher

$$E(\mathrm{R}) \stackrel{\triangle}{=} \sup_{\mathbb{Q}:\mathrm{R}_{\mathbb{Q}}\geq \mathrm{R}} \mathrm{E}_{\mathbb{Q}}.$$

Burnashev [3] analyzed the performance of variable-length block codes with feedback and established inner and outer bounds to their performance. Results of [3] determine the reliability function of variable-length block codes on DMCs for all rates. According to [3], we have the following.

1) If all entries of W are positive, then[13]

$$\mathrm{E}(\mathrm{R}) = \left(1 - \frac{\mathrm{R}}{C}\right) D \qquad \forall \mathrm{R} \in [0, C]$$

where D is maximum Kullback–Leibler divergence between the output distributions of any two input letters

$$D \stackrel{\triangle}{=} \max_{\mathsf{x},\tilde{\mathsf{x}}\in\mathcal{X}} \mathrm{D}\left(W_{\mathsf{x}} \| W_{\tilde{\mathsf{x}}}\right). \tag{9}$$

2) If there are one or more zero entries[14] in W, i.e., if there are two input letters x, $\tilde{\mathsf{x}}$ and an output letter y such that, $W_{\mathsf{x}}(\mathsf{y}) = 0$ and $W_{\tilde{\mathsf{x}}}(\mathsf{y}) > 0$, then for all $\mathrm{R} < C$, for large enough $\mathbf{E}[\mathsf{T}]$, there are rate R variable-length block codes that are error free, i.e., $P_{\mathbf{e}} = 0$.

When $P_{\mathbf{e}} = 0$, all error events can have zero probability at the same time. Consequently, all the UEP problems are answered trivially when there is a zero probability transition. This is why we have assumed that $W_{\mathsf{x}}(\mathsf{y}) > 0$ for all $\mathsf{x} \in \mathcal{X}$ and $\mathsf{y} \in \mathcal{Y}$.

We denote the input letters that get this maximum value of Kullback–Leibler divergence by[15] $\mathbf{a}$ and $\mathbf{r}$

$$D = \mathrm{D}\left(W_{\mathbf{a}} \| W_{\mathbf{r}}\right). \tag{10}$$

III. Problem Statement and Main Results

A. Problem Statement

For each $\mathsf{m} \in \mathcal{M}$, the conditional error probability is defined as[16]

$$P_{\mathbf{e}|\mathsf{m}} \stackrel{\triangle}{=} \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M} \,\middle|\, \mathsf{M} = \mathsf{m}\right]. \tag{11}$$

In the conventional setting, we are interested in either the average or the maximum of the conditional error probability of the messages. The behavior of the minimum conditional error probability is scarcely investigated. Single-message UEP problem attempts to answer that question by determining the tradeoff between exponential decay rates of $P_{\mathbf{e}}$ and $\min_{\mathsf{m}\in\mathcal{M}} P_{\mathbf{e}|\mathsf{m}}$. The operational definition of the problem in terms of reliable sequences is as follows.

Definition 10 (Single-Message Message-Wise UEP Problem): For any reliable sequence $\mathbb{Q}$, the missed detection exponent of the reliable sequence $\mathbb{Q}$ is defined as

$$\mathrm{E}_{\mathrm{md},\mathbb{Q}} = \liminf_{\kappa\to\infty} \frac{-\ln \min_{\mathsf{m}\in\mathcal{M}^{(\kappa)}} P_{\mathbf{e}|\mathsf{m}}^{(\kappa)}}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]} \tag{12}$$

where $P_{\mathbf{e}|\mathsf{m}}^{(\kappa)}$ is the conditional error probability of the message m for the κth code of the reliable sequence $\mathbb{Q}$.

For any rate $\mathrm{R} \in [0, C]$ and error exponent[17] $\mathrm{E} \in [0, (1 - \frac{\mathrm{R}}{C})D]$, the missed detection exponent $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ is defined as

$$\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E}) \stackrel{\triangle}{=} \sup_{\mathbb{Q}:\substack{\mathrm{R}_{\mathbb{Q}}\geq \mathrm{R} \\ \mathrm{E}_{\mathbb{Q}}\geq \mathrm{E}}} \mathrm{E}_{\mathrm{md},\mathbb{Q}}. \tag{13}$$

In variable-length block codes with feedback, the single-message message-wise UEP problem not only answers a curious

[13]Problem is formulated somewhat differently in [3]; as a result, the work in [3] did not deal with the case $\mathbf{E}[\mathsf{T}] = \infty$. The bounds in [3] does not guarantee that the error probability of a variable-length code with infinite expected decoding time is greater than zero; however, this is the case if all the transition probabilities are positive. To see that, consider a channel with positive minimum transition probability λ, i.e., $\lambda = \min_{\mathsf{x},\mathsf{y}} W_{\mathsf{x}}(\mathsf{y}) > 0$. In such a channel, any variable-length code satisfies $P_{\mathbf{e}} \geq \frac{|\mathcal{M}|-1}{|\mathcal{M}|}\mathbf{E}\left[\left(\frac{\lambda}{1-\lambda}\right)^{\mathsf{T}}\right]$; then, $P_{\mathbf{e}} > 0$ as $\lambda > 0$ and $\mathbf{P}[\mathsf{T} < \infty] = 1$. Consequently, both the rate and the error exponent are zero for variable-length block codes with infinite expected decoding time. A more detailed discussion of this fact can be found in Section H1 in the Appendix.

[14]Note that, in this situation, $D = \infty$.

[15]This particular naming of letters is reminiscent of the use of these letters in Yamamoto–Itoh scheme [12]. Although they are named differently in [12], $\mathbf{a}$ is used for accepting and $\mathbf{r}$ is used for rejecting the tentative decision in Yamamoto–Itoh scheme.

[16]Later, in the paper, we consider block codes with erasures. The conditional error probabilities, $P_{\mathbf{e}|\mathsf{m}}$ for $\mathsf{m} \in \mathcal{M}$, are defined slightly differently for them, see (24).

[17]Burnashev's expression for error exponent of variable-length block codes is used explicitly in the definition because we know, as a result of [3], that the error exponents of all reliable sequences are upper bounded by Burnashev's exponent. An alternative definition oblivious to Burnashev's result can simply define $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ for all rate-exponent vectors that are achievable. That definition is equivalent to Definition 10, because of [3].

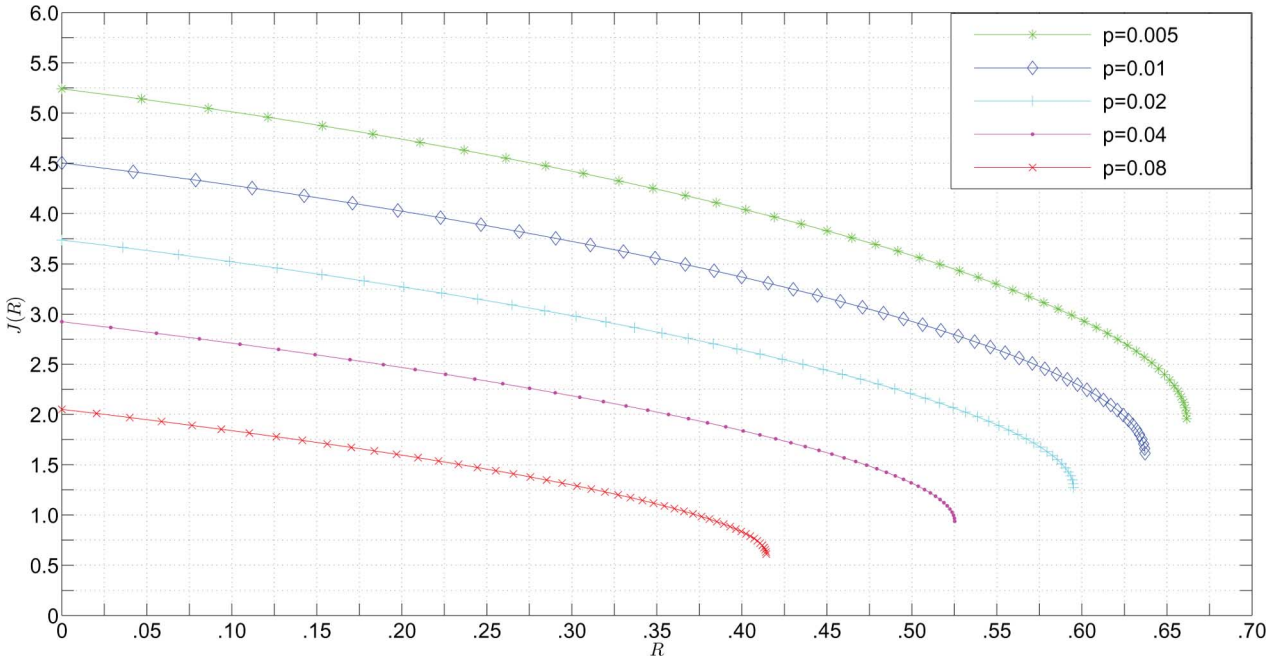

Fig. 1. $J(\mathrm{R})$ function is drawn for BSCs with crossover probabilities $p \in \{0.005, 0.01, 0.02, 0.04, 0.08\}$.

question about the decay rate of the minimum conditional error probability of a code, but also plays a key role in the bit-wise UEP problem, which is our main focus in this paper.

Though they are central in the message-wise UEP problems, the conditional error probabilities of the messages are not relevant in the bit-wise UEP problems. In the bit-wise UEP problems, we analyze the error probabilities of groups of submessages. In order to do that, consider a code with a message set $\mathcal{M}$ of the form

$$\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_\ell.$$

Then, the transmitted message M and decoded message $\widehat{\mathsf{M}}$ of the code are of the form

$$\mathsf{M} = (\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_\ell)$$
$$\widehat{\mathsf{M}} = (\widehat{\mathsf{M}}_1, \widehat{\mathsf{M}}_2, \ldots, \widehat{\mathsf{M}}_\ell)$$

where $\mathsf{M}_j, \widehat{\mathsf{M}}_j \in \mathcal{M}_j$ for all $j = 1, 2, \ldots, \ell$. Furthermore, M_j and $\widehat{\mathsf{M}}_j$ are called jth transmitted submessage and jth decoded submessage, respectively.

The error probabilities we are interested in correspond to erroneous transmission of certain parts of the message. In order to define them succinctly, let us define $\mathcal{M}^j$, M^j, and $\widehat{\mathsf{M}}^j$ for all j between one and ℓ as follows:

$$\mathcal{M}^j \stackrel{\triangle}{=} \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_j$$
$$\mathsf{M}^j \stackrel{\triangle}{=} (\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_j)$$
$$\widehat{\mathsf{M}}^j \stackrel{\triangle}{=} (\widehat{\mathsf{M}}_1, \widehat{\mathsf{M}}_2, \ldots, \widehat{\mathsf{M}}_j).$$

Then, $P_{\mathrm{e}}(j)$ is defined[18] as the probability of the event that $\widehat{\mathsf{M}}^j \neq \mathsf{M}^j$

$$P_{\mathrm{e}}(j) \stackrel{\triangle}{=} \mathbf{P}\left[\widehat{\mathsf{M}}^j \neq \mathsf{M}^j\right] \qquad \text{for } j = 1, 2, \ldots, \ell. \tag{14}$$

[18]Similar to the conditional error probabilities, $P_{\mathrm{e}|\mathsf{m}}$'s for $\mathsf{m} \in \mathcal{M}$, error probabilities of submessages, $P_{\mathrm{e}}(j)$'s for $j = 1, 2, \ldots, \ell$, are defined slightly differently for codes with erasures, see (30).

Note that if $\widehat{\mathsf{M}}^j \neq \mathsf{M}^j$, then $\widehat{\mathsf{M}}^i \neq \mathsf{M}^i$ for all i greater than j. Thus

$$P_{\mathrm{e}}(1) \leq P_{\mathrm{e}}(2) \leq P_{\mathrm{e}}(3) \leq \cdots \leq P_{\mathrm{e}}(\ell). \tag{15}$$

Definition 11 (Bit-Wise UEP Problem for Fixed ℓ): For any positive integer ℓ, let $\mathbb{Q}$ be a reliable sequence whose message sets $\mathcal{M}^{(\kappa)}$ are of the form $\mathcal{M}^{(\kappa)} = \mathcal{M}_1^{(\kappa)} \times \mathcal{M}_2^{(\kappa)} \times \cdots \times \mathcal{M}_\ell^{(\kappa)}$. Then, the entries of the rate vector $\vec{\mathrm{R}}_{\mathbb{Q}}$ and the error exponent vector $\vec{\mathrm{E}}_{\mathbb{Q}}$ are defined as

$$\mathrm{R}_{\mathbb{Q},j} \stackrel{\triangle}{=} \liminf_{\kappa\to\infty} \frac{\ln |\mathcal{M}_j^{(\kappa)}|}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]} \quad \forall j \in \{1, 2, \ldots, \ell\}$$
$$\mathrm{E}_{\mathbb{Q},j} \stackrel{\triangle}{=} \liminf_{\kappa\to\infty} \frac{-\ln P_{\mathrm{e}}(j)^{(\kappa)}}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]} \quad \forall j \in \{1, 2, \ldots, \ell\}.$$

A rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ is achievable if and only if there exists a reliable sequence $\mathbb{Q}$ such that $(\vec{\mathrm{R}}, \vec{\mathrm{E}}) = (\vec{\mathrm{R}}_{\mathbb{Q}}, \vec{\mathrm{E}}_{\mathbb{Q}})$.

This definition of the *bit-wise* UEP problem is slightly different than the one described in the introduction, because $P_{\mathrm{e}}(j)$ is defined as $\mathbf{P}\left[\widehat{\mathsf{M}}^j \neq \mathsf{M}^j\right]$ rather than $\mathbf{P}\left[\widehat{\mathsf{M}}_j \neq \mathsf{M}_j\right]$. Note that if $\widehat{\mathsf{M}}_j \neq \mathsf{M}_j$, then $\widehat{\mathsf{M}}^j \neq \mathsf{M}^j$; consequently, $\mathbf{P}\left[\widehat{\mathsf{M}}^j \neq \mathsf{M}^j\right] \geq \mathbf{P}\left[\widehat{\mathsf{M}}_j \neq \mathsf{M}_j\right]$ for all j's. In addition, if we assume without loss of generality that $\mathbf{P}\left[\widehat{\mathsf{M}}_j \neq \mathsf{M}_j\right] \geq \mathbf{P}\left[\widehat{\mathsf{M}}_i \neq \mathsf{M}_i\right]$ for all $j \geq i$, the union bound implies that $\mathbf{P}\left[\widehat{\mathsf{M}}^j \neq \mathsf{M}^j\right] \leq j\mathbf{P}\left[\widehat{\mathsf{M}}_j \neq \mathsf{M}_j\right]$. Thus, for the case when ℓ is fixed, both formulations of the problem result in exactly the same achievable region of rate-exponent vectors.

The achievable region of rate-exponent vectors could have been defined as the closure of the points of the form $(\vec{\mathrm{R}}_{\mathbb{Q}}, \vec{\mathrm{E}}_{\mathbb{Q}})$ for some reliable sequence $\mathbb{Q}$. Using the definition of $(\vec{\mathrm{R}}_{\mathbb{Q}}, \vec{\mathrm{E}}_{\mathbb{Q}})$'s, one can easily show that, in this case too both

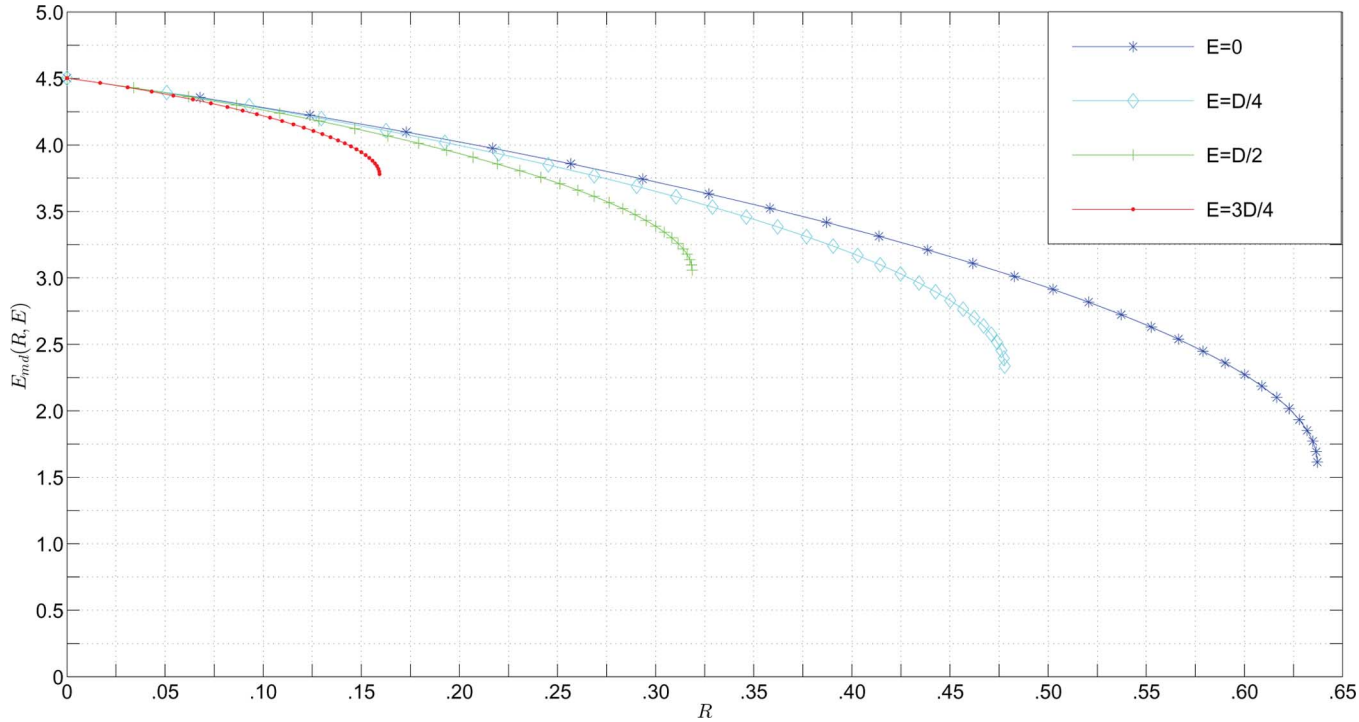

Fig. 2. $\mathrm{E_{md}(R,E)}$ is drawn at various values of the error exponent E as a function of rate R for a BSC with crossover probability $p = 0.01$. Note that when $p = 0.01$, $C = 0.6371$ Nats per channel use and $D = 4.503$. As we increase the exponent of the average error probability, i.e., E, the value of $\mathrm{E_{md}(R,E)}$ decreases, as one would expect.

definitions result in exactly the same achievable region of rate-exponent vectors.

B. Main Results

For variable-length block codes with feedback, the results of both the single-message *message-wise* UEP problem and the *bit-wise* UEP problem are given in terms of the $J(\mathrm{R})$ function defined in the following. The $J(\mathrm{R})$ function is first introduced by[19] Kudryashov [7, eq. (2.6)], while describing the performance of nonblock variable-length codes with feedback and delay constraints. Later, the $J(\mathrm{R})$ function is used in [2] for describing the performance of block codes in single-message *message-wise* UEP problem. It is shown in [2, Appendix D] that for both fixed-length block codes without feedback and variable-length block codes with feedback on DMCs satisfy

$$\mathrm{E_{md}(R,0)} = J(\mathrm{R})\,. \tag{16}$$

Recently Nazer *et al*. obtained the closed-form expression for $\mathrm{E_{md}(R,0)}$ for fixed-length block codes on the additive white Gaussian noise channel, under certain average and peak power constraints [9, Th. 1]. Curiously, equality given in (16) holds in that case too.[20]

Definition 12: For any $\mathrm{R} \in [-\infty, C]$, $J(\mathrm{R})$ is defined as

$$J(\mathrm{R}) \stackrel{\triangle}{=} \max_{\substack{0\le\alpha\le 1\\ \mathsf{x}_1,\mathsf{x}_2\in\mathcal{X}\\ \mu_1,\mu_2\in P(\mathcal{X})}} \alpha \mathrm{D}\left(\bar{\mu}_1 \| W_{\mathsf{x}_1}\right) + (1-\alpha)\mathrm{D}\left(\bar{\mu}_2 \| W_{\mathsf{x}_2}\right) \tag{17}$$
$$\alpha,\mathsf{x}_1,\mathsf{x}_2,\mu_1,\mu_2:\ \alpha \mathrm{I}(\mu_1,W)+(1-\alpha)\mathrm{I}(\mu_2,W)\ge \mathrm{R}$$

where $\bar{\mu}_i(\mathsf{y}) = \sum_{\mathsf{x}} W_{\mathsf{x}}(\mathsf{y})\mu_i(\mathsf{x})$ for $i = 1, 2$.

We have plotted the $J(\mathrm{R})$ function for binary symmetric channels[21] (BSCs) with various crossover probabilities in Fig. 1. Note that as the channel becomes noisier, i.e., as the crossover probability becomes closer to 1/2, the value of $J(\mathrm{R})$ function decreases at all values of rate where it is positive. Furthermore, the highest value of rate where it is positive, i.e., the channel capacity, decreases.

Lemma 1: The function $J(\mathrm{R})$ defined in (17) is a concave, decreasing function such that $J(\mathrm{R}) = D$ for $\mathrm{R} \le 0$.

Proof of Lemma 1 is given in Section A in the Appendix.

Now, let us consider the singe-message *message-wise* UEP problem given in Definition 10.

Theorem 1: For any rate $0 \le \mathrm{R} \le C$ and error exponent $\mathrm{E} \le (1 - \frac{\mathrm{R}}{C})D$, the missed detection exponent $\mathrm{E_{md}(R,E)}$ defined in (13) is equal to[22]

$$\mathrm{E_{md}(R,E)} = \mathrm{E} + \left(1 - \tfrac{\mathrm{E}}{D}\right) J\left(\tfrac{\mathrm{R}}{1-\frac{\mathrm{E}}{D}}\right) \tag{18}$$

where C, D, and $J(\cdot)$ are given in (8), (9), and (17), respectively. Furthermore $\mathrm{E_{md}(R,E)}$ is jointly concave in (R,E) pairs.

We have plotted $\mathrm{E_{md}(R,E)}$ as a function of rate, for various values of E in Fig. 2. When rate is zero, the exponent of the average error probability can be made as high as D. Thus, all

[19] In [7, eq. (2.6)], there is no optimization over the parameter α. Thus, strictly speaking, what is introduced in [7, eq. (2.6)] is $j(\mathrm{R})$ given in (64) rather than $J(\mathrm{R})$ given in (17).

[20] Unlike DMC for these channels, it is possible to obtain a closed-form expression in terms of the rate and the power constraints.

[21] Recall that in a BSC with crossover probability p, $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1\}$, and $W_{\mathsf{x}}(\mathsf{y}) = (1-p)\mathbb{1}_{\{\mathsf{x}=\mathsf{y}\}} + p\mathbb{1}_{\{\mathsf{x}\neq\mathsf{y}\}}$.

[22] For the case when $\mathrm{R} = 0$ and $\mathrm{E} = D$, the $\left(1 - \frac{\mathrm{E}}{D}\right) J\left(\frac{\mathrm{R}}{1-\frac{\mathrm{E}}{D}}\right)$ term should be interpreted as 0, i.e., $\left(1 - \frac{\mathrm{E}}{D}\right) J\left(\frac{\mathrm{R}}{1-\frac{\mathrm{E}}{D}}\right)\Big|_{\substack{\mathrm{R}=0\\ \mathrm{E}=D}} = 0$.

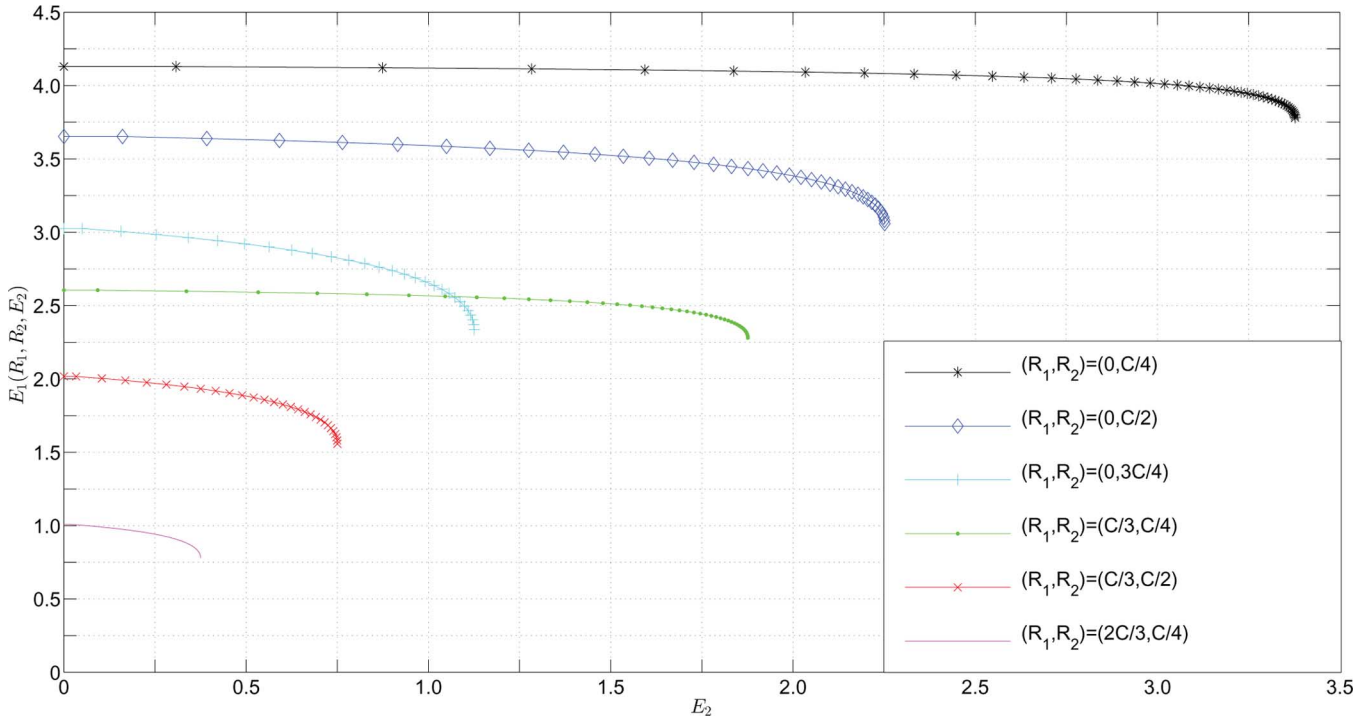

Fig. 3. $E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$ is drawn for various values of rate pairs $(\mathrm{R}_1, \mathrm{R}_2)$ as a function of error exponent E_2 for a BSC with crossover probability $p = 0.01$. Recall that when $p = 0.01$, $C = 0.6371$ Nats per channel use and $D = 4.503$.

the curves meet at $(0, D)$ point. But for all positive rates, the exponent of the average error probability makes a difference; as E increases, $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ decreases. Furthermore, for any given rate R, the exponent of the average error probability can only be as high as $(1 - \frac{\mathrm{R}}{C})D$. This is why the curves corresponding to higher values of E have smaller support on rate axis.

Proof of Theorem 1 is presented in Section I in the Appendix.

Similar to the single-message *message-wise* UEP problem, the solution of the *bit-wise* UEP problem is given in terms of the $J(\mathrm{R})$ function.

Theorem 2: A rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ is achievable if and only if there exists a $\vec{\eta}$ such that[23]

$$\mathrm{E}_i \le (1 - \sum_{j=1}^{\ell} \eta_j) D + \sum_{j=i+1}^{\ell} \eta_j J\left(\frac{\mathrm{R}_j}{\eta_j}\right) \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{19a}$$

$$\mathrm{R}_i \le C\eta_i \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{19b}$$

$$\eta_i \ge 0 \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{19c}$$

$$\sum_{j=1}^{\ell} \eta_j \le 1 \tag{19d}$$

where C, D, and $J(\cdot)$ are given in (8), (9), and (17), respectively. Furthermore, the set of all achievable rate-exponent vectors is convex.

Proof of Theorem 2 is presented in Section J in the Appendix.

For the special case when there are only two submessages, the condition given in Theorem 2 for the achievability of a rate vector error exponent vector pair can be turned into an analytical expression for the optimal E_1 in terms of R_1, R_2, and E_2. In order to see why, note that revealing the region of achievable $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_1, \mathrm{E}_2)$ vectors is equivalent to revealing the region of achievable $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$'s and the value of the maximum achievable E_1 for all the $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$'s in the achievable region.

Corollary 1: For any rate pair $(\mathrm{R}_1, \mathrm{R}_2)$ such that $\mathrm{R}_1 + \mathrm{R}_2 \le C$ and error exponent E_2 such that $\mathrm{E}_2 \le (1 - \frac{\mathrm{R}_1 + \mathrm{R}_2}{C})D$, the optimal value of E_1 is given by[24]

$$E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2) = \mathrm{E}_2 + \left(1 - \frac{\mathrm{R}_1}{C} - \frac{\mathrm{E}_2}{D}\right) J\left(\frac{\mathrm{R}_2}{1 - \frac{\mathrm{R}_1}{C} - \frac{\mathrm{E}_2}{D}}\right) \tag{20}$$

where C, D, and $J(\cdot)$ are given in (8), (9), and (17), respectively. Furthermore $E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$ is concave in $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$.

Note that for the $E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$ given in (20), $E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2) \ge \mathrm{E}_2$ for all $(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$ triples such that $\mathrm{R}_1 + \mathrm{R}_2 \le C$ and $\mathrm{E}_2 \le (1 - \frac{\mathrm{R}_1 + \mathrm{R}_2}{C})D$. Furthermore, inequality is strict as long as $\mathrm{R}_2 > 0$. We have drawn $E_1(\mathrm{R}_1, \mathrm{R}_2, \mathrm{E}_2)$ for various $(\mathrm{R}_1, \mathrm{R}_2)$ pairs as a function of E_2 in Fig. 3.

IV. Achievability

In both the single-message *message-wise* UEP problem and the *bit-wise* UEP problem, the codes that achieve the optimal performance employ a number of different ideas at the same time. In order to avoid introducing all of those ideas at once,

[23]For the case when $\mathrm{R}_j = 0$ and $\eta_j = 0$, the $\eta_j J\left(\frac{\mathrm{R}_j}{\eta_j}\right)$ term should be interpreted as 0, i.e., $\eta_j J\left(\frac{\mathrm{R}_j}{\eta_j}\right)\Big|_{\substack{\mathrm{R}_j = 0 \\ \eta_j = 0}} = 0$.

[24]For the case when $\mathrm{R}_2 = 0$ and $\mathrm{E}_2 = (1 - \frac{\mathrm{R}_1}{C})D$, the second term on the right-hand side of (20) should be interpreted as zero, i.e., $\left(1 - \frac{\mathrm{R}_1}{C} - \frac{\mathrm{E}_2}{D}\right) J\left(\frac{\mathrm{R}_2}{1 - \frac{\mathrm{R}_1}{C} - \frac{\mathrm{E}_2}{D}}\right)\Big|_{\mathrm{R}_2 = 0,\ \mathrm{E}_2 = (1 - \frac{\mathrm{R}_1}{C})D} = 0$

we first describe two families of codes and analyze the probabilities of various error events in those two families of codes. Later, we use those two families of codes as the building blocks for the codes that achieve the optimal performance in the UEP problems we are interested in. Before going into a more detailed description and analysis of those codes, let us first give a birds eye plan for this section.

1) *A Single-Message Message-wise UEP Scheme without Feedback:* First, in Section IV-A, we consider a family of fixed-length codes without feedback. We prove that these codes can achieve any rate R less than channel capacity, with vanishing[25] error probability P_{e} while having a minimum conditional error probability, $\min_{\mathsf{m}} P_{\mathbf{e}|\mathsf{m}}$, as low as $e^{-nJ(\mathrm{R})}$. The main drawback of this family of codes is that the decay rate of the average error probability P_{e} has to be subexponential in this family of codes.
2) *Control Phase and Error-Erasure Decoding:* In Section IV-B, in order to obtain nonzero exponential decay for the average error probability, we use a method introduced by Yamamoto and Itoh in [12]. We append the fixed-length codes described in Section IV-A with a control phase and use an error-erasure decoder. This new family of codes with control phase and error-erasure decoding is shown, in Section IV-B, to achieve any rate R less than the channel capacity C with exponentially decaying average error probability $P_{\mathbf{e}}$, exponentially decaying minimum conditional error probability $\min_{\mathsf{m}} P_{\mathbf{e}|\mathsf{m}}$, and vanishing erasure probability $P_{\mathbf{x}}$.
3) *Single-Message Message-wise UEP for Variable-Length Codes:* In Section IV-C, we obtain variable-length codes for single-message *message-wise* UEP problem using the codes described in Section IV-B. In order to do that, we use the fixed-length codes with feedback and erasures described in Section IV-B, repetitively until a nonerasure decoding happens. This idea too was employed by Yamamoto and Itoh in [12].
4) *Bit-wise UEP for Variable-Length Codes:* In Section IV-D, we first use the codes described in Section IV-A and the control phase discussed in Section IV-B to obtain a family of fixed-length codes with feedback and erasures which has *bit-wise* UEP, i.e., which has different bounds on error probabilities for different submessages. While using the codes described in Section IV-A, we employ an implicit acceptance explicit rejection scheme first introduced in [7] by Kudryashov. Once we obtain a fixed-length code with erasures and *bit-wise* UEP, we use a repeat at erasures scheme like the one described in Section IV-C to obtain a variable-length code with *bit-wise* UEP.

[25]Vanishing with increasing block length.

The achievability results we derive in this section are revealed to be the optimal ones, in terms of the decay rates of error probabilities with expected decoding time $\mathbf{E}[\mathsf{T}]$, as a result of the outer bounds we derive in Section V.

A. Single-Message Message-Wise UEP Scheme Without Feedback

In this section, we describe a family of fixed-length block codes without feedback that achieves any rate R less than capacity with small error probability while having an exponentially small $\min_{\mathsf{m}} P_{\mathbf{e}|\mathsf{m}}$, for sufficiently large block length n. We describe these codes in terms of a time-sharing constant $\alpha \in [0,1]$, two input letters $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, and two probability distributions on the input alphabet, $\mu_1, \mu_2 \in P(\mathcal{X})$.

In order to point out that certain sequence of input letters is a codeword or part of a codeword for message m, we put (m) after it. Hence, we denote the codeword for m by $\mathsf{x}^n(\mathsf{m})$ in a given code and by $\mathsf{X}^n(\mathsf{m})$ in a code ensemble, as a random quantity.

Let us start with describing the encoding scheme. The codeword of the first message, i.e., $\mathsf{x}^n(1)$, is x_1 in first $n_\alpha = \lfloor \alpha n \rfloor$ time instances and x_2 in the rest, i.e., $\mathsf{x}_\tau(1) = \mathsf{x}_1$ for $\tau = 1, \ldots, n_\alpha$ and $\mathsf{x}_\tau(1) = \mathsf{x}_2$ for $\tau = (n_\alpha + 1), \ldots, n$. The codewords of the other messages are described via a random coding argument. In the ensemble of codes, we are considering all entries of all codewords other than the first codeword, i.e., $\mathsf{X}_\tau(\mathsf{m})$ $\forall \tau \in [1,n], \forall \mathsf{m} \neq 1$, are generated independently of other codewords and other entries of the same codeword. In the first n_α time instances, $\mathsf{X}_\tau(\mathsf{m})$ is generated using μ_1, in the rest using μ_2, i.e., $\mathbf{P}[\mathsf{X}_\tau(\mathsf{m}) = \mathsf{x}] = \mu_1(\mathsf{x})$ for $\tau = 1, \ldots, n_\alpha$ and $\mathbf{P}[\mathsf{X}_\tau(\mathsf{m}) = \mathsf{x}] = \mu_2(\mathsf{x})$ for $\tau = (n_\alpha + 1), \ldots, n$.

Let us begin the description of the decoding scheme, by specifying the decoding region of the first message $\mathcal{G}[1]$: it is the set of all output sequences y^n whose empirical distribution is not typical with $(\alpha, \bar{\mu}_1, \bar{\mu}_2)$. More precisely, the decoding region of the first message $\mathcal{G}[1]$ is given by (21) shown at the bottom of the page where Δ is the total variation distance defined in (1), $\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}$ and $\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}$ are the empirical distributions of $\mathsf{y}_1^{n_\alpha}$ and $\mathsf{y}_{n_\alpha+1}^{n}$ defined in (4), and $\bar{\mu}_1$ and $\bar{\mu}_2$ are probability distributions on $\mathcal{Y}$, i.e., $\bar{\mu}_1, \bar{\mu}_2 \in P(\mathcal{Y})$, such that $\bar{\mu}_i(\mathsf{y}) = \sum_{\mathsf{x}} \mu_i(\mathsf{x}) W_{\mathsf{x}}(\mathsf{y})$.

For other messages, $\mathsf{m} \neq 1$, decoding regions $\mathcal{G}[\mathsf{m}]$ are the set of all output sequences for which $\mathrm{Q}_{\{\mathsf{x}^n(\mathsf{m}),\mathsf{y}^n\}}$ is typical with $(\alpha, \mu_1 W, \mu_2 W)$ and $\mathrm{Q}_{\{\mathsf{x}^n(\widetilde{\mathsf{m}}),\mathsf{y}^n\}}$ is not typical with $(\alpha, \mu_1 W, \mu_2 W)$ for any $\widetilde{\mathsf{m}} \neq \mathsf{m}$. To be precise, the decoding region of the messages other than the first message are

$$\mathcal{G}[\mathsf{m}] = \mathcal{B}[\mathsf{x}^n(\mathsf{m})] \bigcap \left(\cap_{\widetilde{\mathsf{m}} \neq \mathsf{m}} \overline{\mathcal{B}[\mathsf{x}^n(\widetilde{\mathsf{m}})]} \right) \qquad \forall \mathsf{m} \in \{2, 3, \ldots, |\mathcal{M}|\} \quad (22)$$

where for all $\mathsf{x}^n \in \mathcal{X}^n$, $\mathcal{B}[\mathsf{x}^n]$ is the set of all y^n's for which $(\mathsf{x}^n, \mathsf{y}^n)$ is typical with $(\alpha, \mu_1 W, \mu_2 W)$ shown at the bottom

$$\mathcal{G}[1] = \left\{ \mathsf{y}^n : n_\alpha \Delta\left(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}, \bar{\mu}_1 \right) + (n - n_\alpha) \Delta\left(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}, \bar{\mu}_2 \right) \geq |\mathcal{X}||\mathcal{Y}| \sqrt{n \ln(1+n)} \right\} \quad (21)$$

of the page where Δ is the total variation distance defined in (1), $\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha},\mathsf{y}_1^{n_\alpha}\}}$ and $\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n},\mathsf{y}_{n_\alpha+1}^{n}\}}$ are the empirical distributions of $(\mathsf{x}_1^{n_\alpha},\mathsf{y}_1^{n_\alpha})$ and $(\mathsf{x}_{n_\alpha+1}^{n},\mathsf{y}_{n_\alpha+1}^{n})$ defined in (4), and $\mu_1 W$ and $\mu_2 W$ are probability distributions on $\mathcal{X}\times\mathcal{Y}$, i.e., $\mu_1 W \in P(\mathcal{X}\times\mathcal{Y})$ and $\mu_2 W \in P(\mathcal{X}\times\mathcal{Y})$.

In Section B in the Appendix, we have analyzed the conditional error probabilities $P_{\mathsf{e}|\mathsf{m}}$ for the previously described code and proved Lemma 2 given in the following.

Lemma 2: For any block length n, time-sharing constant $\alpha\in[0,1]$, input letters $\mathsf{x}_1,\mathsf{x}_2\in\mathcal{X}$, and input distributions $\mu_1,\mu_2\in P(\mathcal{X})$, there exists a length n block code such that

$$\begin{aligned}
&|\mathcal{M}| \geq e^{n(\alpha \mathrm{I}(\mu_1,W)+(1-\alpha)\mathrm{I}(\mu_2,W)-\varepsilon_n)}\\
&P_{\mathsf{e}|1} \leq e^{-n(\alpha \mathrm{D}(\bar{\mu}_1\|W_{\mathsf{x}_1})+(1-\alpha)\mathrm{D}(\bar{\mu}_2\|W_{\mathsf{x}_2})-\varepsilon_n)}\\
&P_{\mathsf{e}|\mathsf{m}} \leq \varepsilon_n \qquad\qquad \mathsf{m}=2,3,\ldots,|\mathcal{M}|
\end{aligned}$$

where $\bar{\mu}_1(\mathsf{y})=\sum_{\mathsf{x}} W_{\mathsf{x}}(\mathsf{y})\mu_1(\mathsf{x})$, $\bar{\mu}_2(\mathsf{y})=\sum_{\mathsf{x}} W_{\mathsf{x}}(\mathsf{y})\mu_2(\mathsf{x})$ and $\varepsilon_n=\frac{10|\mathcal{X}||\mathcal{Y}|(\ln\frac{e}{\lambda})\sqrt{\ln(1+n)}}{\sqrt{n}}$.

Given the channel W, if we discard the error terms ε_n, for a given value of rate, $0\leq \mathrm{R}\leq C$, we can optimize exponent of $P_{\mathsf{e}|1}$ over the time-sharing constant α, the input letters x_1, x_2, and input distributions μ_1, μ_2. Evidently, the optimization problem we get is the one given for the definition of $J(\mathrm{R})$, in (17). Thus, Lemma 2 implies that for any $\mathrm{R}\in[0,C]$ and block length n, there exists a length n code such that $|\mathcal{M}|\geq e^{n(\mathrm{R}-\varepsilon_n)}$, $P_{\mathsf{e}|\mathsf{m}}\leq\varepsilon_n$ for $\mathsf{m}=2,3,\ldots,|\mathcal{M}|$ and $P_{\mathsf{e}|1}\leq e^{-n(J(\mathrm{R})-\varepsilon_n)}$.

One curious question is whether or not the exponent of $P_{\mathsf{e}|1}$ can be increased by including more than two phases. Carathéodory's theorem answers that question negatively, i.e., to obtain the largest value of $J(\mathrm{R})$ one does not need to do time sharing between more than two input-letter–input-distribution pairs.

B. Control Phase and Error-Erasure Decoding

The family of codes described in Lemma 2 has a large exponent for the conditional error probability of the first message, i.e., $P_{\mathsf{e}|1}$. But the conditional error probabilities of other messages, $P_{\mathsf{e}|\mathsf{m}}$ for $\mathsf{m}\neq 1$, decay subexponentially. In order to facilitate an exponential decay of $P_{\mathsf{e}|\mathsf{m}}$ for $\mathsf{m}\neq 1$ with block length, we append the codes described in Lemma 2 with a control phase and allow erasures. The idea of using a control phase and an error-erasure decoding, in establishing achievability results for variable-length code, was first employed by Yamamoto and Itoh in [12].

In order to explain what we mean by the control phase, let us describe our encoding scheme and decoding rule briefly. First, a code from the family of codes described in Section IV-A is used to transmit M and the receiver makes a tentative decision ${}_{\mathrm{t}}\widehat{\mathsf{M}}$ using the decoder of the very same code. The transmitter knows ${}_{\mathrm{t}}\widehat{\mathsf{M}}$ because of the feedback link. In the remaining time instances, i.e., in the control phase, the transmitter sends the input letter $\mathbf{a}$ if ${}_{\mathrm{t}}\widehat{\mathsf{M}}=\mathsf{M}$, the input letter $\mathbf{r}$ if ${}_{\mathrm{t}}\widehat{\mathsf{M}}\neq\mathsf{M}$. The input letters $\mathbf{a}$ and $\mathbf{r}$ are described in (10). At the end of the control phase, the receiver checks whether or not the output sequence in the control phase is typical with $W_{\mathbf{a}}$, if it is then $\widehat{\mathsf{M}}={}_{\mathrm{t}}\widehat{\mathsf{M}}$ otherwise an erasure is declared.

Lemma 3 given in the following states the results of the performance analysis of the previously described code. In order to understand what is stated in Lemma 3 accurately, let us make a brief digression and elaborate on the codes with erasure. We have assumed in our models until now that $\widehat{\mathsf{M}}\in\mathcal{M}$. However, there are many interesting problems in which this might not hold. In codes with erasures, for example, we replace $\widehat{\mathsf{M}}\in\mathcal{M}$ with $\widehat{\mathsf{M}}\in\widehat{\mathcal{M}}$ where $\widehat{\mathcal{M}}=\mathcal{M}\cup\{\mathbf{x}\}$ and $\mathbf{x}$ is the erasure symbol. Furthermore, in codes with erasures for each $\mathsf{m}\in\mathcal{M}$, the conditional error probability $P_{\mathsf{e}|\mathsf{m}}$ and conditional erasure probability $P_{\mathbf{x}|\mathsf{m}}$ are defined as follows:

$$P_{\mathsf{e}|\mathsf{m}}=\mathbf{P}\Big[\widehat{\mathsf{M}}\notin\{\mathsf{m},\mathbf{x}\}\Big|\,\mathsf{M}=\mathsf{m}\Big]\qquad\forall\mathsf{m}\in\mathcal{M}\tag{24a}$$

$$P_{\mathbf{x}|\mathsf{m}}=\mathbf{P}\Big[\widehat{\mathsf{M}}=\mathbf{x}\Big|\,\mathsf{M}=\mathsf{m}\Big]\qquad\forall\mathsf{m}\in\mathcal{M}.\tag{24b}$$

Note that definitions of $P_{\mathsf{e}|\mathsf{m}}$ and $P_{\mathbf{x}|\mathsf{m}}$ given previously can be seen as the generalizations of the corresponding definitions in block codes without erasures. In erasure free codes, aforementioned definitions are equivalent to corresponding definitions there.

Lemma 3: For any block length n, rate $0\leq\mathrm{R}\leq C$, and error exponent $0\leq\mathrm{E}\leq(1-\frac{\mathrm{R}}{C})D$, there exists a length n block code with erasures such that

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

where $\varepsilon_n=\frac{10|\mathcal{X}||\mathcal{Y}|(\ln\frac{e}{\lambda})\sqrt{\ln(1+n)}}{\sqrt{n}}$.

Proof of Lemma 3 is given in Section C in the Appendix.

Note that in Lemma 3, unlike $P_{\mathsf{e}|\mathsf{m}}$'s which decrease exponentially with n, $P_{\mathbf{x}|\mathsf{m}}$'s decays as $\frac{\ln n}{\sqrt{n}}$. It is possible to tweak the proof so as to have a nonzero exponent for $P_{\mathbf{x}|\mathsf{m}}$'s, see [8]. But this can only be done at the expense of $P_{\mathsf{e}|\mathsf{m}}$'s. Our aim, however, is to achieve the optimal performance in variable-length block codes. As we will see in the following section, for that what matters is exponents of error probabilities and having vanishing erasure probabilities. The rate at which erasure probability decays does not effect the performance of variable-length block codes in terms of error exponents.

$$\mathcal{B}[\mathsf{x}^n]=\{\mathsf{y}^n: n_\alpha\Delta\Big(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha},\mathsf{y}_1^{n_\alpha}\}},\mu_1W\Big)+(n-n_\alpha)\Delta\Big(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n},\mathsf{y}_{n_\alpha+1}^{n}\}},\mu_2W\Big)\leq|\mathcal{X}||\mathcal{Y}|\sqrt{n\ln(1+n)}\}\tag{23}$$

C. Single-Message Message-Wise UEP Achievability

In this section, we construct variable-length block codes for the single-message *message-wise* UEP problem using Lemma 3. In first n time units, the variable-length encoding scheme uses a fixed-length block code with erasures which has the performance described in Lemma 3. If the decoded message of the fixed-length code is in the message set, i.e., if $\widehat{\mathsf{M}} \in \mathcal{M}$, then decoded message of the fixed-length code becomes the decoded message of the variable-length code. If the decoded message of the fixed-length code is the erasure symbol, i.e., if $\widehat{\mathsf{M}} = \mathbf{x}$, then the encoder uses the fixed-length code again in the second n time units. By repeating this scheme until the decoded message of the fixed-length code is in $\mathcal{M}$, i.e., $\widehat{\mathsf{M}} \in \mathcal{M}$, we obtain a variable-length code.

Let L be the number of times the fixed-length code is used until a $\widehat{\mathsf{M}} \in \mathcal{M}$ is observed. Then, given the message M, L is a geometrically distributed random variable with success probability $(1 - P_{\mathbf{x}|\mathsf{M}})$ where $P_{\mathbf{x}|\mathsf{M}}$ is the conditional erasure probability of the fixed-length code given the message M. Then, the conditional probability distribution and the conditional expected value of L given M are

$$\mathbf{P}[\mathsf{L} = \mathsf{l} | \mathsf{M}] = (1 - P_{\mathbf{x}|\mathsf{M}})(P_{\mathbf{x}|\mathsf{M}})^{\mathsf{l}-1} \quad \mathsf{l} = 1, 2, \ldots \tag{25a}$$

$$\mathbf{E}[\mathsf{L} | \mathsf{M}] = (1 - P_{\mathbf{x}|\mathsf{M}})^{-1}. \tag{25b}$$

Furthermore, the conditional expected value of decoding time and the conditional error probability given the message M are

$$\mathbf{E}[\mathsf{T} | \mathsf{M}] = n\mathbf{E}[\mathsf{L} | \mathsf{M}] \tag{26a}$$

$$\mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M} \middle| \mathsf{M}\right] = P_{\mathrm{e}|\mathsf{M}}\mathbf{E}[\mathsf{L} | \mathsf{M}] \tag{26b}$$

where n is the block length of the fixed-length code and $P_{\mathrm{e}|\mathsf{M}}$ is the conditional error probability given the message M for the fixed-length code.

Thus, as result of (25b) and (26) and Lemma 3, we know that for any rate $\mathrm{R} \in [0, C]$, error exponent $\mathrm{E} \in [0, (1 - \frac{\mathrm{R}}{C})D]$, there exists a reliable sequence $\mathbb{Q}$ such that $\mathrm{R}_{\mathbb{Q}} = \mathrm{R}$, $\mathrm{E}_{\mathbb{Q}} = \mathrm{E}$, and

$$\mathrm{E}_{\mathrm{md}\,\mathbb{Q}} = \mathrm{E} + (1 - \tfrac{\mathrm{E}}{D})J\left(\tfrac{\mathrm{R}}{1 - \mathrm{E}/D}\right). \tag{27}$$

We show in Section V-C that for any reliable sequence $\mathbb{Q}$ with rate $\mathrm{R}_{\mathbb{Q}} = \mathrm{R}$ and error exponent $\mathrm{E}_{\mathbb{Q}} = \mathrm{E}$, $\mathrm{E}_{\mathrm{md},\mathbb{Q}}$ is upper bounded by the expression on the right-hand side of (27).

D. Bit-Wise UEP Achievability

In this section, we first use the family of codes described in Section IV-A and the control phase idea described in Section IV-B to construct fixed-length block codes with erasures which have *bit-wise* UEP. Then, we use them with a repeat until nonerasure decoding scheme, similar to the one described in Section IV-C, to obtain variable-length block codes with *bit-wise* UEP.

Let us start with describing the encoding scheme for the fixed-length block code with *bit-wise* UEP. If there are ℓ submessages, i.e., if $\mathcal{M} = (\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_\ell)$, then the encoding scheme has $\ell + 1$ phases with lengths $n_1, n_2, \ldots, n_{\ell+1}$ such that $n_1 + n_2 + \cdots + n_{\ell+1} = n$.

1) In the first phase, a length n_1 code from the family of codes described in Section IV-A is used. The message set of the code ${}_{\mathrm{t}}\mathcal{M}_1$ is $\mathcal{M}_1 \cup \{|\mathcal{M}_1| + 1\}$ and the message ${}_{\mathrm{t}}\mathsf{M}_1$ of the code is determined by the first submessage ${}_{\mathrm{t}}\mathsf{M}_1 = \mathsf{M}_1 + 1$. At the end of first phase, receiver uses the decoder of the length n_1 code to get a tentative decision ${}_{\mathrm{t}}\widehat{\mathsf{M}}_1$ which is known by the transmitter at the beginning of the second phase because of the feedback link.
2) In the second phase, a length n_2 code from the family of codes described in Section IV-A, with the message set ${}_{\mathrm{t}}\mathcal{M}_2 = \mathcal{M}_2 \cup \{|\mathcal{M}_2| + 1\}$, is used. If ${}_{\mathrm{t}}\widehat{\mathsf{M}}$ is decoded correctly at the end of the first phase, then the message ${}_{\mathrm{t}}\mathsf{M}_2$ of the code used in the second phase is determined by the second submessage as ${}_{\mathrm{t}}\mathsf{M}_2 = \mathsf{M}_2 + 1$, else ${}_{\mathrm{t}}\mathsf{M}_2 = 1$. At the end of the second phase, the receiver uses the decoder of the second phase code to get the tentative decision ${}_{\mathrm{t}}\widehat{\mathsf{M}}_2$ which is known by the transmitter at the beginning of the third phase because of the feedback link.
3) In phases 3 to ℓ, the previously described scheme is used. In phase i, a length n_i code, with the message set ${}_{\mathrm{t}}\mathcal{M}_i = \mathcal{M}_i \cup \{|\mathcal{M}_i| + 1\}$, from the family of codes described in Section IV-A is used. The message of the length n_i code ${}_{\mathrm{t}}\mathsf{M}_i$ is $\mathsf{M}_i + 1$ if ${}_{\mathrm{t}}\widehat{\mathsf{M}}_{i-1} = {}_{\mathrm{t}}\mathsf{M}_{i-1}$, 1 otherwise for $i = 3, 4, \ldots, \ell$.
4) The last phase is a $n_{\ell+1}$ long control phase, i.e., a $n_{\ell+1}$ long code with the message set ${}_{\mathrm{t}}\mathcal{M}_{\ell+1} = \{1, 2\}$ is used in the last phase. The codewords for the first and second messages are $n_{\ell+1}$ long sequences of input letters $\mathbf{r}$ and $\mathbf{a}$, respectively, where $\mathbf{r}$ and $\mathbf{a}$ are described in (10). The tentative decision in the last phase ${}_{\mathrm{t}}\widehat{\mathsf{M}}_{\ell+1}$ is equal to the first message if the output sequence in the last phase is not typical with $W_{\mathbf{a}}$, the second message otherwise. The message of the $n_{\ell+1}$ long code ${}_{\mathrm{t}}\mathsf{M}_{\ell+1}$ is equal to 2 if ${}_{\mathrm{t}}\widehat{\mathsf{M}}_\ell = {}_{\mathrm{t}}\mathsf{M}_\ell$, 1 otherwise.

Note that if we define ${}_{\mathrm{t}}\widehat{\mathsf{M}}_0$, ${}_{\mathrm{t}}\mathsf{M}_0$ and $\mathsf{M}_{\ell+1}$ all to be 1, i.e., ${}_{\mathrm{t}}\widehat{\mathsf{M}}_0 = {}_{\mathrm{t}}\mathsf{M}_0 = \mathsf{M}_{\ell+1} = 1$, we can write the following rule for determining the ${}_{\mathrm{t}}\mathsf{M}_i$'s for $i = 1$ to $\ell + 1$

$${}_{\mathrm{t}}\mathsf{M}_i = 1 + \mathbb{1}_{\left\{{}_{\mathrm{t}}\widehat{\mathsf{M}}_{i-1} = {}_{\mathrm{t}}\mathsf{M}_{i-1}\right\}}\mathsf{M}_i \quad i = 1, 2, \ldots, (\ell + 1). \tag{28}$$

It is important however to keep in mind that the last phase is a control phase and the codes in the first ℓ phases are from the family of codes described in Section IV-A.

Note that during the phases $i = 2$ to ℓ, erroneous transmission of ${}_{\mathrm{t}}\mathsf{M}_{i-1}$ is conveyed using ${}_{\mathrm{t}}\mathsf{M}_i = 1$; hence, the transmission of M_i through ${}_{\mathrm{t}}\mathsf{M}_i$, i.e., ${}_{\mathrm{t}}\mathsf{M}_i = 1 + \mathsf{M}_i$, is a tacit approval of the tentative decision ${}_{\mathrm{t}}\widehat{\mathsf{M}}_{i-1}$. Because of this, the aforementioned encoding scheme is said to have an implicit acceptance explicit rejection property. The idea of implicit acceptance explicit rejection was first introduced by Kudryashov in [7] in the context of nonblock variable-length codes with feedback and delay constraints.

After finishing the description of the encoding scheme, we are ready to describe the decoding scheme. The receiver determines the decoded message using the tentative decisions, ${}_{\mathrm{t}}\widehat{\mathsf{M}}_i$ for $i = 1$ to $\ell + 1$. If one or more of the tentative decisions are equal to 1, then an erasure is declared. If all $\ell + 1$ tentative decision are

different from 1, then $\widehat{\mathsf{M}}_i = {}_{\mathsf{t}}\widehat{\mathsf{M}}_i - 1$ for all $i = 1, 2, \ldots, \ell$. Hence, the decoding rule is

$$(\widehat{\mathsf{M}}_1, \widehat{\mathsf{M}}_2, \ldots, \widehat{\mathsf{M}}_\ell) = \begin{cases} ({}_{\mathsf{t}}\widehat{\mathsf{M}}_1 - 1, {}_{\mathsf{t}}\widehat{\mathsf{M}}_2 - 1, \ldots, {}_{\mathsf{t}}\widehat{\mathsf{M}}_\ell - 1) & \text{if } \prod_{i=1}^{\ell+1}({}_{\mathsf{t}}\widehat{\mathsf{M}}_i - 1) > 0 \\ \mathbf{x} & \text{if } \prod_{i=1}^{\ell+1}({}_{\mathsf{t}}\widehat{\mathsf{M}}_i - 1) = 0. \end{cases} \tag{29}$$

For *bit-wise* UEP codes with erasure, the definition of $P_{\mathbf{e}}(i)$ is slightly different from the original one given in (14)

$$P_{\mathbf{e}}(i) = \mathbf{P}\left[\{\widehat{\mathsf{M}}^i \neq \mathsf{m}^i, \widehat{\mathsf{M}} \neq \mathbf{x}\}\right]. \tag{30}$$

With this alternative definition in mind, let us define $P_{\mathbf{e}|\mathsf{m}}(i)$ as the conditional probability of the erroneous transmission of any one of the first i submessage when $\mathsf{M} = \mathsf{m}$

$$P_{\mathbf{e}|\mathsf{m}}(i) = \mathbf{P}\left[\{\widehat{\mathsf{M}}^i \neq \mathsf{m}^i, \widehat{\mathsf{M}} \neq \mathbf{x}\}\middle|\, \mathsf{M} = \mathsf{m}\right]. \tag{31}$$

The error analysis of the previously described fixed-length codes, presented in Section D in the Appendix, leads to Lemma 4 given in the following.

Lemma 4: For block length n, any integer $\ell \leq \frac{n}{\ln(1+n)}$, rate vector $\vec{\mathrm{R}}$, and time-sharing vector $\vec{\eta}$ such that

$$\mathrm{R}_i \leq C\eta_i \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{32a}$$
$$\eta_i \geq 0 \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{32b}$$
$$\sum\nolimits_{i=1}^{\ell} \eta_i \leq 1 \tag{32c}$$

there exists a length n block code such that

$$|\mathcal{M}_i| \geq e^{n(\mathrm{R}_i - \varepsilon_{n,\ell})} \quad \forall i \in \{1, 2, \ldots \ell\}$$
$$|\mathcal{M}^i| \geq e^{n(-\varepsilon_{n,\ell} + \sum_{j=1}^{i} \mathrm{R}_j)} \quad \forall i \in \{1, 2, \ldots \ell\}$$
$$P_{\mathbf{e}|\mathsf{m}}(i) \leq \varepsilon_{n,\ell} e^{-n\left(-\varepsilon_{n,\ell} + \sum_{j=i+1}^{\ell+1} \eta_j J\left(\frac{\mathrm{R}_j}{\eta_j}\right)\right)} \quad \forall \mathsf{m} \in \mathcal{M},\ i \in \{1, 2, \ldots \ell\}$$
$$P_{\mathbf{x}|\mathsf{m}} \leq \varepsilon_{n,\ell} \quad \forall \mathsf{m} \in \mathcal{M}$$

where

$$\eta_{\ell+1} = 1 - \sum_{i=1}^{\ell} \eta_i, \mathrm{R}_{\ell+1} = 0, \varepsilon_{n,\ell} = \frac{10|\mathcal{X}||\mathcal{Y}| \ln(\frac{e}{\lambda})\sqrt{\ln(1+n)}}{\sqrt{n}} \sqrt{1+\ell}.$$

Recall the repeat at erasures scheme described in Section IV-C. If we use that scheme to obtain a variable-length code from the fixed-length *bit-wise* UEP code described in Lemma 4, we obtain a variable-length code with UEP such that

$$\mathbf{E}[\mathsf{T} \,|\, \mathsf{M}] = \frac{n}{1 - P_{\mathbf{x}|\mathsf{M}}} \tag{33a}$$
$$\mathbf{P}\left[\widehat{\mathsf{M}}^i \neq \mathsf{M}^i \middle|\, \mathsf{M}\right] \leq \frac{P_{\mathbf{e}|\mathsf{M}}(i)}{1 - P_{\mathbf{x}|\mathsf{M}}} \quad i = 1, 2, \ldots, \ell. \tag{33b}$$

As a result of (33) and Lemma 4, we know that for any rate vector $\vec{\mathrm{R}}$, error exponent vector $\vec{\mathrm{E}}$, and time-sharing vector $\vec{\eta}$ such that

$$\mathrm{E}_i \leq (1 - \sum_{j=1}^{\ell} \eta_j) D + \sum_{j=i+1}^{\ell} \eta_j J\left(\frac{\mathrm{R}_j}{\eta_j}\right) \forall i \in \{1, 2, \ldots, \ell\} \tag{34a}$$
$$\mathrm{R}_i \leq C\eta_i \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{34b}$$
$$\eta_i \geq 0 \quad \forall i \in \{1, 2, \ldots, \ell\} \tag{34c}$$
$$\sum_{j=1}^{\ell} \eta_j \leq 1 \tag{34d}$$

there exists a reliable sequence $\mathbb{Q}$ such that $(\vec{\mathrm{R}}_{\mathbb{Q}}, \vec{\mathrm{E}}_{\mathbb{Q}}) = (\vec{\mathrm{R}}, \vec{\mathrm{E}})$. Thus, the existence of the time-sharing vector $\vec{\eta}$ satisfying the constraints given in (34) is a sufficient condition for the achievability of a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$. We show in Section V-D that the existence of a time-sharing vector $\vec{\eta}$ satisfying the constraints given in (34) is also a necessary condition for the achievability of a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$.

V. Converse

Berlin *et al.* [1] used the error probability of a random binary query posed at a stopping time for bounding the error probability of a variable-length block code. Later, similar techniques have been applied in [2] for establishing outer bounds in UEP problems. Our approach is similar to that of [1] and [2]; we, too, use error probabilities of random queries posed at stopping times for establishing outer bounds. Our approach, nevertheless, is novel because of the error events we choose to analyze and the bounding techniques we use. Furthermore, the relation we establish in Lemma 5 between the error probabilities and the decay rate of the conditional entropy of the messages with time is a brand new tool for UEP problems.

For rigorously and unambiguously generalizing the technique used in [1] and [2], we introduce the concept of anticipative list decoders (ALDs) in Section V-A. Then, in Section V-B, we bound the probabilities of certain error events associated with ALDs from the following. This bound, i.e., Lemma 5, is used in Sections V-C and V-D to derive tight outer bounds for the performance of variable-length block codes in the single-message *message-wise* UEP problem and in the *bit-wise* UEP problem, respectively.

A. Anticipative List Decoders (ALDs)

In this section, we first introduce the concepts of ALDs and nontrivial ALDs. After that, we show that for a given variable-length code, any nontrivial ALD $(\widetilde{\mathsf{T}}, \mathcal{A})$ can be used to define a probability distribution, $\mathtt{P}_{\{\mathcal{A}\}}$, on $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$. Finally, we use $\mathtt{P}_{\{\mathcal{A}\}}$ to define the probability measure $\mathbf{P}_{\{\mathcal{A}\}}[\cdot]$ for the events in $\wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}})$. Both the nontrivial ALDs $(\widetilde{\mathsf{T}}, \mathcal{A})$ and the probability measures $\mathbf{P}_{\{\mathcal{A}\}}[\cdot]$ associated with them play key roles in Lemma 5 of Section V-B.

An ALD for a variable-length code is a list decoder $\mathcal{A}$ that decodes at a stopping time $\widetilde{\mathsf{T}}$ that is always less than or equal to the decoding time of the code T. The ALDs are used to formulate

questions about the transmitted message or the decoded message, in terms of a subset of the message set $\mathcal{M}$ that is chosen at a stopping time $\widetilde{\mathsf{T}}$. For example, let $\mathcal{A}$ be the set of all $\mathsf{m} \in \mathcal{M}$ whose posterior probability at time one is larger than $1/|\mathcal{M}|$. Evidently, for all values of Y_1, $\mathcal{A}$ is a subset of $\mathcal{M}$, but it is not necessarily the same subset for all values of Y_1. Indeed, $\mathcal{A}$ is a function from $\mathcal{Y}_1$ to the power set of $\mathcal{M}$ and $(\widetilde{\mathsf{T}}, \mathcal{A})$ is an ALD, for which $\widetilde{\mathsf{T}} = 1$. Formal definition, for ALDs, is given in the following. In order to avoid separate treatment in certain special cases, we include the case when $\widetilde{\mathsf{T}} = 0$ and $\mathcal{A}$ is fixed subset of $\mathcal{M}$, in the definition.

Definition 13 (ALD): For a variable-length code with decoding time T, a pair $(\widetilde{\mathsf{T}}, \mathcal{A})$ is called an ALD if

1) either $\widetilde{\mathsf{T}}$ is the constant random variable 0 and $\mathcal{A}$ is a fixed subset of $\mathcal{M}$, i.e.,
$$\widetilde{\mathsf{T}} = 0$$
$$\mathcal{A} \in \wp(\mathcal{M})$$

2) or $\widetilde{\mathsf{T}}$ is a stopping time, which is smaller than T with probability one, and $\mathcal{A}$ is a $\wp(\mathcal{M})$-valued function defined on $\mathcal{Y}^{\widetilde{\mathsf{T}}}$, i.e.,
$$\mathbf{P}\left[\widetilde{\mathsf{T}} \leq \mathsf{T}\right] = 1$$
$$\mathcal{A} : \mathcal{Y}^{\widetilde{\mathsf{T}}} \to \wp(\mathcal{M}).$$

Definition of ALD does not require $\mathcal{A}$ to be of some fixed size, nor it requires $\mathcal{A}$ to include more likely or less likely messages. Thus, for certain values of $\mathsf{Y}^{\widetilde{\mathsf{T}}}$, $\mathcal{A}$ might not include any $\mathsf{m} \in \mathcal{M}$ with positive posterior probability. In other words, for some values of $\mathsf{Y}^{\widetilde{\mathsf{T}}}$, we might have

$$\mathbf{P}\left[\mathsf{M} \in \mathcal{A}(\mathsf{Y}^{\widetilde{\mathsf{T}}}) \middle| \mathsf{Y}^{\widetilde{\mathsf{T}}} = \mathsf{y}^{\tilde{\mathsf{t}}}\right] = 0.$$

The ALD's in which such $\mathsf{y}^{\tilde{\mathsf{t}}}$'s have zero probability are called nontrivial ALD's.

Definition 14 (Nontrivial ALD): An ALD $(\widetilde{\mathsf{T}}, \mathcal{A})$ is called a nontrivial anticipative list decoder (*NALD*) if $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}(\mathsf{Y}^{\widetilde{\mathsf{T}}}) \middle| \mathsf{Y}^{\widetilde{\mathsf{T}}}\right] > 0$ with probability one, i.e.,
$$\mathbf{P}\left[\mathbf{P}\left[\mathsf{M} \in \mathcal{A}(\mathsf{Y}^{\widetilde{\mathsf{T}}}) \middle| \mathsf{Y}^{\widetilde{\mathsf{T}}}\right] > 0\right] = 1. \tag{35}$$

In the following, for any variable-length code and an associated NALD $(\widetilde{\mathsf{T}}, \mathcal{A})$, we define a probability distribution $\mathtt{P}_{\{\mathcal{A}\}}$ on $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$ and a probability measure $\mathbf{P}_{\{\mathcal{A}\}}[\cdot]$ for the events in $\wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}})$. For doing that, first note that the probability measure generated by the code, i.e., $\mathbf{P}[\cdot]$, can be used to define a probability distribution $\mathtt{P}$ on $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$ as follows:

$$\mathtt{P}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \stackrel{\triangle}{=} \mathbf{P}\left[\mathsf{M} = \mathsf{m}, \mathsf{Y}^{\mathsf{T}} = \mathsf{y}^{\mathsf{t}}\right] \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*} \tag{36}$$

where $\mathcal{Y}^{\mathsf{T}*}$ is a countable set for any stopping time, given in (6b).

As T is a stopping time, the probability of any event Γ in $\wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}})$ under $\mathbf{P}[\cdot]$, i.e., $\mathbf{P}[\Gamma]$, is equal to

$$\mathbf{P}[\Gamma] = \sum_{(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \in \Gamma \cap (\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*})} \mathtt{P}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}). \tag{37}$$

Evidently, we can extend the definition of $\mathtt{P}$ and assume that $\mathtt{P}$ is zero whenever y^{t} is in $\mathcal{Y}^{\infty}_{\{\mathsf{T}=\infty\}}$, i.e.,

$$\mathtt{P}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \stackrel{\triangle}{=} 0 \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\infty}_{\{\mathsf{T}=\infty\}}. \tag{38}$$

This extension is neither necessary nor relevant for calculating the probabilities of the events in $\wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}})$, because T is a stopping time, i.e., $\mathbf{P}[\mathsf{T} < \infty] = 1$.

Definition 15: Given a variable-length code with decoding time T, for any *NALD* $(\widetilde{\mathsf{T}}, \mathcal{A})$ let $\mathtt{P}_{\{\mathcal{A}\}}$ be[26]

$$\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \stackrel{\triangle}{=} \mathtt{P}(\mathsf{y}^{\tilde{\mathsf{t}}}) \frac{\mathtt{P}(\mathsf{m}|\mathsf{y}^{\tilde{\mathsf{t}}}) \mathbb{1}_{\{\mathsf{m} \in \mathcal{A}(\mathsf{y}^{\tilde{\mathsf{t}}})\}}}{\sum_{\widetilde{\mathsf{m}} \in \mathcal{M}} \mathtt{P}(\widetilde{\mathsf{m}}|\mathsf{y}^{\tilde{\mathsf{t}}}) \mathbb{1}_{\{\widetilde{\mathsf{m}} \in \mathcal{A}(\mathsf{y}^{\tilde{\mathsf{t}}})\}}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}_{\tilde{\mathsf{t}}+1}|\mathsf{y}^{\tilde{\mathsf{t}}}, \mathsf{m})$$
$$\forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}. \tag{39}$$

Note that Definition 15 is a parametric definition in the sense that it assigns a $\mathtt{P}_{\{\mathcal{A}\}}$ for all NALDs $(\widetilde{\mathsf{T}}, \mathcal{A})$. While proving outer bounds, we will employ not one but multiple *NALD*'s and use them in conjunction with our new result, i.e., Lemma 5. But before introducing Lemma 5, let us elaborate on the relations between marginal and conditional distributions of $\mathtt{P}_{\{\mathcal{A}\}}$ and $\mathtt{P}$.

For $\mathtt{P}_{\{\mathcal{A}\}}$ defined in (39), we have

$$\sum_{\mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) = 1.$$

Hence, $\mathtt{P}_{\{\mathcal{A}\}}$ is a probability distribution on $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$, i.e., $\mathtt{P}_{\{\mathcal{A}\}} \in P(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*})$.

Note that the marginal distributions of $\mathtt{P}_{\{\mathcal{A}\}}$ and $\mathtt{P}$ are the same on $\mathcal{Y}^{\widetilde{\mathsf{T}}*}$. Furthermore, for all $\mathsf{y}^{\tilde{\mathsf{t}}} \in \mathcal{Y}^{\widetilde{\mathsf{T}}*}$ and $\mathsf{m} \in \mathcal{M}$, the conditional distributions of $\mathtt{P}_{\{\mathcal{A}\}}$ and $\mathtt{P}$ are the same on $\mathcal{Y}^{\mathsf{T}*}_{\widetilde{\mathsf{T}}*}$. The probability distributions $\mathtt{P}_{\{\mathcal{A}\}}$ and $\mathtt{P}$ differ only in their conditional distributions on $\mathcal{M}$ given $\mathsf{y}^{\tilde{\mathsf{t}}}$. More specifically

$$\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{y}^{\tilde{\mathsf{t}}}) = \mathtt{P}(\mathsf{y}^{\tilde{\mathsf{t}}}) \qquad \forall \mathsf{y}^{\tilde{\mathsf{t}}} \in \mathcal{Y}^{\widetilde{\mathsf{T}}*} \tag{40a}$$

$$\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}|\mathsf{y}^{\tilde{\mathsf{t}}}) = \frac{\mathtt{P}(\mathsf{m}|\mathsf{y}^{\tilde{\mathsf{t}}}) \mathbb{1}_{\{\mathsf{m} \in \mathcal{A}(\mathsf{y}^{\tilde{\mathsf{t}}})\}}}{\sum_{\widetilde{\mathsf{m}} \in \mathcal{M}} \mathtt{P}(\widetilde{\mathsf{m}}|\mathsf{y}^{\tilde{\mathsf{t}}}) \mathbb{1}_{\{\widetilde{\mathsf{m}} \in \mathcal{A}(\mathsf{y}^{\tilde{\mathsf{t}}})\}}} \quad \forall \mathsf{y}^{\tilde{\mathsf{t}}} \in \mathcal{Y}^{\widetilde{\mathsf{T}}*}, \forall \mathsf{m} \in \mathcal{M} \tag{40b}$$

$$\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{y}^{\mathsf{t}}_{\tilde{\mathsf{t}}+1}|\mathsf{y}^{\tilde{\mathsf{t}}}, \mathsf{m}) = \mathtt{P}(\mathsf{y}^{\mathsf{t}}_{\tilde{\mathsf{t}}+1}|\mathsf{y}^{\tilde{\mathsf{t}}}, \mathsf{m}) \qquad \forall \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}, \forall \mathsf{m} \in \mathcal{M}. \tag{40c}$$

Using the parametric definition of probability distribution $\mathtt{P}_{\{\mathcal{A}\}}$ on $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$, we define a probability measure $\mathbf{P}_{\{\mathcal{A}\}}[\cdot]$ for the events in $\wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}})$ as follows:

$$\mathbf{P}_{\{\mathcal{A}\}}[\Gamma] \stackrel{\triangle}{=} \sum_{(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \in \Gamma \cap (\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*})} \mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \quad \forall \Gamma \in \wp(\mathcal{M} \times \mathcal{Y}^{\mathsf{T}}). \tag{41}$$

[26]There is a slight abuse of notation in (39); if $\widetilde{\mathsf{T}}$ is not a stopping time but rather a constant random variable $\widetilde{\mathsf{T}} = 0$, $\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}})$ should be interpreted as

$$\mathtt{P}_{\{\mathcal{A}\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \stackrel{\triangle}{=} \frac{\mathbb{1}_{\{\mathsf{m} \in \mathcal{A}\}}}{|\mathcal{A}|} \mathtt{P}(\mathsf{y}^{\mathsf{t}}|\mathsf{m}) \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}.$$

Evidently, we can extend the definition of $\mathsf{P}_{\{\mathcal{A}\}}$ to $\mathcal{M}\times\mathcal{Y}^{\mathsf{T}}$ by defining it to be zero on $\mathcal{M}\times\mathcal{Y}^{\infty}_{\{\mathsf{T}=\infty\}}$, i.e.,

$$\mathsf{P}_{\{\mathcal{A}\}}(\mathsf{m},\mathsf{y}^{\mathsf{t}})\stackrel{\triangle}{=}0 \quad \forall \mathsf{m}\in\mathcal{M},\mathsf{y}^{\mathsf{t}}\in\mathcal{Y}^{\infty}_{\{\mathsf{T}=\infty\}}. \tag{42}$$

As in the case of P, this extension is neither necessary nor relevant for calculating the probabilities $\mathbf{P}_{\{\mathcal{A}\}}[\Gamma]$ given in (41).

B. Error Probability and Decay Rate of Entropy

In this section, we lower bound the probability of the event that the decoded message $\widehat{\mathsf{M}}$ is in $\mathcal{A}$ under the probability measure $\mathbf{P}_{\{\mathcal{A}\}}[\cdot]$, i.e., $\mathbf{P}_{\{\mathcal{A}\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}(\mathsf{Y}^{\widetilde{\mathsf{T}}})\right]$. The bounds we derive depend on the decay rate of the conditional entropy of the messages in the interval between $\widetilde{\mathsf{T}}$ and T.

Before even stating our bound, we need to specify what we mean by the conditional entropy of the messages. While defining the conditional entropy, many authors do take an average over the sample values of the conditioned random variable and obtain a constant. We, however, do not take an average over the conditioned random variable and define conditional entropy as a random variable itself, which is a function of the random variable that is conditioned on[27]

$$\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\tau})\stackrel{\triangle}{=}\sum_{\mathsf{m}\in\mathcal{M}}\mathbf{P}[\mathsf{M}=\mathsf{m}|\,\mathsf{Y}^{\tau}]\ln\tfrac{1}{\mathbf{P}[\mathsf{M}=\mathsf{m}|\mathsf{Y}^{\tau}]}. \tag{43}$$

Using the probability distribution P defined in (36), we see that the conditional entropy defined in (43) is equal to

$$\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\tau})=\mathbf{E}\left[\ln\tfrac{1}{\mathsf{P}(\mathsf{M}|\mathsf{Y}^{\tau})}\,\middle|\,\mathsf{Y}^{\tau}\right]. \tag{44}$$

Lemma 5: For any variable-length block code with finite expected decoding time, $\mathbf{E}[\mathsf{T}]<\infty$, let $(\mathsf{T}_1,\mathcal{A}_1),(\mathsf{T}_2,\mathcal{A}_2),\ldots,(\mathsf{T}_k,\mathcal{A}_k)$ be k NALDs[28] such that

$$\mathbf{P}[\{0\le\mathsf{T}_1\le\mathsf{T}_2\le\cdots\le\mathsf{T}_k\le\mathsf{T}\}]=1. \tag{45}$$

Then, for all i in $\{1,2,\ldots,k\}$ such that $\left(\mathbf{P}\left[\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]+P_{\mathbf{e}}\right)\le 1/2$ we have (46) shown at the bottom of the page where $\mathsf{T}_0=0$, $\mathsf{T}_{k+1}=\mathsf{T}$, and for all j in $\{1,2,\ldots,(k+1)\}$, r_j's are given by

$$\mathrm{r}_j=\left\{\begin{array}{ll}0 & \text{if } \mathbf{P}[\mathsf{T}_j=\mathsf{T}_{j-1}]=1\\ \frac{\mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_{j-1}})-\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_j})\right]}{\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]} & \text{if } \mathbf{P}[\mathsf{T}_j=\mathsf{T}_{j-1}]<1\end{array}\right\}. \tag{47}$$

Proof of Lemma 5 is presented in Section E in the Appendix.

Before presenting the application of Lemma 5 in UEP problems, let us elaborate on its hypothesis and ramifications. We assumed that $(\mathsf{T}_i,\mathcal{A}_i)$ are all NALDs. Thus, for each $(\mathsf{T}_i,\mathcal{A}_i)$, the set of all $\mathsf{y}^{\mathsf{t}_i}\in\mathcal{Y}^{\mathsf{T}_i}$ such that the transmitted message is guaranteed to be outside $\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})$ has zero probability and there is an associated probability measure $\mathbf{P}_{\{\mathcal{A}_i\}}[\cdot]$ given in (41). Furthermore, $\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$ is the probability of the event that decoded message $\widehat{\mathsf{M}}$ is not in $\mathcal{A}_i$ under the probability measure $\mathbf{P}_{\{\mathcal{A}_i\}}[\cdot]$.

Condition given in (45) ensures that the decoding times of the k NALD's we are considering, $\mathsf{T}_1,\mathsf{T}_2,\ldots,\mathsf{T}_k$, are reached in their indexing order and before the decoding time of the variable-length code T. Any $\mathsf{T}_1,\mathsf{T}_2,\ldots,\mathsf{T}_k$ satisfying (45) divides the time interval between 0 and T into $k+1$ disjoint intervals. The duration of these intervals as well as the decrease of the conditional entropy during them is random. For the jth interval, the expected values of the duration and the decrease in the conditional entropy are given by $\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]$ and $\mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_{j-1}})-\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_j})\right]$, respectively. Hence, r_j's defined in (47) are rate of decrease of the conditional entropy of the messages per unit time in different intervals.

Lemma 5 bounds the probability of $\widehat{\mathsf{M}}$ being outside $\mathcal{A}_i$ under $\mathbf{P}_{\{\mathcal{A}_i\}}[\cdot]$ from below in terms of r_j's and $\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]$'s for $j>i$. The bound on $\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$ also depends on $\mathbf{P}\left[\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$ and $P_{\mathbf{e}}$. But the particular choice of $\mathcal{A}_j$'s for $j\neq i$ has no effect on the bound. This feature of the bound is its main merit over bounds resulting from the previously suggested techniques.

C. Single-Message Message-Wise UEP Converse

In this section, we bound the conditional error probabilities of the messages, i.e., $P_{\mathbf{e}|\mathsf{m}}$'s, from below uniformly over the message set $\mathcal{M}$ in a variable-length block code with average error probability $P_{\mathbf{e}}$, using Lemma 5. Resulting outer bound reveals that the inner bound we obtained in Section IV-C for the single-message *message-wise UEP* problem is tight.

Consider a variable-length block code with finite expected decoding time, i.e., $\mathbf{E}[\mathsf{T}]<\infty$. In order to bound $P_{\mathbf{e}|\mathsf{m}}$, defined

$$\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]\ge\exp\left(\frac{-h\left(P_{\mathbf{e}}+\mathbf{P}\left[\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]\right)-\sum_{j=i+1}^{k+1}\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]\,J(\mathrm{r}_j)}{1-P_{\mathbf{e}}-\mathbf{P}[\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})]}\right) \tag{46}$$

[27]Recall the standard notation in probability theory about the conditional expectations and conditional probabilities. Let H be a real-valued random variable and G be a random quantity that takes values from a finite set $\mathcal{G}$, such that $\mathbf{P}[\mathsf{G}=\mathsf{g}]>0$ for all $\mathsf{g}\in\mathcal{G}$. Then, unlike $\mathbf{E}[\mathsf{H}]$, which is constant, $\mathbf{E}[\mathsf{H}|\,\mathsf{G}]$ is a random variable. Thus, an equation of the form $\mathsf{Z}=\mathbf{E}[\mathsf{H}|\,\mathsf{G}]$ implies not the equality of two constants but the equality of two random variables, i.e., it means that $\mathsf{z}=\mathbf{E}[\mathsf{H}|\,\mathsf{G}=\mathsf{g}]$ for all $\mathsf{g}\in\mathcal{G}$. Similarly, let $\mathcal{H}_1$ be a set of sample values of the random variable H; then, unlike $\mathbf{P}[\mathsf{H}\in\mathcal{H}_1]$, which is a constant, $\mathbf{P}[\mathsf{H}\in\mathcal{H}_1|\,\mathsf{G}]$ is a random variable. Equations (43) and (44) are such equations. Explaining conditional expectations and conditional probabilities are beyond the scope of this paper; readers who are not sufficiently fluent with these concepts are encouraged to read [10, Ch. I, Sec. VIII], which deal the case where random variables can take finitely many values. Appropriately generalized formal treatment of the subject in terms of sigma fields is presented in [10, Ch. II, Sec. VII].

[28]Recall ALD's and NALD's are defined in Definitions 13 and 14, respectively.

in (11), from the following, we apply Lemma 5 for $k = 2$ with $(\mathsf{T}_1, \mathcal{A}_1)$, $(\mathsf{T}_2, \mathcal{A}_2)$ given in the following.

1) Let T_1 be zero and $\mathcal{A}_1$ be $\{\mathsf{m}\}$, i.e.,

$$\mathsf{T}_1 = 0 \tag{48}$$

$$\mathcal{A}_1 = \{\mathsf{m}\}. \tag{49}$$

2) Let T_2 be the first time instance before T such that one message, not necessarily the one chosen for $\mathcal{A}_1$, i.e., m, has a posterior probability $1-\delta$ or higher

$$\mathsf{T}_2 \stackrel{\triangle}{=} \min\{\tau : \max_{\widetilde{\mathsf{m}}} \mathbf{P}[\mathsf{M} = \widetilde{\mathsf{m}} | \, \mathsf{Y}^\tau] \geq (1-\delta) \text{ or } \tau = \mathsf{T}\}. \tag{50}$$

Let $\mathcal{A}_2$ be the set of all messages whose posterior probability at time T_2 is less then $(1-\delta)$

$$\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \stackrel{\triangle}{=} \{\widetilde{\mathsf{m}} \in \mathcal{M} : \mathbf{P}\left[\mathsf{M} = \widetilde{\mathsf{m}} | \, \mathsf{Y}^{\mathsf{T}_2}\right] < (1-\delta)\}. \tag{51}$$

We apply Lemma 5 for $(\mathsf{T}_1, \mathcal{A}_1)$ and $(\mathsf{T}_2, \mathcal{A}_2)$ given in (48)–(51). Then, using the fact that $J(\cdot) \leq D$, we get

$$\ln P_{\mathbf{e}|\mathsf{m}} \geq \frac{-h\left(P_{\mathbf{e}} + |\mathcal{M}|^{-1}\right) - \mathbf{E}[\mathsf{T}_2] J\left(\frac{\mathbf{E}\left[\mathsf{H}(\mathsf{M}) - \mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right]}{\mathbf{E}[\mathsf{T}_2]}\right) - \mathbf{E}[\mathsf{T} - \mathsf{T}_2] D}{1 - P_{\mathbf{e}} - |\mathcal{M}|^{-1}} \tag{52a}$$

$$\ln \mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right] \geq \frac{-h\left(P_{\mathbf{e}} + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]\right) - \mathbf{E}[\mathsf{T} - \mathsf{T}_2] D}{1 - P_{\mathbf{e}} - \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]}. \tag{52b}$$

If $\delta < 1/2$, one can show $\mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]$ is roughly equal to $P_{\mathbf{e}}/\delta$. Thus, inequality in (52b) becomes a lower bound on $\mathbf{E}[\mathsf{T} - \mathsf{T}_2]$ in terms of $P_{\mathbf{e}}$. It can be shown that the lower bound (52a) takes its smallest value for the smallest value of $\mathbf{E}[\mathsf{T} - \mathsf{T}_2]$. Then, using Fano's inequality for $\mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right]$, we obtain Lemma 6 given in the following.

A complete proof of Lemma 6 for variable-length block codes with finite expected decoding time is presented in Section F in the Appendix. For variable-length block codes with infinite expected decoding time, Lemma 6 follows from the lower bounds on $P_{\mathbf{e}}$ and $P_{\mathbf{e}|\mathsf{m}}$ derived in Sections H1 and H2 in the Appendix.

Lemma 6: For any variable-length block code and positive δ such that $P_{\mathbf{e}} + \delta + \frac{P_{\mathbf{e}}}{\delta} + |\mathcal{M}|^{-1} \leq 1/2$

$$-\frac{\ln P_{\mathbf{e}|\mathsf{m}}}{\mathbf{E}[\mathsf{T}]} \leq \mathrm{E} + (1 - \tfrac{\mathrm{E}-\tilde{\epsilon}}{D}) J\left(\frac{\mathrm{R}-\tilde{\epsilon}}{1-\frac{\mathrm{E}-\tilde{\epsilon}}{D}}\right) \qquad \forall \mathsf{m} \in \mathcal{M} \tag{53}$$

where $\mathrm{R} = \frac{|\mathcal{M}|}{\mathbf{E}[\mathsf{T}]}$, $\mathrm{E} = \frac{-\ln P_{\mathbf{e}}}{\mathbf{E}[\mathsf{T}]}$, $\tilde{\epsilon} = \frac{\tilde{\epsilon}_1 D + \tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$, $\tilde{\epsilon}_1 = P_{\mathbf{e}} + \delta + \frac{P_{\mathbf{e}}}{\delta} + |\mathcal{M}|^{-1}$ and $\tilde{\epsilon}_2 = \frac{h(\tilde{\epsilon}_1) - \ln \lambda \delta}{\mathbf{E}[\mathsf{T}]}$.

Lemma 6 is a generalization of [2, Th. 8] and [2, Lemma 1]. While deriving bounds given in [2, Th. 8] and [2, Lemma 1], no attention is paid to the fact that the rate of decrease of the conditional entropy of the messages can be different in different time intervals. As a result, both [2, Th. 8] and [2, Lemma 1] are tight only when the error exponent is very close to zero. While deriving the bound given in Lemma 6, on the other hand, the variation in the rate the conditional entropy decreases in different intervals is taken into account. Hence, the outer bound given in Lemma 6 matches the inner bound given in Section IV-C for all achievable values of error exponent, $0 \leq \mathrm{E} \leq (1 - \frac{\mathrm{R}}{C})D$.

Consider a reliable sequence of codes $\mathbb{Q}$ with rate $\mathrm{R}_{\mathbb{Q}}$ and error exponent $\mathrm{E}_{\mathbb{Q}}$. Then, if we apply Lemma 6 with $\delta = \frac{1}{\ln(1/P_{\mathbf{e}})}$, we get

$$\mathrm{E}_{\mathrm{md},\mathbb{Q}} \leq \mathrm{E}_{\mathbb{Q}} + (1 - \tfrac{\mathrm{E}_{\mathbb{Q}}}{D}) J\left(\frac{\mathrm{R}_{\mathbb{Q}}}{1 - \mathrm{E}_{\mathbb{Q}}/D}\right). \tag{54}$$

Note that the upper bound on $\mathrm{E}_{\mathrm{md},\mathbb{Q}}$'s given in (54) is achievable by at least one $\mathbb{Q}$ described in Section IV-C.

D. Bit-Wise UEP Converse

In this section, we apply Lemma 5 to a variable-length block code with a message set $\mathcal{M}$ of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_\ell$, in order to obtain lower bounds on $P_{\mathbf{e}}(i)$'s for $i = 1, 2, \ldots, \ell$ in terms of the sizes of the submessage sets $|\mathcal{M}_1|, |\mathcal{M}_2|, \ldots, |\mathcal{M}_\ell|$ and the expected decoding time $\mathbf{E}[\mathsf{T}]$. When applied to reliable code sequences, these bounds on $P_{\mathbf{e}}(i)$'s in terms of $|\mathcal{M}_i|$'s and $\mathbf{E}[\mathsf{T}]$ give a necessary condition for the achievability of a rate vector and error exponent vector pair $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ that matches the sufficient condition for the achievability derived in Section IV-D.

In order to bound $P_{\mathbf{e}}(i)$'s, we use Lemma 5 with ℓ *NALD*'s, $(\mathsf{T}_1, \mathcal{A}_1), \ldots, (\mathsf{T}_\ell, \mathcal{A}_\ell)$. Let us start with defining T_i's and $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$'s.

1) For any i in $\{1, 2, \ldots, \ell\}$, let T_i be the first time instance that a member of $\mathcal{M}^i$ gains a posterior probability larger than or equal to $(1-\delta)$ if it happens before T, T otherwise

$$\mathsf{T}_i \stackrel{\triangle}{=} \min\{\tau : \max_{\mathsf{m}^i} \mathbf{P}\left[\mathsf{M}^i = \mathsf{m}^i \big| \, \mathsf{Y}^\tau\right] \geq 1-\delta \text{ or } \tau = \mathsf{T}\}. \tag{55}$$

2) For any i in $\{1, 2, \ldots, \ell\}$, let $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$ be the set of all messages of the form $\mathsf{m} = (\mathsf{m}^i, \mathsf{m}_{i+1}, \ldots, \mathsf{m}_\ell)$ for which posterior probability of m^i is less than $(1-\delta)$ at T_i

$$\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \stackrel{\triangle}{=} \{(\mathsf{m}^i, \mathsf{m}_{i+1}, \ldots, \mathsf{m}_\ell) \in \mathcal{M} : \mathbf{P}\left[\mathsf{M}^i = \mathsf{m}^i \big| \, \mathsf{Y}^{\mathsf{T}_i}\right] < 1 - \delta\}. \tag{56}$$

If we apply Lemma 5 for $(\mathsf{T}_1, \mathcal{A}_1), \ldots, (\mathsf{T}_\ell, \mathcal{A}_\ell)$ defined in (55) and (56), we obtain lower bounds on $\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$'s in terms of $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$'s and r_j's and $\mathbf{E}[\mathsf{T}_j - \mathsf{T}_{j+1}]$'s for $j > i$. In order to turn these bounds into bounds on $P_{\mathbf{e}}(i)$'s, we bound $\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$'s and $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$'s from above.

1) The posterior probability of a message at time $\tau + 1$ cannot be smaller than λ times its value at time τ because $\min_{\mathsf{x} \in \mathcal{X}, \mathsf{y} \in \mathcal{Y}} W_{\mathsf{x}}(\mathsf{y}) = \lambda$. Thus, if $\delta < 1/2$, one can bound $\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$'s from above

$$\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right] < \tfrac{1}{\lambda\delta} P_{\mathbf{e}}(i) \quad \forall i \in \{1, 2, \ldots, \ell\}. \tag{57}$$

2) Note that if at T_i there is a m^i with posterior probability $(1-\delta)$, then $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \big| \, \mathsf{Y}^{\mathsf{T}_i}\right] \leq \delta$. If at T_i there is no m^i with posterior probability $(1-\delta)$, then $\mathbf{P}\left[\widehat{\mathsf{M}}^i \neq \mathsf{M}^i \big| \, \mathsf{Y}^{\mathsf{T}_i}\right] \geq \delta$. Using these facts, one can bound $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]$ from above

$$\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right] \leq \tfrac{P_{\mathbf{e}}}{\delta} + \delta \quad \forall i \in \{1, 2, \ldots, \ell\}. \tag{58}$$

More detailed derivations of the inequalities given in (57) and (58) can be found in Section G in the Appendix.

Using (57) and (58) together with Lemma 5, we can conclude that

$$\ln P_{\mathrm{e}}(i) \geq \ln(\lambda\delta) + \frac{-h(P_{\mathrm{e}}+\delta+P_{\mathrm{e}}/\delta)-\sum_{j=i+1}^{\ell+1}\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]J(\mathrm{r}_j)}{1-P_{\mathrm{e}}-\delta-P_{\mathrm{e}}/\delta} \quad \forall i \in \{1,2,\ldots,\ell\} \tag{59}$$

provided that $P_{\mathrm{e}}+\delta+P_{\mathrm{e}}/\delta \leq 1/2$, where r_j's are defined in (47).

Note that the lower bound on $P_{\mathrm{e}}(i)$'s given in (59) takes different values depending on the rate of decrease of the conditional entropy of the messages in different intervals, i.e., r_j's, and the expected duration of different intervals, i.e., $\mathbf{E}[\mathsf{T}_j-\mathsf{T}_{j-1}]$'s. Making a worst case assumption on the rate of decrease of entropy and the durations of the intervals, one can obtain Lemma 7 given in the following.

A complete proof of Lemma 7 for variable-length block codes with finite expected decoding time is presented in Section G in the Appendix. For variable-length block codes with infinite expected decoding time, Lemma 7 follows from the lower bounds on P_{e} and $P_{\mathrm{e}}(i)$'s derived in Sections H1 and H3 in the Appendix.

Lemma 7: For any variable-length block code with feedback with a message set $\mathcal{M}$ of the form[29] $\mathcal{M}=\mathcal{M}_1\times\mathcal{M}_2\times\cdots,\mathcal{M}_\ell$ and for any positive δ such that $P_{\mathrm{e}}+\delta+\frac{P_{\mathrm{e}}}{\delta}\leq\frac{1}{5}$, we have

$$(1-\tilde{\epsilon}_3)\mathrm{E}_i-\tilde{\epsilon}_5 \leq (1-\sum_{j=1}^{\ell}\eta_j)D + \sum_{j=i+1}^{\ell}\eta_j J\Big(\frac{(1-\tilde{\epsilon}_3)\mathrm{R}_j}{\eta_j}\Big) \quad i=1,2,\ldots,\ell \tag{60a}$$

$$(1-\tilde{\epsilon}_3)\mathrm{R}_i-\tilde{\epsilon}_4\mathbb{1}_{\{i=1\}} \leq C\eta_i \quad i=1,2,\ldots,\ell \tag{60b}$$

for some time-sharing vector $\vec{\eta}$ such that

$$\eta_i \geq 0 \quad i=1,2,\ldots,\ell \tag{61a}$$

$$\sum_{j=1}^{\ell}\eta_j \leq 1 \tag{61b}$$

where $\mathrm{R}_i=\frac{|\ln\mathcal{M}_i|}{\mathbf{E}[\mathsf{T}]}$, $\mathrm{E}_i=\frac{-\ln P_{\mathrm{e}}(i)}{\mathbf{E}[\mathsf{T}]}$, $\tilde{\epsilon}_3=P_{\mathrm{e}}+\delta+\frac{P_{\mathrm{e}}}{\delta}$, $\tilde{\epsilon}_4=\frac{h(\tilde{\epsilon}_3)}{\mathbf{E}[\mathsf{T}]}$, and $\tilde{\epsilon}_5=\frac{h(\tilde{\epsilon}_3)-\ln\lambda\delta}{\mathbf{E}[\mathsf{T}]}$.

For any reliable sequence $\mathbb{Q}$ whose message sets $\mathcal{M}^{(\kappa)}$ are of the form $\mathcal{M}^{(\kappa)}=\mathcal{M}_1^{(\kappa)}\times\mathcal{M}_2^{(\kappa)}\times\cdots\times\mathcal{M}_\ell^{(\kappa)}$, if we set δ to $\delta=\frac{1}{-\ln P_{\mathrm{e}}}$, Lemma 7 implies that there exists a $\vec{\eta}$ such that[30]

$$\mathrm{E}_{\mathbb{Q},i} \leq (1-\sum_{j=1}^{k}\eta_j)D+\sum_{j=i+1}^{\ell}\eta_j J\Big(\frac{\mathrm{R}_{\mathbb{Q},j}}{\eta_j}\Big) \quad \forall i\in\{1,2,\ldots,\ell\} \tag{62a}$$

$$\mathrm{R}_{\mathbb{Q},i} \leq C\eta_i \quad \forall i\in\{1,2,\ldots,\ell\} \tag{62b}$$

$$\eta_i \geq 0 \quad \forall i\in\{1,2,\ldots,\ell\} \tag{62c}$$

$$\sum_{j=1}^{\ell}\eta_j \leq 1. \tag{62d}$$

[29]We tacitly assume, without loss of generality, that $|\mathcal{M}_1|\geq 2$.

[30]This fact is far from trivial; yet, it is intuitive to all who has worked with sequences of vectors in a bounded subset of $\mathbf{R}^\ell$ where $\mathbf{R}^\ell$ is the ℓ-dimensional real vector space with the norm $\left\|\vec{X}\right\|=\sup_j|x_j|$ For details, see Section J in the Appendix.

Recall that a rate-exponent vector $(\vec{\mathrm{R}},\vec{\mathrm{E}})$ is achievable only if there exists a reliable code sequence $\mathbb{Q}$ such that $(\vec{\mathrm{R}}_{\mathbb{Q}},\vec{\mathrm{E}}_{\mathbb{Q}})=(\vec{\mathrm{R}},\vec{\mathrm{E}})$. Thus, a rate-exponent vector $(\vec{\mathrm{R}},\vec{\mathrm{E}})$ is achievable only if there exists a time-sharing vector $\vec{\eta}$ satisfying (34). In other words, the sufficient condition for the achievability of $(\vec{\mathrm{R}},\vec{\mathrm{E}})$ we have derived in Section IV-D is also a necessary condition.

VI. Conclusion

We have considered the single-message *message-wise* and the fixed ℓ *bit-wise* UEP problems and characterized the achievable rate error exponent regions completely for both of the problems.

In the *bit-wise* UEP problem, we have observed that encoding schemes decoupling the communication and bulk of the error correction both at the transmitter and at the receiver can achieve optimal performance. This result is extending the similar observations made for conventional variable-length block coding schemes without UEP. However, for doing that one needs to go beyond the idea of communication phase and control phase introduced in [12], and harness the implicit confirmation explicit rejection schemes, introduced by Kudryashov in [7].

For the converses results, we have introduced a new technique for establishing outer bounds to the performance of the variable-length block codes, which can be used in both *message-wise* and *bit-wise* UEP problems.[31]

We were only interested in *bit-wise* UEP problem in this paper. We have analyzed single-message *message-wise* UEP problem, because it is closely related to *bit-wise* UEP problem and its analysis allowed us to introduce the ideas we use for *bit-wise* UEP, gradually. However, it seems using the technique employed in [2, Th. 9] on the achievability side and Lemma 5 on the converse side, one might be able to determine the achievable region of rate-exponent vectors for variable-length block codes in *message-wise* UEP problem. Such a work would allow us to determine the gains of feedback and variable-length decoding, because Csiszár [5] had already solved the problem for fixed-length block codes.

Arguably, the most important shortcoming of our *bit-wise* UEP result is that it only addresses the case when the number of groups of bits ℓ is a fixed integer. However, this has more to do with the formal definition of the problem we have chosen in Section III than our analysis and nonasymptotic results given in Sections IV and V, i.e., Lemmas 4 and 7.

Using the rate-exponent vectors for representing the performance of a reliable sequence with *bit-wise* UEP is apt only when the number of groups of bits are fixed or bounded. When the number of groups of bits ℓ in a reliable sequence diverge with increasing κ, i.e., when $\lim_{\kappa\to\infty}\ell_\kappa=\infty$, the rate-exponent vector formulation becomes fundamentally inapt. Consider, for example, a reliable sequence in which $|\mathcal{M}_i^{(\kappa)}|=\lceil e^{\mathbf{E}[\mathsf{T}^{(\kappa)}]\frac{\mathrm{R}}{\ell_\kappa}}\rceil$. The rate of this reliable sequence is R; yet, the rate of all of the submessages are zero. Thus, when ℓ_κ diverges, the rate vector does not have the same operational relevance or meaning it has when ℓ_κ is fixed or bounded. In order to characterize the change of error performance among submessages in the case when ℓ_κ

[31]We have not employed the bound in any hybrid problem but it seems that the result is abstract enough to be employed even in those problems with judicious choice of NALD's.

diverges, one needs to come up with an alternative formulation of the problem, in terms of cumulative rate of submessages.

Our nonasymptotic results are useful to some extend even when ℓ_κ diverges. Although infinite dimensional rate-exponent vectors fall short of representing all achievable performances, one can still use Lemma 4 of Section IV and Lemma 7 of Section V to characterize the set of achievable rate vector error exponent vector pairs.

1) As a result of Lemma 7, the necessary condition given in (19) is still a necessary condition for the achievability of rate-exponent vector.
2) Using Lemma 4, we see that the sufficient condition given in (19) is still a sufficient condition as long as the number of submessages in the reliable sequence satisfy $\limsup_{n\to\infty} \frac{\ell_n}{n/\ln n} = 0$.

Thus, for the case when $\ell \sim o(\frac{\mathbf{E}[\mathsf{T}]}{\ln \mathbf{E}[\mathsf{T}]})$, i.e., $\limsup_{\kappa\to\infty} \frac{\ell_\kappa}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]/\ln \mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]} = 0$, the condition given in (19) is still a necessary and sufficient condition for the achievability of a rate-exponent vector.

APPENDIX

A. Proof of Lemma 1

Proof: Note that $J(\mathrm{R})$ defined in (17) is also equal to

$$\begin{aligned} J(\mathrm{R}) &= \max_{\substack{0\le\alpha\le 1\\ \mathsf{x}_1,\mathsf{x}_2\in\mathcal{X}\\ \mu_1,\mu_2\in P(\mathcal{X})\\ \alpha,\mathsf{x}_1,\mathsf{x}_2,\mu_1,\mu_2,\mathrm{R}_1,\mathrm{R}_2:\ \mathrm{R}_1,\mathrm{R}_2\in[0,C]\\ \mathrm{I}(\mu_1,W)\ge\mathrm{R}_1\\ \mathrm{I}(\mu_2,W)\ge\mathrm{R}_2\\ \alpha\mathrm{R}_1+(1-\alpha)\mathrm{R}_2=\mathrm{R}}} \alpha\mathrm{D}\left(\bar{\mu}_1\|W_{\mathsf{x}_1}\right)+(1-\alpha)\mathrm{D}\left(\bar{\mu}_2\|W_{\mathsf{x}_2}\right) \\ &= \max_{\substack{0\le\alpha\le 1\\ \alpha,\mathrm{R}_1,\mathrm{R}_2:\ \mathrm{R}_1,\mathrm{R}_2\in[0,C]\\ \alpha\mathrm{R}_1+(1-\alpha)\mathrm{R}_2=\mathrm{R}}} \alpha j(\mathrm{R}_1)+(1-\alpha)j(\mathrm{R}_2) \end{aligned} \tag{63}$$

where $j(\mathrm{R})$ is given by

$$j(\mathrm{R}) \stackrel{\triangle}{=} \max_{\substack{\mathsf{x}\in\mathcal{X}\\ \alpha,\mathsf{x},\mu:\ \mu\in P(\mathcal{X})\\ \mathrm{I}(\mu,W)\ge\mathrm{R}}} \mathrm{D}\left(\bar{\mu}\|W_{\mathsf{x}}\right) \qquad \forall \mathrm{R}\in C. \tag{64}$$

Note that $j(\mathrm{R})$ is a bounded real-valued function of a real variable. Therefore, Carathéodory's Theorem implies that considering two point convex combinations suffices in order make $j(\mathrm{R})$ a concave function. In other words, for any k, we have

$$\begin{aligned} &\max_{\substack{0\le\alpha\le 1\\ \alpha,\mathrm{R}_1,\mathrm{R}_2:\ \mathrm{R}_1,\mathrm{R}_2\in[0,C]\\ \alpha\mathrm{R}_1+(1-\alpha)\mathrm{R}_2=\mathrm{R}}} \alpha j(\mathrm{R}_1)+(1-\alpha)j(\mathrm{R}_2) \\ &\qquad = \max_{\substack{0\le\alpha_i\le 1\ \forall i\\ 0\le\mathrm{R}_i\le C\ \forall i\\ \alpha_1,\ldots,\alpha_k,\\ \mathrm{R}_1,\ldots,\mathrm{R}_k:\ \sum_i\alpha_i=1\\ \sum_i\alpha_i\mathrm{R}_i=\mathrm{R}}} \sum_{i=1}^{k}\alpha_i j(\mathrm{R}_i). \end{aligned} \tag{65}$$

Then, the concavity of $J(\mathrm{R})$ follows from (63)–(65).

Evidently, if the constraint set in a maximization is curtailed, then resulting maximum value cannot increase. Hence, $J(\mathrm{R})$ function defined in (17) is a decreasing function of R.

As a result of the definition of D given in (9) and the convexity of Kullback–Leibler divergence, we have $D \ge J(0)$. On the other hand, $\mathrm{D}\left(\bar{\mu}\|W_{\mathsf{x}}\right) = D$ and $\mathrm{I}\left(\mu, W\right) \ge 0$ for $\mathsf{x} = \mathbf{r}$ and $\mu(\cdot) = \mathbb{1}_{\{\cdot=\mathbf{a}\}}$ where $\mathbf{a}$ and $\mathbf{r}$ are described in (10). Therefore, we have $j(0) \ge D$. Using the fact that $J(\mathrm{R}) \ge j(\mathrm{R})$, we conclude that $J(0) = j(0) = D$. ■

B. Proof of Lemma 2

Proof: We prove the lemma for a slightly more general setting and establish a result that will be easier to make use of in the proofs of other achievability results. Let $\mathcal{G}_\gamma[1]$, $\mathcal{G}_\gamma[\mathsf{m}]$, and $\mathcal{B}_\gamma[\mathsf{x}^n]$ be (66a)–(66c) shown at the bottom of the page. Note that $\mathcal{G}[1]$, $\mathcal{G}[\mathsf{m}]$, and $\mathcal{B}[\mathsf{x}^n]$ given, (21)–(23) are simply the $\mathcal{G}_\gamma[1]$, $\mathcal{G}_\gamma[\mathsf{m}]$, and $\mathcal{B}_\gamma[\mathsf{x}^n]$ for $\gamma = |\mathcal{X}||\mathcal{Y}|\sqrt{n\ln(1+n)}$.

For all $\mathsf{y}^n \notin \mathcal{G}_\gamma[1]$, we have

$$\begin{aligned} &n_\alpha \mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}\middle\| W_{\mathsf{x}_1}\right) + (n-n_\alpha)\mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}\middle\| W_{\mathsf{x}_2}\right) \\ &\quad = n_\alpha \mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}\middle\| \bar{\mu}_1\right) + (n-n_\alpha)\mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}\middle\| \bar{\mu}_2\right) \\ &\qquad + n_\alpha \sum_{\mathsf{y}} \mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}(\mathsf{y}) \ln \tfrac{\bar{\mu}_1(\mathsf{y})}{W_{\mathsf{x}_1}(\mathsf{y})} \\ &\qquad + (n-n_\alpha)\sum_{\mathsf{y}} \mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}(\mathsf{y}) \ln \tfrac{\bar{\mu}_2(\mathsf{y})}{W_{\mathsf{x}_2}(\mathsf{y})} \\ &\quad \stackrel{(a)}{\ge} n_\alpha \sum_{\mathsf{y}} \mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}(\mathsf{y}) \ln \tfrac{\bar{\mu}_1(\mathsf{y})}{W_{\mathsf{x}_1}(\mathsf{y})} \\ &\qquad + (n-n_\alpha)\sum_{\mathsf{y}} \mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}(\mathsf{y}) \ln \tfrac{\bar{\mu}_2(\mathsf{y})}{W_{\mathsf{x}_2}(\mathsf{y})} \\ &\quad \stackrel{(b)}{\ge} n_\alpha \mathrm{D}\left(\bar{\mu}_1\|W_{\mathsf{x}_1}\right) + (n-n_\alpha)\mathrm{D}\left(\bar{\mu}_2\|W_{\mathsf{x}_2}\right) + 2\gamma\ln\lambda. \end{aligned}$$

$$\mathcal{G}_\gamma[1] = \left\{\mathsf{y}^n : n_\alpha \Delta\left(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}, \bar{\mu}_1\right) + (n-n_\alpha)\Delta\left(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^{n}\}}, \bar{\mu}_2\right) \ge \gamma\right\} \tag{66a}$$

$$\mathcal{G}_\gamma[\mathsf{m}] = \mathcal{B}_\gamma[\mathsf{x}^n(\mathsf{m})] \bigcap \left(\cap_{\widetilde{\mathsf{m}}\neq\mathsf{m}} \overline{\mathcal{B}_\gamma[\mathsf{x}^n(\widetilde{\mathsf{m}})]}\right) \qquad \forall \mathsf{m}\in\{2,3,\ldots,|\mathcal{M}|\} \tag{66b}$$

$$\mathcal{B}_\gamma[\mathsf{x}^n] = \{\mathsf{y}^n : n_\alpha \Delta\left(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha},\mathsf{y}_1^{n_\alpha}\}}, \mu_1 W\right) + (n-n_\alpha)\Delta\left(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n},\mathsf{y}_{n_\alpha+1}^{n}\}}, \mu_2 W\right) < \gamma\} \tag{66c}$$

Inequality (a) follows from the nonnegativity of the Kullback–Leibler divergence. In order to see why (b) holds, first recall that $\min_{\mathsf{x},\mathsf{y}} W_{\mathsf{x}}(\mathsf{y}) = \lambda$. Hence, $|\ln \frac{\bar{\mu}_1(\mathsf{y})}{W_{\mathsf{x}_1}(\mathsf{y})}| \leq \ln \frac{1}{\lambda}$ and $|\ln \frac{\bar{\mu}_2(\mathsf{y})}{W_{\mathsf{x}_2}(\mathsf{y})}| \leq \ln \frac{1}{\lambda}$. Then, the inequality (b) follows from the definitions of total variation Δ and $\mathcal{G}_\gamma[1]$, given in (1) and (66a) and the fact that $\mathsf{y}^n \notin \mathcal{G}_\gamma[1]$.

Note that the conditional error probability of the first message is given by

$$\begin{aligned} P_{\mathbf{e}|1} &= \mathbf{P}\Big[\hat{\mathsf{M}} \neq 1 \Big| \mathsf{M} = 1\Big] \\ &= \sum_{\mathsf{y}^n \notin \mathcal{G}_\gamma[1]} \mathbf{P}[\mathsf{Y}^n = \mathsf{y}^n | \mathsf{M} = 1]. \end{aligned}$$

Recall that the codeword of the message $\mathsf{M} = 1$ is the concatenation of n_α x_1's and $(n - n_\alpha)$ x_2's where $n_\alpha = \lfloor n\alpha \rfloor$. Hence, the probability of all y^n's whose empirical distribution in first n_α times instances is $\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}$ and whose empirical distribution in $[(n_\alpha + 1), n]$ is $\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^n\}}$ is upper bounded by $e^{-n_\alpha \mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}} \middle\| W_{\mathsf{x}_1}\right) - (n-n_\alpha)\mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^n\}} \middle\| W_{\mathsf{x}_2}\right)}$. Furthermore, there are less than $(n_\alpha + 1)^{|\mathcal{Y}|}$ distinct empirical distributions in the first phase and there are less than $(n - n_\alpha + 1)^{|\mathcal{Y}|}$ distinct empirical distributions in the second phase. Thus, we have the equation shown at the bottom of the page where $\varepsilon_2(\gamma, n) = \frac{-2\gamma \ln \lambda + D + 2|\mathcal{Y}| \ln(n+1)}{n}$.

The codewords and the decoding regions of the remaining messages are specified using a random coding argument together with an empirical typicality decoder. Consider an ensemble of codes in which first n_α entries of all the codewords are independent and identically distributed (i.i.d.) with input distribution μ_1 and the rest of the entries are i.i.d. with the input distribution μ_2.

For any message m other than the first one, i.e., $\mathsf{m} \neq 1$, the decoding region is $\mathcal{G}_\gamma[\mathsf{m}]$ given in (66b). In other words, for any message m, other than the first one, the decoding region is the set of output sequences for which $(\mathsf{x}^n(\mathsf{m}), \mathsf{y}^n)$ is typical with $(\alpha, \mu_1 W, \mu_2 W)$, i.e., $\mathsf{y}^n \in \mathcal{B}_\gamma[\mathsf{x}^n(\mathsf{m})]$, and $(\mathsf{x}^n(\widetilde{\mathsf{m}}), \mathsf{y}^n)$ is not typical with $(\alpha, \mu_1 W, \mu_2 W)$, i.e., $\mathsf{y}^n \in \mathcal{B}_\gamma[\mathsf{x}^n(\widetilde{\mathsf{m}})]$, for any $\widetilde{\mathsf{m}} \neq \mathsf{m}$.

Since the decoding regions of different messages are disjoint, the previously described code does not decode to more than one message. Disjointness of decoding regions of messages $2, 3, \ldots, |\mathcal{M}|$ follows from the definitions of $\mathcal{G}_\gamma[2], \mathcal{G}_\gamma[3], \ldots, \mathcal{G}_\gamma[|\mathcal{M}|]$, given in (66b). In order to see why $\mathcal{G}_\gamma[1] \cap (\cup_{\mathsf{m}\neq 1} \mathcal{G}_\gamma[\mathsf{m}]) = \emptyset$ holds, note that for any pair probability of distributions, the total variation between them is lower bounded by the total variation between their marginals. In particular

$$\begin{aligned} \Delta\Big(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}(\mathsf{m}), \mathsf{y}_1^{n_\alpha}\}}, \mu_1 W\Big) &\geq \Delta\Big(\mathrm{Q}_{\{\mathsf{y}_1^{n_\alpha}\}}, \bar{\mu}_1\Big) \\ \Delta\Big(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^n(\mathsf{m}), \mathsf{y}_{n_\alpha+1}^n\}}, \mu_2 W\Big) &\geq \Delta\Big(\mathrm{Q}_{\{\mathsf{y}_{n_\alpha+1}^n\}}, \bar{\mu}_2\Big). \end{aligned}$$

Then, as results of definitions of $\mathcal{G}_\gamma[1]$, $\mathcal{B}_\gamma[\mathsf{x}^n]$, and $\mathcal{G}_\gamma[\mathsf{m}]$ for $\mathsf{m} \neq 1$ given in (66a), (66c), and (66b), we have

$$\mathcal{G}_\gamma[1] \cap \mathcal{G}_\gamma[\mathsf{m}] = \emptyset \qquad \mathsf{m} = 2, 3, \ldots |\mathcal{M}|.$$

Then, for $\mathsf{m} \in \{2, 3, \ldots, |\mathcal{M}|\}$, the average of the conditional error probability of mth message over the ensemble is upper bounded as

$$\begin{aligned} \mathbf{E}\big[P_{\mathbf{e}|\mathsf{m}}\big] \leq\ & \mathbf{P}[\mathsf{Y}^n \notin \mathcal{B}_\gamma[\mathsf{X}^n(\mathsf{m})] | \mathsf{M} = \mathsf{m}] \\ & + \sum_{\widetilde{\mathsf{m}} \neq \mathsf{m}} \mathbf{P}[\mathsf{Y}^n \in \mathcal{B}_\gamma[\mathsf{X}^n(\widetilde{\mathsf{m}})] | \mathsf{M} = \mathsf{m}]. \quad (67) \end{aligned}$$

Let us start with bounding $\mathbf{P}[\mathsf{Y}^n \notin \mathcal{B}_\gamma[\mathsf{X}^n(\mathsf{m})] | \mathsf{M} = \mathsf{m}]$. Let $\mathsf{S}_1(\mathsf{x}, \mathsf{y})$ and $\mathsf{S}_2(\mathsf{x}, \mathsf{y})$ be

$$\mathsf{S}_1(\mathsf{x}, \mathsf{y}) \stackrel{\triangle}{=} n_\alpha |\mathrm{Q}_{\{\mathsf{X}_1^{n_\alpha}(\mathsf{m}), \mathsf{Y}_1^{n_\alpha}\}}(\mathsf{x}, \mathsf{y}) - \mu_1(\mathsf{x}) W_{\mathsf{x}}(\mathsf{y})| \quad (68a)$$

$$\mathsf{S}_2(\mathsf{x}, \mathsf{y}) \stackrel{\triangle}{=} (n - n_\alpha) |\mathrm{Q}_{\{\mathsf{X}_{n_\alpha+1}^n(\mathsf{m}), \mathsf{Y}_{n_\alpha+1}^n\}}(\mathsf{x}, \mathsf{y}) - \mu_2(\mathsf{x}) W_{\mathsf{x}}(\mathsf{y})|. \quad (68b)$$

As a result of the definition of total variation distance given in (1) and aforementioned definitions, we have

$$\begin{aligned} n_\alpha \Delta\Big(\mathrm{Q}_{\{\mathsf{X}_1^{n_\alpha}(\mathsf{m}), \mathsf{Y}_1^{n_\alpha}\}}, \mu_1 W\Big) + (n - n_\alpha) \Delta\Big(\mathrm{Q}_{\{\mathsf{X}_{n_1+1}^n(\mathsf{m}), \mathsf{Y}_{n_1+1}^n\}}, \mu_2 W\Big) \\ = \tfrac{1}{2} \sum\nolimits_{\mathsf{x},\mathsf{y}} [\mathsf{S}_1(\mathsf{x}, \mathsf{y}) + \mathsf{S}_2(\mathsf{x}, \mathsf{y})]. \end{aligned}$$

Thus, the definition of $\mathcal{B}_\gamma[\mathsf{x}^n(\mathsf{m})]$ given in (66c) implies that

$$\begin{aligned} &\mathbf{P}[\mathsf{Y} \notin \mathcal{B}_\gamma[\mathsf{X}^n(\mathsf{m})] | \mathsf{M} = \mathsf{m}] \\ &\qquad = \mathbf{P}\left[\sum_{\mathsf{x},\mathsf{y}} [\mathsf{S}_1(\mathsf{x}, \mathsf{y}) + \mathsf{S}_2(\mathsf{x}, \mathsf{y})] \geq 2\gamma \middle| \mathsf{M} = \mathsf{m}\right]. \quad (69) \end{aligned}$$

If for all $\mathsf{x} \in \mathcal{X}$, $\mathsf{y} \in \mathcal{Y}$, and $j \in \{1, 2\}$, $\mathsf{S}_j(\mathsf{x}, \mathsf{y}) \leq \gamma |\mathcal{X}|^{-1} |\mathcal{Y}|^{-1}$, then $\sum_{\mathsf{x},\mathsf{y}} [\mathsf{S}_1(\mathsf{x}, \mathsf{y}) + \mathsf{S}_2(\mathsf{x}, \mathsf{y})] \leq 2\gamma$. Thus, if $\mathsf{Y} \notin \mathcal{B}_\gamma[\mathsf{X}^n(\mathsf{m})]$, then for at least one $(\mathsf{x}, \mathsf{y}, j)$ triple $\mathsf{S}_j(\mathsf{x}, \mathsf{y}) \geq \gamma |\mathcal{X}|^{-1} |\mathcal{Y}|^{-1}$. Using the union bound, we get

$$\begin{aligned} &\mathbf{P}\Big[\sum\nolimits_{\mathsf{x},\mathsf{y}} [\mathsf{S}_1(\mathsf{x}, \mathsf{y}) + \mathsf{S}_2(\mathsf{x}, \mathsf{y})] \geq 2\gamma \Big| \mathsf{M} = \mathsf{m}\Big] \\ &\qquad \leq \sum\nolimits_{\mathsf{x},\mathsf{y},j} \mathbf{P}\Big[\mathsf{S}_j(\mathsf{x}, \mathsf{y}) \geq \tfrac{\gamma}{|\mathcal{X}||\mathcal{Y}|} \Big| \mathsf{M} = \mathsf{m}\Big]. \quad (70) \end{aligned}$$

For bounding $\mathbf{P}\Big[\mathsf{S}_j(\mathsf{x}, \mathsf{y}) \geq \frac{\gamma}{|\mathcal{X}||\mathcal{Y}|} \Big| \mathsf{M} = \mathsf{m}\Big]$, we can simply use Chebyshev's inequality; however, in order to get better error terms, we use a standard concentration result about the sums of bounded random variables, [4, Th. 5.3].

Lemma 8: Let $\mathsf{Z}_1, \mathsf{Z}_2, \ldots, \mathsf{Z}_k$ be independent random variables satisfying $|\mathsf{Z}_i - \mathbf{E}[\mathsf{Z}_i]| \leq c_i$ for all $1 \leq i \leq k$. Then

$$\mathbf{P}\left[\left|\sum\nolimits_{i=1}^k (\mathsf{Z}_i - \mathbf{E}[\mathsf{Z}_i])\right| > \gamma\right] \leq 2e^{-\frac{\gamma^2}{2\sum_{i=1}^k c_i}}.$$

$$\begin{aligned} P_{\mathbf{e}|1} &\leq (n_\alpha + 1)^{|\mathcal{Y}|} (n - n_\alpha + 1)^{|\mathcal{Y}|} e^{-n_\alpha \mathrm{D}(\bar{\mu}_1 \| W_{\mathsf{x}_1}) + (n-n_\alpha)\mathrm{D}(\bar{\mu}_2 \| W_{\mathsf{x}_2}) - 2\gamma \ln \lambda} \\ &\leq e^{-n(\alpha \mathrm{D}(\bar{\mu}_1 \| W_{\mathsf{x}_1}) + (1-\alpha)\mathrm{D}(\bar{\mu}_2 \| W_{\mathsf{x}_2}) - \varepsilon_2(\gamma, n))} \end{aligned}$$

For all $\mu_1 \in P(\mathcal{X})$, $\mathsf{x} \in \mathcal{X}$, and $\mathsf{y} \in \mathcal{Y}$, we have $c_i = 1$ for all $i = 1, 2, \ldots, n_\alpha$; thus

$$\mathbf{P}\Big[\mathsf{S}_1(\mathsf{x},\mathsf{y}) \geq \tfrac{\gamma}{|\mathcal{X}||\mathcal{Y}|}\Big|\, \mathsf{M}=\mathsf{m}\Big] \leq 2e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n_\alpha}} \leq 2e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}. \quad (71)$$

Similarly

$$\mathbf{P}\Big[\mathsf{S}_2(\mathsf{x},\mathsf{y}) \geq \tfrac{\gamma}{|\mathcal{X}||\mathcal{Y}|}\Big|\, \mathsf{M}=\mathsf{m}\Big] \leq 2e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}. \quad (72)$$

Using (69)–(72), we get

$$\mathbf{P}[\mathsf{Y} \notin \mathcal{B}_\gamma[\mathsf{X}^n(\mathsf{m})]|\, \mathsf{M}=\mathsf{m}] \leq 4|\mathcal{X}|\mathcal{Y}|e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}. \quad (73)$$

Now, we focus on $\mathbf{P}[\mathsf{Y}^n \in \mathcal{B}_\gamma[\mathsf{X}^n(\widetilde{\mathsf{m}})]|\, \mathsf{M}=\mathsf{m}]$ terms. Note that all y^n in $\mathcal{B}_\gamma[\mathsf{x}^n(\widetilde{\mathsf{m}})]$ satisfy

$$n_\alpha \Delta\Big(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}(\widetilde{\mathsf{m}}),\mathsf{y}_1^{n_\alpha}\}}, \mu_1 W\Big) + (n-n_1)\Delta\Big(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n}(\widetilde{\mathsf{m}}),\mathsf{y}_{n_\alpha+1}^{n}\}}, \mu_2 W\Big) \leq \gamma. \quad (74)$$

On the other hand, when $\mathsf{M}=\mathsf{m}$, $\mathsf{X}^n(\widetilde{\mathsf{m}})$ and Y^n are independent and their distribution is given by (75) shown at the bottom of the page. Furthermore, the number of $(\mathsf{x}_1^{n_\alpha}(\widetilde{\mathsf{m}}), \mathsf{y}_1^{n_\alpha})$ sequences with an empirical distribution $\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}(\widetilde{\mathsf{m}}),\mathsf{y}_1^{n_\alpha}\}}$ is upper bounded as $e^{n_\alpha \mathrm{H}\left(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}(\widetilde{\mathsf{m}}),\mathsf{y}_1^{n_\alpha}\}}\right)}$. In addition, there are at most $(n_\alpha+1)^{|\mathcal{X}||\mathcal{Y}|}$ different empirical distributions. Using these two bounds and their counterparts for $(\mathsf{x}_{n_\alpha+1}^{n}(\widetilde{\mathsf{m}}), \mathsf{y}_{n_\alpha+1}^{n})$ together with (74) and (75) we get (76) shown at the bottom of the page. Hence, if (see the third equation at the bottom of the page) then

$$\sum\nolimits_{\widetilde{\mathsf{m}}\neq\mathsf{m}} \mathbf{P}[\mathsf{Y}^n \in \mathcal{B}_\gamma[\mathsf{X}^n(\widetilde{\mathsf{m}})]|\, \mathsf{M}=\mathsf{m}] \leq 4|\mathcal{X}|\mathcal{Y}|e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}. \quad (77)$$

Thus, the average P_{e} over the ensemble can be bounded using (67), (70), and (77) as

$$\mathbf{E}[P_{\mathrm{e}}] \leq 8|\mathcal{X}|\,|\mathcal{Y}|e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}.$$

But if the ensemble average of the error probability is upper bounded like this, there is at least one code that has this low error probability. Furthermore, half of its messages have conditional error probabilities less than twice this average. Thus, for any block length n, time-sharing constant $\alpha \in [0,1]$, input letters $\mathsf{x}_1, \mathsf{x}_2 \in \mathcal{X}$, input distributions $\mu_1, \mu_2 \in P(\mathcal{X})$, there exists a length n code such that

$$|\mathcal{M}\setminus\{1\}| \geq e^{n(\alpha \mathrm{I}(\mu_1,W)+(1-\alpha)\mathrm{I}(\mu_2,W)-\varepsilon_1(\gamma,n))} \quad (78a)$$

$$P_{\mathrm{e}|1} \leq e^{-n(\alpha \mathrm{D}(\bar{\mu}_1\|W_{\mathsf{x}_1})+(1-\alpha)\mathrm{D}(\bar{\mu}_2\|W_{\mathsf{x}_2})-\varepsilon_2(\gamma,n))} \quad (78b)$$

$$P_{\mathrm{e}|\mathsf{m}} \leq \varepsilon_3(\gamma,n) \qquad \mathsf{m}=2,3,\ldots,|\mathcal{M}| \quad (78c)$$

where

$$\varepsilon_1(\gamma,n) = \tfrac{C-\ln(2|\mathcal{X}||\mathcal{Y}|)+2|\mathcal{X}||\mathcal{Y}|\ln(n+1)-2\gamma\ln\lambda)}{n} + \tfrac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n^2} \quad (79a)$$

$$\varepsilon_2(\gamma,n) = \tfrac{D+2|\mathcal{Y}|\ln(n+1)-2\gamma\ln\lambda}{n} \quad (79b)$$

$$\varepsilon_3(\gamma,n) = 16|\mathcal{X}|\,|\mathcal{Y}|e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}. \quad (79c)$$

Lemma 2 follows from (78) and the fact that

$$\varepsilon_i(\gamma,n)\big|_{\gamma=|\mathcal{X}||\mathcal{Y}|\sqrt{n\ln(1+n)}} \leq \tfrac{9|\mathcal{X}||\mathcal{Y}|(1-\ln\lambda)\sqrt{\ln(1+n)}}{\sqrt{n}} \qquad i=1,2,3 \quad (80)$$

for $\varepsilon_1(\gamma,n)$, $\varepsilon_2(\gamma,n)$, and $\varepsilon_3(\gamma,n)$ given in (79). ■

$$\begin{aligned}\mathbf{P}[(\mathsf{X}^n(\widetilde{\mathsf{m}}),\mathsf{Y}^n)=(\mathsf{x}^n(\widetilde{\mathsf{m}}),\mathsf{y}^n)|\,\mathsf{M}=\mathsf{m}] &= \prod\nolimits_{i=1}^{n_\alpha}\mu_1(\mathsf{x}_i(\widetilde{\mathsf{m}}))\bar{\mu}_1(\mathsf{y}_i)\prod\nolimits_{j=n_\alpha+1}^{n}\mu_2(\mathsf{x}_j(\widetilde{\mathsf{m}}))\bar{\mu}_2(\mathsf{y}_j)\\ &= e^{-n_\alpha \mathrm{D}\left(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}\widetilde{(\mathsf{m})},\mathsf{y}_1^{n_\alpha}\}}\middle\|\mu_1\bar{\mu}_1\right)} e^{-n_\alpha \mathrm{H}\left(\mathrm{Q}_{\{\mathsf{x}_1^{n_\alpha}\widetilde{(\mathsf{m})},\mathsf{y}_1^{n_\alpha}\}}\right)}\\ &\quad e^{-(n-n_\alpha)\mathrm{D}\left(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n}\widetilde{(\mathsf{m})},\mathsf{y}_{n_\alpha+1}^{n}\}}\middle\|\mu_2\bar{\mu}_2\right)} e^{-(n-n_\alpha)\mathrm{H}\left(\mathrm{Q}_{\{\mathsf{x}_{n_\alpha+1}^{n}\widetilde{(\mathsf{m})},\mathsf{y}_{n_\alpha+1}^{n}\}}\right)}\end{aligned} \quad (75)$$

$$\begin{aligned}\mathbf{P}[\mathsf{Y}^n \in \mathcal{B}_\gamma[\mathsf{X}^n(\widetilde{\mathsf{m}})]|\,\mathsf{M}=\mathsf{m}] &\leq (n_\alpha+1)^{|\mathcal{X}||\mathcal{Y}|}(n-n_\alpha+1)^{|\mathcal{X}||\mathcal{Y}|}e^{-n_\alpha \mathrm{D}(\mu_1 W\|\mu_1\bar{\mu}_1)-(n-n_\alpha)\mathrm{D}(\mu_2 W\|\mu_2\bar{\mu}_2)-2\gamma\ln\lambda}\\ &= (n_\alpha+1)^{|\mathcal{X}||\mathcal{Y}|}(n-n_\alpha+1)^{|\mathcal{X}||\mathcal{Y}|}e^{-n_\alpha \mathrm{I}(\mu_1,W)-(n-n_\alpha)\mathrm{I}(\mu_2,W)-2\gamma\ln\lambda}\\ &\leq e^{-n(\alpha\mathrm{I}(\mu_1,W)+(1-\alpha)\mathrm{I}(\mu_2,W))}e^{C+2|\mathcal{X}||\mathcal{Y}|\ln(n+1)-2\gamma\ln\lambda}\end{aligned} \quad (76)$$

$$|\mathcal{M}\setminus\{1\}| = 4|\mathcal{X}|\,|\mathcal{Y}|e^{-\frac{\gamma^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2 n}}e^{n(\alpha\mathrm{I}(\mu_1,W)+(1-\alpha)\mathrm{I}(\mu_2,W))}e^{-C-2|\mathcal{X}||\mathcal{Y}|\ln(n+1)+2\gamma\ln\lambda}$$

C. Proof of Lemma 3

Proof: Let n_1 be $n_1 = \lceil (1-\frac{\mathrm{E}}{D})n \rceil$. Recall that we have assumed $\mathrm{E} \le (1-\frac{\mathrm{R}}{C})D$; then, we have $\frac{\mathrm{R}}{C} \le 1-\frac{\mathrm{E}}{D}$. Consequently, $\frac{\mathrm{R}}{C} \le \frac{n_1}{n}$ and $\frac{n}{n_1}\mathrm{R} \le C$. On the other hand, as a result of (78) and the definition of $J(\cdot)$ given in (17), for any positive integer n_1, positive real number γ_1, rate $\widetilde{\mathrm{R}} \le C$, there exists a length n_1 code such that

$$|\mathcal{M}| - 1 \ge e^{n_1[\widetilde{\mathrm{R}} - \varepsilon_1(\gamma_1,n_1)]} \tag{81a}$$

$$P_{\mathbf{e}|1} \le e^{-n_1[J\left(\widetilde{\mathrm{R}}\right) - \varepsilon_2(\gamma_1,n_1)]} \tag{81b}$$

$$P_{\mathbf{e}|\mathsf{m}} \le \varepsilon_3(\gamma_1,n_1) \qquad \mathsf{m} = 2,3,\ldots,|\mathcal{M}| \tag{81c}$$

where $\varepsilon_1(\gamma_1,n_1)$, $\varepsilon_2(\gamma_1,n_1)$, and $\varepsilon_3(\gamma_1,n_1)$ are given in (79).

We use such a code in the first phase with $\widetilde{\mathrm{R}} = \frac{n}{n_1}\mathrm{R}$ and call its decoded message ${}_{\mathrm{t}}\widehat{\mathsf{M}}$, the tentative decision. Then, as a result of (81) and the fact that[32] $n_1 J\left(\frac{n}{n_1}\mathrm{R}\right) \ge n(1-\frac{\mathrm{E}}{D})J\left(\frac{\mathrm{R}}{1-\mathrm{E}/D}\right)$, we get

$$|\mathcal{M}| - 1 \ge e^{n\mathrm{R} - n_1\varepsilon_1(\gamma_1,n_1)} \tag{82a}$$

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m} \,\middle|\, \mathsf{M}=1\right] \le e^{-n(1-\frac{\mathrm{E}}{D})J\left(\frac{\mathrm{R}}{1-\mathrm{E}/D}\right) + n_1\varepsilon_2(\gamma_1,n_1)} \tag{82b}$$

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m} \,\middle|\, \mathsf{M}=\mathsf{m}\right] \le \varepsilon_3(\gamma_1,n_1) \qquad \mathsf{m} = 2,3,\ldots,|\mathcal{M}|. \tag{82c}$$

The transmitter knows what the tentative decision is and determines the channel inputs in the last $(n-n_1)$ time instances depending on its correctness. If ${}_{\mathrm{t}}\widehat{\mathsf{M}} = \mathsf{M}$, the channel inputs in the last $(n-n_1)$ time instances are all $\mathbf{a}$; if ${}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{M}$, the channel inputs in the last $(n-n_1)$ time instances are all $\mathbf{r}$.

After observing Y^n, receiver checks whether the empirical distribution of the channel output in the last $(n-n_1)$ time units is typical with $W_{\mathbf{a}}$ if it is then $\widehat{\mathsf{M}} = {}_{\mathrm{t}}\widehat{\mathsf{M}}$ otherwise $\widehat{\mathsf{M}} = \mathbf{x}$. Hence, the decoding region for erasures is given by

$$\mathcal{G}_\gamma[\mathbf{x}] = \{\mathsf{y}^n : (n-n_1)\Delta\left(\mathrm{Q}_{\{\mathsf{y}_{n_1+1}^n\}}, W_{\mathbf{a}}\right) \ge \gamma_2\}.$$

Let us start with bounding $\mathbf{P}\left[\widehat{\mathsf{M}} = \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} = \mathsf{m}, \mathsf{M} = \mathsf{m}\right]$, i.e., the probability of erasure for correct tentative decision. First note that

$$(n-n_1)\Delta\left(\mathrm{Q}_{\{\mathsf{Y}_{n_1+1}^n\}}, W_{\mathbf{a}}\right) = \tfrac{1}{2}\sum\nolimits_{\mathsf{y}} \mathsf{S}(\mathsf{y})$$

where $\mathsf{S}(\mathsf{y}) = (n-n_1)|\mathrm{Q}_{\{\mathsf{Y}_{n_1+1}^n\}}(\mathsf{y}) - W_{\mathbf{a}}(\mathsf{y})|$. Then, following an analysis similar to that one presented between (69) and (73), we get

$$\mathbf{P}\left[\widehat{\mathsf{M}} = \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} = \mathsf{m}, \mathsf{M} = \mathsf{m}\right] \le 2|\mathcal{Y}| e^{-\frac{\gamma_2^2}{2|\mathcal{Y}|^2(n-n_1)}} = \varepsilon_3(\gamma_2, n-n_1) \qquad \forall \mathsf{m} \in \mathcal{M}. \tag{83}$$

[32]Recall that $n_1 \ge (1-\mathrm{E}/D)n$ and $J(\cdot)$ is a nonincreasing and positive function.

In order to bound the probability of nonerasure decoding when tentative decision is incorrect, note that

$$\begin{aligned}\mathbf{P}&\left[\mathsf{Y}_{n_1+1}^n = \mathsf{y}_{n_1+1}^n \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m}, \mathsf{M} = \mathsf{m}\right] \\ &= \prod_{j=n_1+1}^{n} W_{\mathbf{r}}(\mathsf{y}_j) \\ &= e^{-(n-n_1)\mathrm{D}\left(\mathrm{Q}_{\{\mathsf{y}_{n_1+1}^n\}} \middle\| \bar{\mu}_2\right)} e^{-(n-n_1)\mathrm{H}\left(\mathrm{Q}_{\{\mathsf{y}_{n_1+1}^n\}}\right)}.\end{aligned}$$

Then, following an analysis similar to the one between (75) and (76), we get

$$\begin{aligned}\mathbf{P}&\left[\widehat{\mathsf{M}} \ne \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m}, \mathsf{M} = \mathsf{m}\right] \\ &\le \min\{(n-n_1+1)^{|\mathcal{Y}|} e^{-(n-n_1)D - 2\gamma_2 \ln\lambda}, 1\} \\ &\le \min\{e^{-n\mathrm{E} + |\mathcal{Y}|\ln(n-n_1) + D - 2\gamma_2\ln\lambda}, 1\} \\ &\le \min\{e^{-n\mathrm{E} + (n-n_1)\varepsilon_2(\gamma_2, n-n_1)}, 1\} \quad \forall \mathsf{m} \in \mathcal{M}.\end{aligned} \tag{84}$$

Furthermore, the conditional error and erasure probabilities can be bounded in terms of $\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m} \,\middle|\, \mathsf{M} = \mathsf{m}\right]$, $\mathbf{P}\left[\widehat{\mathsf{M}} \ne \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m}, \mathsf{M} = \mathsf{m}\right]$, and $\mathbf{P}\left[\widehat{\mathsf{M}} = \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} = \mathsf{m}, \mathsf{M} = \mathsf{m}\right]$ as follows:

$$P_{\mathbf{e}|\mathsf{m}} = \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m} \,\middle|\, \mathsf{M} = \mathsf{m}\right] \mathbf{P}\left[\widehat{\mathsf{M}} \ne \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m}, \mathsf{M} = \mathsf{m}\right] \qquad \forall \mathsf{m} \in \mathcal{M} \tag{85a}$$

$$P_{\mathbf{x}|\mathsf{m}} \le \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}} \ne \mathsf{m} \,\middle|\, \mathsf{M} = \mathsf{m}\right] + \mathbf{P}\left[\widehat{\mathsf{M}} = \mathbf{x} \,\middle|\, {}_{\mathrm{t}}\widehat{\mathsf{M}} = \mathsf{m}, \mathsf{M} = \mathsf{m}\right] \qquad \forall \mathsf{m} \in \mathcal{M}. \tag{85b}$$

Using (82)–(85), we get

$$|\mathcal{M}| - 1 \ge e^{n\mathrm{R} - n_1\varepsilon_1(\gamma_1,n_1)} \tag{86a}$$

$$P_{\mathbf{e}|1} \le e^{-n(1-\frac{\mathrm{E}}{D})J\left(\frac{\mathrm{R}}{1-\mathrm{E}/D}\right) + n_1\varepsilon_2(\gamma_1,n_1)} \min\{e^{-n\mathrm{E} + n_2\varepsilon_2(\gamma_2,n_2)}, 1\} \tag{86b}$$

$$P_{\mathbf{x}|1} \le e^{-n(1-\frac{\mathrm{E}}{D})J\left(\frac{\mathrm{R}}{1-\mathrm{E}/D}\right) + n_1\varepsilon_2(\gamma_1,n_1)} + \varepsilon_3(\gamma_2,n_2) \tag{86c}$$

$$P_{\mathbf{e}|\mathsf{m}} \le \varepsilon_3(\gamma,n_1) \min\{e^{-n\mathrm{E} + |\mathcal{Y}|\ln(n+1) + n_2\varepsilon_2(\gamma_2,n_2)}, 1\} \qquad \mathsf{m} \ne 1 \tag{86d}$$

$$P_{\mathbf{x}|\mathsf{m}} \le \varepsilon_3(\gamma_1,n_1) + \varepsilon_3(\gamma_2,n_2) \quad \mathsf{m} \ne 1 \tag{86e}$$

where $n_2 = n - n_1$.

We set $\gamma_j = |\mathcal{X}||\mathcal{Y}|\sqrt{5n_j\ln(1+n)}$ for $j = 1, 2$ and obtain

$$n_j\varepsilon_1(\gamma_j,n_j) \le 2|\mathcal{X}||\mathcal{Y}|(\ln(n+1) - \sqrt{5n\ln(1+n)}\ln\lambda) + (5/2)\ln(1+n) \tag{87a}$$

$$n_j\varepsilon_2(\gamma_j,n_j) \le 2|\mathcal{X}||\mathcal{Y}|(\ln(n+1) - \sqrt{5n\ln(1+n)}\ln\lambda) + D \tag{87b}$$

$$\varepsilon_3(\gamma_j,n_j) \le 16|\mathcal{X}||\mathcal{Y}|/(1+n)^{5/2}. \tag{87c}$$

Lemma 3 follows from the identities $|\mathcal{X}| \ge 2$, $|\mathcal{Y}| \ge 2$, $D \le \ln(\frac{1}{\lambda})$, $n \ge 1$ and (86) and (87). ■

D. Proof of Lemma 4

Proof: Note that given the encoding scheme summarized in (28) and the decoding rule given in (29), if $\widehat{\mathsf{M}} = \mathbf{x}$, then there is a $i \leq \ell+1$ such that ${}_{\mathrm{t}}\widehat{\mathsf{M}}_j = {}_{\mathrm{t}}\mathsf{M}_j$ for all $j < i$ and ${}_{\mathrm{t}}\widehat{\mathsf{M}}_i \neq {}_{\mathrm{t}}\mathsf{M}_i$. Thus, the conditional erasure probability $P_{\mathbf{x}|\mathsf{m}}$ is upper bounded as shown in (88) at the bottom of the page. Similarly, if $\widehat{\mathsf{M}} \neq \mathbf{x}$ and $\widehat{\mathsf{M}}^i \neq \mathsf{M}^i$, then for all $j > i$, ${}_{\mathrm{t}}\mathsf{M}_j = 1$ and ${}_{\mathrm{t}}\widehat{\mathsf{M}}_j \neq 1$; furthermore, there is a $k \leq i$ such that ${}_{\mathrm{t}}\widehat{\mathsf{M}}_j = {}_{\mathrm{t}}\mathsf{M}_j$ for all $j < k$ and ${}_{\mathrm{t}}\widehat{\mathsf{M}}_k \neq {}_{\mathrm{t}}\mathsf{M}_k$. Hence, one can bound $P_{\mathbf{e}|\mathsf{m}}(i)$ as shown in (89) at the bottom of the page.

In the first ℓ phases, we use $n_i = \lfloor \eta_i n \rfloor$ long codes with rate $\frac{\mathrm{R}_i}{\eta_i}$ with the performance given in (81). Thus, for $1 \leq i \leq \ell$, we have

$$|{}_{\mathrm{t}}\mathcal{M}_i| \geq 1 + e^{n\mathrm{R}_i - C - n_i\varepsilon_1(\gamma_i,n_i)} \tag{90a}$$

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_i \neq 1 \middle| {}_{\mathrm{t}}\mathsf{M}_i = 1\right] \leq e^{-n\eta_i J\left(\frac{\mathrm{R}_i}{\eta_i}\right) + D - n_i\varepsilon_2(\gamma_i,n_i)} \tag{90b}$$

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_i \neq {}_{\mathrm{t}}\mathsf{M}_i \middle| {}_{\mathrm{t}}\mathsf{M}_i = 1 + \mathsf{m}_i\right] \leq \varepsilon_{3(\gamma_i,n_i)} \quad \mathsf{m}_i = 1,2,3,\ldots,(|{}_{\mathrm{t}}\mathcal{M}_i| - 1) \tag{90c}$$

where $\varepsilon_{1(\gamma_i,n_i)}$, $\varepsilon_{2(\gamma_i,n_i)}$, $\varepsilon_{3(\gamma_i,n_i)}$ are given in (79).

In order to derive bounds corresponding to the ones given in (90) for the last phase, let us give the decoding regions for 1 and 2 for the length $n_{\ell+1}$ code employed between $(n+1-n_{\ell+1})$ and n

$$\mathcal{G}_\gamma[1] = \{\mathsf{y}_{n+1-n_{\ell+1}}^n : n_{\ell+1}\Delta\Big(\mathrm{Q}_{\{\mathsf{y}_{n+1-n_{\ell+1}}^n\}}, W_{\mathbf{a}}\Big) \geq \gamma_{\ell+1}\}$$
$$\mathcal{G}_\gamma[2] = \{\mathsf{y}_{n+1-n_{\ell+1}}^n : n_{\ell+1}\Delta\Big(\mathrm{Q}_{\{\mathsf{y}_{n+1-n_{\ell+1}}^n\}}, W_{\mathbf{a}}\Big) < \gamma_{\ell+1}\}.$$

Following an analysis similar to the one leading to (83) and (84), we get

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_{\ell+1} \neq 1 \middle| {}_{\mathrm{t}}\mathsf{M}_{\ell+1} = 1\right] \leq e^{-n_{\ell+1}D + n_{\ell+1}\varepsilon_2(\gamma_{\ell+1},n_{\ell+1})} \tag{91a}$$

$$\mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_{\ell+1} \neq 2 \middle| {}_{\mathrm{t}}\mathsf{M}_{\ell+1} = 2\right] \leq \varepsilon_{3(\gamma_{\ell+1},n_{\ell+1})}. \tag{91b}$$

Using (88)–(91), we obtain, see obtain (92a)–(92d) shown at the bottom of the page. If we set $\gamma_i = |\mathcal{X}||\mathcal{Y}|\sqrt{4n_i\ln(1+n)}$ for $i = 1,2,\ldots,(\ell+1)$ for $\varepsilon_{1(\gamma_i,n_i)}$, $\varepsilon_{2(\gamma_i,n_i)}$, and $\varepsilon_{3(\gamma_i,n_i)}$ given in (79), we have

$$\begin{aligned} n_i\varepsilon_{1(\gamma_i,n_i)} &\leq 2\ln(1+n) + 2|\mathcal{X}||\mathcal{Y}|\ln(n_i+1) \\ &\quad - 2|\mathcal{X}||\mathcal{Y}|\sqrt{4n_i\ln(1+n)}\ln\lambda \\ n_i\varepsilon_{2(\gamma_i,n_i)} &\leq D + 2|\mathcal{X}||\mathcal{Y}|\ln(n_i+1) \\ &\quad - 2|\mathcal{X}||\mathcal{Y}|\sqrt{4n_i\ln(1+n)}\ln\lambda \\ \varepsilon_{3(\gamma_i,n_i)} &\leq 16|\mathcal{X}||\mathcal{Y}|/(1+n)^2. \end{aligned}$$

Using the concavity of $\sqrt{\mathsf{z}}$ function, we can conclude that

$$\begin{aligned} &\sum_{i=1}^{\ell+1} \tfrac{n_i\varepsilon_1(\gamma_i,n_i)+C}{n} \\ &\leq 2|\mathcal{X}||\mathcal{Y}|\left(\tfrac{(\ell+1)\ln(1+n)}{n} - \tfrac{(\ell+1)}{n}2\sqrt{\tfrac{n}{\ell+1}\ln(1+n)}\ln\lambda\right) \\ &\quad + \tfrac{2(\ell+1)\ln(1+n)}{n} + \tfrac{(\ell+1)}{n}C \end{aligned} \tag{93a}$$

$$\begin{aligned} &\sum_{i=1}^{\ell+1} \tfrac{n_i\varepsilon_2(\gamma_i,n_i)+D}{n} \\ &\leq 2|\mathcal{X}||\mathcal{Y}|\left(\tfrac{(\ell+1)\ln(1+n)}{n} - \tfrac{(\ell+1)}{n}2\sqrt{\tfrac{n}{\ell+1}\ln(1+n)}\ln\lambda\right) \\ &\quad + \tfrac{\ell+1}{n}2D \end{aligned} \tag{93b}$$

$$\sum_{i=1}^{\ell+1} \varepsilon_{3(\gamma_i,n_i)} \leq 8|\mathcal{X}||\mathcal{Y}|\tfrac{\ell+1}{1+n}. \tag{93c}$$

$$\begin{aligned} P_{\mathbf{x}|\mathsf{m}} &\leq \sum\nolimits_{i=1}^{\ell+1} \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_i \neq (1+\mathsf{m}_i) \middle| \mathsf{M} = \mathsf{m}, {}_{\mathrm{t}}\widehat{\mathsf{M}}_1 = {}_{\mathrm{t}}\mathsf{M}_1, \ldots, {}_{\mathrm{t}}\widehat{\mathsf{M}}_{i-1} = {}_{\mathrm{t}}\mathsf{M}_{i-1}\right] \\ &= \sum\nolimits_{i=1}^{\ell+1} \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_i \neq {}_{\mathrm{t}}\mathsf{M}_i \middle| {}_{\mathrm{t}}\mathsf{M}_i = 1 + \mathsf{m}_i\right] \end{aligned} \tag{88}$$

$$P_{\mathbf{e}|\mathsf{m}}(i) \leq \left[\sum\nolimits_{j=1}^{i} \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_j \neq {}_{\mathrm{t}}\mathsf{M}_j \middle| {}_{\mathrm{t}}\mathsf{M}_j = 1 + \mathsf{m}_j\right]\right] \prod\nolimits_{j=i+1}^{\ell+1} \mathbf{P}\left[{}_{\mathrm{t}}\widehat{\mathsf{M}}_j \neq 1 \middle| {}_{\mathrm{t}}\mathsf{M}_j = 1\right] \tag{89}$$

$$|\mathcal{M}_i| \geq e^{n\mathrm{R}_i - n_i\varepsilon_1(\gamma_i,n) - C} \qquad \forall i = 1,2,\ldots,\ell \tag{92a}$$

$$|\mathcal{M}^i| \geq e^{n\sum_{j=1}^{i}\mathrm{R}_j} e^{-\sum_{j=1}^{i}(n_j\varepsilon_1(\gamma_j,n_j)+C)} \qquad \forall i = 1,2,\ldots,\ell \tag{92b}$$

$$P_{\mathbf{e}|\mathsf{m}}(i) \leq \sum_{j=1}^{i} \varepsilon_{3(\gamma_j,n_j)} \min\left\{1, e^{-n\sum_{j=i+1}^{\ell+1}\eta_i J\left(\frac{\mathrm{R}}{\eta_i}\right)} e^{n\sum_{j=i+1}^{\ell+1} n_j\varepsilon_2(\gamma_j,n_j)+D}\right\} \qquad \forall i = 1,2,\ldots,\ell, \forall \mathsf{m} \in \mathcal{M} \tag{92c}$$

$$P_{\mathbf{x}|\mathsf{m}} \leq \sum_{j=1}^{\ell+1} \varepsilon_{3(\gamma_j,n_j)} \qquad \forall \mathsf{m} \in \mathcal{M} \tag{92d}$$

Then, Lemma 4 follows from (92) and (93) for any $\ell \leq \frac{n}{\ln(n+1)}$. ■

E. Proof of Lemma 5

Proof: For $\mathtt{P}$ defined in (36) as a result of (37), we have

$$\mathbf{P}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] = \sum_{\mathsf{y}^{\mathsf{t}} \in \left\{\mathsf{y}^{\mathsf{t}}:\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right\} \cap \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \tag{94a}$$

$$\mathbf{P}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] = \sum_{\mathsf{y}^{\mathsf{t}} \in \left\{\mathsf{y}^{\mathsf{t}}:\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right\} \cap \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}). \tag{94b}$$

For $\mathtt{P}_{\{\mathcal{A}\}}$ defined in (39) as a result of (41), we have

$$\mathbf{P}_{\{\mathcal{A}_i\}}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] = \sum_{\mathsf{y}^{\mathsf{t}} \in \left\{\mathsf{y}^{\mathsf{t}}:\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right\} \cap \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}_{\{\mathcal{A}\}}(\mathsf{y}^{\mathsf{t}}) \tag{95a}$$

$$\mathbf{P}_{\{\mathcal{A}_i\}}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] = \sum_{\mathsf{y}^{\mathsf{t}} \in \left\{\mathsf{y}^{\mathsf{t}}:\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right\} \cap \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}_{\{\mathcal{A}\}}(\mathsf{y}^{\mathsf{t}}). \tag{95b}$$

Using (94) and (95) together with the data processing inequality for Kullback–Leibler divergence, we get

$$\begin{aligned}\sum_{\mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} & \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \ln \tfrac{\mathtt{P}(\mathsf{y}^{\mathsf{t}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{y}^{\mathsf{t}})} \\ &\geq \mathbf{P}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] \ln \frac{\mathbf{P}\left[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]}{\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]} \\ &\quad + \mathbf{P}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] \ln \frac{\mathbf{P}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]}{\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]}.\end{aligned}$$

Since $0 \leq \mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right] \leq 1$, we have

$$\begin{aligned}\sum_{\mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \ln \tfrac{\mathtt{P}(\mathsf{y}^{\mathsf{t}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{y}^{\mathsf{t}})} &\geq -h\Big(\mathbf{P}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big]\Big) \\ &\quad + \Big(1 - \mathbf{P}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big]\Big) \ln \frac{1}{\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]}. \end{aligned} \tag{96}$$

Note that if $\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$ and $\mathsf{M} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$, then $\widehat{\mathsf{M}} \neq \mathsf{M}$. Consequently

$$\begin{aligned}\mathbf{P}\Big[\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] &= \mathbf{P}\Big[\Big\{\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}), \mathsf{M} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big\}\Big] \\ &\quad + \mathbf{P}\Big[\Big\{\widehat{\mathsf{M}} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}), \mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big\}\Big] \\ &\leq P_{\mathrm{e}} + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]. \end{aligned} \tag{97}$$

Since the binary entropy function $h(\cdot)$ is increasing on the interval $[0, 1/2]$ if $P_{\mathrm{e}} + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right] \leq 1/2$, (96) and (97) imply

$$\begin{aligned}\sum_{\mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \ln \tfrac{\mathtt{P}(\mathsf{y}^{\mathsf{t}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{y}^{\mathsf{t}})} &\geq -h\big(P_{\mathrm{e}} + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]\big) \\ &\quad + \big(1 - P_{\mathrm{e}} - \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]\big) \ln \frac{1}{\mathbf{P}_{\{\mathcal{A}_i\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\right]}. \end{aligned} \tag{98}$$

Let B, B^*, and B_τ be

$$\mathsf{B} \stackrel{\triangle}{=} \ln \frac{\mathtt{P}(\mathsf{Y}^{\mathsf{T}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T}})} \tag{99a}$$

$$\mathsf{B}^* \stackrel{\triangle}{=} \ln \frac{\mathtt{P}(\mathsf{Y}^{\mathsf{T}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T}})} \mathbb{1}_{\{\mathsf{T} < \infty\}} \tag{99b}$$

$$\mathsf{B}_\tau \stackrel{\triangle}{=} \ln \frac{\mathtt{P}(\mathsf{Y}^{\mathsf{T} \wedge \tau})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T} \wedge \tau})} \qquad \forall \tau \in \{1, 2, \ldots\} \tag{99c}$$

where $\mathsf{T} \wedge \tau$ is the minimum of T and τ.

Note that as τ goes to infinity, $\mathsf{B}_\tau \to \mathsf{B}$ and $\mathsf{B}_\tau \to \mathsf{B}^*$ with probability one. Since $|\mathsf{B}_\tau| \leq \mathsf{T} \ln \frac{1}{\lambda}$ and $\mathbf{E}[\mathsf{T}] < \infty$, we can apply the dominated convergence theorem [10, Th. 3, p. 187] to obtain

$$\mathbf{E}[\mathsf{B}] = \mathbf{E}[\mathsf{B}^*] = \lim_{\tau \to \infty} \mathbf{E}[\mathsf{B}_\tau]. \tag{100}$$

Finally for B and B^* defined in (99), we have

$$\mathbf{E}[\mathsf{B}] = \mathbf{E}\Big[\ln \tfrac{\mathtt{P}(\mathsf{Y}^{\mathsf{T}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T}})}\Big] \tag{101a}$$

$$\mathbf{E}[\mathsf{B}^*] = \sum_{\mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \ln \tfrac{\mathtt{P}(\mathsf{y}^{\mathsf{t}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{y}^{\mathsf{t}})}. \tag{101b}$$

Thus, as a result of (100) and (101), we have

$$\mathbf{E}\Big[\ln \tfrac{\mathtt{P}(\mathsf{Y}^{\mathsf{T}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T}})}\Big] = \sum_{\mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}} \mathtt{P}(\mathsf{y}^{\mathsf{t}}) \ln \tfrac{\mathtt{P}(\mathsf{y}^{\mathsf{t}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{y}^{\mathsf{t}})}. \tag{102}$$

Furthermore, using the definition of $\mathtt{P}_{\{\mathcal{A}_i\}}$ given in (39), we get

$$\begin{aligned}\mathbf{E}\Big[\ln \tfrac{\mathtt{P}(\mathsf{Y}^{\mathsf{T}})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}^{\mathsf{T}})}\Big] &= \mathbf{E}\Big[\ln \tfrac{\mathtt{P}(\mathsf{Y}_{\mathsf{T}_{j+1}}^{\mathsf{T}} | \mathsf{Y}^{\mathsf{T}_i})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\mathsf{T}_{j+1}}^{\mathsf{T}} | \mathsf{Y}^{\mathsf{T}_i})}\Big] \\ &= \sum\nolimits_{j=i}^{k} \xi_{i,j} \end{aligned} \tag{103}$$

where for all $i \geq 1$ and $j > i$

$$\xi_{i,j} \stackrel{\triangle}{=} \left\{ \begin{array}{ll} 0 & if\ \mathbf{P}[\mathsf{T}_{j+1} = \mathsf{T}_j] = 1 \\ \mathbf{E}\left[\ln \frac{\mathtt{P}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}} | \mathsf{Y}^{\mathsf{T}_j})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}} | \mathsf{Y}^{\mathsf{T}_j})}\right] & if\ \mathbf{P}[\mathsf{T}_{j+1} = \mathsf{T}_j] < 1 \end{array} \right\}. \tag{104}$$

Assume for the moment that

$$\xi_{i,j} \leq \mathbf{E}[\mathsf{T}_{j+1} - \mathsf{T}_j]\, J(\mathrm{r}_{j+1}) \tag{105}$$

where $\mathsf{T}_{k+1} = \mathsf{T}$ and r_j is defined in (47).

Then, Lemma (5) follows from (98), (102), (103), and (105).

Above, we have proved Lemma 5 by assuming that the inequality given in (105) holds for all i in $\{1, 2, \ldots, k\}$ and j in $\{(i+1), \ldots, (k+1)\}$; in the following, we prove that fact.

First note that if $\mathbf{P}[\mathsf{T}_{j+1} = \mathsf{T}_j] = 1$, then as result of (47) and (103), (105) is equivalent to $0 \leq 0J(0)$ which holds trivially. Thus, we assume hence forth that $\mathbf{P}[\mathsf{T}_{j+1} = \mathsf{T}_j] < 1$, which implies $\mathbf{E}[\mathsf{T}_{j+1} - \mathsf{T}_j] > 0$.

Let us consider the stochastic sequence

$$\mathsf{U}_\tau = \left[- \ln \frac{\mathtt{P}(\mathsf{Y}_{\mathsf{T}_j+1}^{\tau} | \mathsf{Y}^{\mathsf{T}_j})}{\mathtt{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\mathsf{T}_j+1}^{\tau} | \mathsf{Y}^{\mathsf{T}_j})} + \sum_{k=\mathsf{T}_j+1}^{\tau} J(\mathsf{I}(\mathsf{M}; \mathsf{Y}_k | \mathsf{Y}^{k-1})) \right] \mathbb{1}_{\{\tau > \mathsf{T}_j\}} \tag{106}$$

where $\mathsf{I}\left(\mathsf{M}; \mathsf{Y}_k \,\middle|\, \mathsf{Y}^{k-1}\right)$ is the conditional mutual information between M and Y_k given Y^{k-1}, defined as

$$\mathsf{I}\left(\mathsf{M}; \mathsf{Y}_k \,\middle|\, \mathsf{Y}^{k-1}\right) \stackrel{\triangle}{=} \mathbf{E}\left[\ln \tfrac{\mathtt{P}(\mathsf{Y}_k | \mathsf{M}, \mathsf{Y}^{k-1})}{\mathtt{P}(\mathsf{Y}_k | \mathsf{Y}^{k-1})} \,\middle|\, \mathsf{Y}^{k-1}\right].$$

Note that as it was the case for conditional entropy, while defining the conditional mutual information, we do not take the average over the conditioned random variable. Thus, $\mathsf{I}\left(\mathsf{M}; \mathsf{Y}_k \,\middle|\, \mathsf{Y}^{k-1}\right)$ is itself a random variable.

For U_τ defined in (106), we have

$$\mathsf{U}_{\tau+1}-\mathsf{U}_\tau = \left(-\ln\frac{\mathsf{P}(\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)}{\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)}+J(\mathsf{I}(\mathsf{M};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau))\right)\mathbb{1}_{\{\tau\geq\mathsf{T}_j\}}. \quad (107)$$

Conditioned on Y^τ random variables $\mathsf{M}-\mathsf{X}_{\tau+1}-\mathsf{Y}_{\tau+1}$ form a Markov chain: thus, as a result of the data processing inequality for the mutual information, we have $\mathsf{I}(\mathsf{X}_{\tau+1};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau) > \mathsf{I}(\mathsf{M};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)$. Since $J(\cdot)$ is a decreasing function, this implies that

$$J(\mathsf{I}(\mathsf{M};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau))\geq J(\mathsf{I}(\mathsf{X}_{\tau+1};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)). \quad (108)$$

Furthermore, because of the definitions of $J(\cdot)$, P, and $\mathsf{P}_{\{\mathcal{A}\}}$ given in (17), (36), and (39), the convexity of Kullback–Leibler divergence and Jensen's inequality, we have

$$J(\mathsf{I}(\mathsf{X}_{\tau+1};\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau))\geq\mathbf{E}\left[\ln\frac{\mathsf{P}(\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)}{\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\tau+1}|\mathsf{Y}^\tau)}\middle|\mathsf{Y}^\tau\right]. \quad (109)$$

Using (107)–(109), we get

$$\mathbf{E}[\mathsf{U}_{\tau+1}|\mathsf{Y}^\tau]\geq\mathsf{U}_\tau. \quad (110)$$

Recall that $\min_{\mathsf{x},\mathsf{y}}W_{\mathsf{x}}(\mathsf{y})=\lambda$ and $|J(\cdot)|\leq D$. Thus, as a result of (107), we have

$$\mathbf{E}[|\mathsf{U}_{\tau+1}-\mathsf{U}_\tau||\mathsf{Y}^\tau]\leq\ln\tfrac{1}{\lambda}+D. \quad (111)$$

As a result of (110) and (111) and the fact that $\mathsf{U}_0=0$, U_τ is a submartingale.

Recall that we have assumed that $\mathbf{P}[\mathsf{T}_{j+1}\leq\mathsf{T}]=1$ and $\mathbf{E}[\mathsf{T}]<\infty$; consequently

$$\mathbf{E}[\mathsf{T}_{j+1}]<\infty. \quad (112)$$

Because of (111) and (112), we can apply a version of Doob's optional stopping theorem [10, Th. 2, p. 487] to the submartingale U_τ and the stopping time T_{j+1} to obtain $\mathbf{E}[\mathsf{U}_{\mathsf{T}_{j+1}}]\geq\mathbf{E}[\mathsf{U}_0]=0$. Consequently

$$\mathbf{E}\left[\ln\frac{\mathsf{P}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}|\mathsf{Y}^{\mathsf{T}_j})}{\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}|\mathsf{Y}^{\mathsf{T}_j})}\right]\leq\mathbf{E}\left[\sum\nolimits_{\tau=\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}J\big(\mathsf{I}\big(\mathsf{M};\mathsf{Y}_\tau\big|\mathsf{Y}^{\tau-1}\big)\big)\right]. \quad (113)$$

Note that as a result of the concavity of $J(\cdot)$ and Jensen's inequality, we have

$$\begin{aligned}&\mathbf{E}\left[\sum\nolimits_{\tau=\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}J\big(\mathsf{I}\big(\mathsf{M};\mathsf{Y}_\tau\big|\mathsf{Y}^{\tau-1}\big)\big)\right]\\&=\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]\,\mathbf{E}\left[\sum\nolimits_{\tau\geq1}\frac{\mathbb{1}_{\{\mathsf{T}_{j+1}\geq\tau>\mathsf{T}_j\}}J\big(\mathsf{I}\big(\mathsf{M};\mathsf{Y}_\tau\big|\mathsf{Y}^{\tau-1}\big)\big)}{\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]}\right]\\&\leq\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]\,J\left(\frac{\mathbf{E}\left[\sum_{\tau\geq1}\mathbb{1}_{\{\mathsf{T}_{j+1}\geq\tau>\mathsf{T}_j\}}\mathsf{I}\big(\mathsf{M};\mathsf{Y}_\tau\big|\mathsf{Y}^{\tau-1}\big)\right]}{\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]}\right).\end{aligned} \quad (114)$$

In order to calculate the argument of $J(\cdot)$ in (114), consider the stochastic sequence

$$\mathsf{V}_\tau=\mathsf{H}(\mathsf{M}|\mathsf{Y}^\tau)+\sum\nolimits_{j=1}^{\tau}\mathsf{I}\left(\mathsf{M};\mathsf{Y}_j\middle|\mathsf{Y}^{j-1}\right). \quad (115)$$

Clearly, $\mathbf{E}[\mathsf{V}_{\tau+1}|\mathsf{Y}^\tau]=\mathsf{V}_\tau$ and $\mathbf{E}[|\mathsf{V}_\tau|]\leq\ln|\mathcal{M}|+C\tau<\infty$. Hence, V_τ is a martingale.

Furthermore

$$\mathbf{E}[|\mathsf{V}_{\tau+1}-\mathsf{V}_\tau||\mathsf{Y}^\tau]\leq\ln|\mathcal{M}|+C. \quad (116)$$

Recall that we have assumed that $\mathbf{P}[\mathsf{T}_j\leq\mathsf{T}_{j+1}\leq\mathsf{T}]=1$ and $\mathbf{E}[\mathsf{T}]<\infty$; consequently

$$\mathbf{E}[\mathsf{T}_j]\leq\mathbf{E}[\mathsf{T}_{j+1}]<\infty. \quad (117)$$

As a result of (116) and (117), we can apply Doob's optimal stopping theorem, [10, Th. 2, p. 487] to V_τ both at stopping time T_j and at stopping time T_{j+1}, i.e., $\mathbf{E}[\mathsf{V}_{\mathsf{T}_{j+1}}]=\mathbf{E}[\mathsf{V}_0]$ and $\mathbf{E}[\mathsf{V}_{\mathsf{T}_j}]=\mathbf{E}[\mathsf{V}_0]$. Consequently

$$\mathbf{E}\left[\sum_{\tau\geq1}\mathbb{1}_{\{\mathsf{T}_{j+1}\geq\tau>\mathsf{T}_j\}}\mathsf{I}\left(\mathsf{M};\mathsf{Y}_\tau\middle|\mathsf{Y}^{\tau-1}\right)\right]=\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_j})-\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_{j+1}})\big]. \quad (118)$$

Using (113), (114), and (118)

$$\mathbf{E}\left[\ln\frac{\mathsf{P}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}|\mathsf{Y}^{\mathsf{T}_j})}{\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{Y}_{\mathsf{T}_j+1}^{\mathsf{T}_{j+1}}|\mathsf{Y}^{\mathsf{T}_j})}\right]\leq\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]\,J\left(\frac{\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_j})-\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_{j+1}})\big]}{\mathbf{E}[\mathsf{T}_{j+1}-\mathsf{T}_j]}\right). \quad (119)$$

Hence, inequality given in (105) is not only when $\mathbf{P}[\mathsf{T}_{j+1}=\mathsf{T}_j]=1$ but also when $\mathbf{P}[\mathsf{T}_{j+1}=\mathsf{T}_j]<1$. ■

F. Proof of Lemma 6 for the Case $\mathbf{E}[\mathsf{T}]<\infty$

Proof: In order to bound $P_{\mathbf{e}|\mathsf{m}}$ from below, we apply Lemma 5 for $(\mathsf{T}_1,\mathcal{A}_1)$ and $(\mathsf{T}_2,\mathcal{A}_2)$ given in (48)–(51) and use the fact that $J(\cdot)\leq D$ to get

$$\ln P_{\mathbf{e}|\mathsf{m}}\geq\frac{-h\big(P_{\mathbf{e}}+|\mathcal{M}|^{-1}\big)-\mathbf{E}[\mathsf{T}_2]\,J\left(\frac{\mathbf{E}\left[\mathsf{H}(\mathsf{M})-\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right]}{\mathbf{E}[\mathsf{T}_2]}\right)-\mathbf{E}[\mathsf{T}-\mathsf{T}_2]D}{1-P_{\mathbf{e}}-|\mathcal{M}|^{-1}} \quad (120a)$$

$$\ln\mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]\geq\frac{-h\big(P_{\mathbf{e}}+\mathbf{P}\big[\mathsf{M}\in\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\big]\big)-\mathbf{E}[\mathsf{T}-\mathsf{T}_2]D}{1-P_{\mathbf{e}}-\mathbf{P}\big[\mathsf{M}\in\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\big]} \quad (120b)$$

provided that $|\mathcal{M}|^{-1}+P_{\mathbf{e}}\leq1/2$ and $\mathbf{P}\left[\mathsf{M}\in\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]+P_{\mathbf{e}}\leq1/2$.

We start with bounding $\mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}}\notin\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]$ from above and $\mathbf{P}\left[\mathsf{M}\notin\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]$ from below.

1) Since $\min_{\mathsf{x}\in\mathcal{X},\mathsf{y}\in\mathcal{Y}}W_{\mathsf{x}}(\mathsf{y})=\lambda$, the posterior probability of a message at time $\tau+1$ cannot be smaller than λ times the posterior probability of the same message at time τ. Hence,

for the stopping time T_2 defined in (50), random[33] set $\mathcal{A}_2$ defined in (51), and $\delta < \frac{1}{2}$, we have

$$\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathsf{Y}^{\mathsf{T}_2} = \mathsf{y}^{\mathsf{t}_2}\right] > \lambda\delta \quad \forall \mathsf{y}^{\mathsf{t}_2} \in \mathcal{Y}^{\mathsf{T}_2 *}. \tag{121}$$

As a result of the definition of $\mathsf{P}_{\{\mathcal{A}_2\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}})$ given in (39), we have

$$\mathsf{P}_{\{\mathcal{A}_2\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) < \mathsf{P}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \frac{\mathbb{1}_{\{\mathsf{m} \in \mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})\}}}{\lambda\delta} \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}. \tag{122}$$

If the decoded message $\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}})$ is not in $\mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})$ and message m is in $\mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})$, then $\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}}) \neq \mathsf{m}$

$$\mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}}) \notin \mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})\right\}} \mathbb{1}_{\{\mathsf{m} \in \mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})\}} \leq \mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}}) \neq \mathsf{m}\right\}} \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}}. \tag{123}$$

Using (122) and (123), we get

$$\mathsf{P}_{\{\mathcal{A}_2\}}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}}) \notin \mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})\right\}} < \mathsf{P}(\mathsf{m}, \mathsf{y}^{\mathsf{t}}) \frac{\mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}}) \neq \mathsf{m}\right\}}}{\lambda\delta} \quad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}. \tag{124}$$

If we sum over all $(\mathsf{m}, \mathsf{y}^{\mathsf{t}})$'s in $\mathcal{M} \times \mathcal{Y}^{\mathsf{T}*}$ and use (37) and (41), we get

$$\mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right] < \frac{\mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M}\right]}{\lambda\delta} = \frac{P_{\mathrm{e}}}{\lambda\delta}. \tag{125}$$

2) The probability of an event Γ_1 is lower bounded by the probability of its intersection with any event Γ_2, i.e., $\mathbf{P}[\Gamma_1] \geq \mathbf{P}[\{\Gamma_1, \Gamma_2\}]$:

$$\begin{aligned} P_{\mathrm{e}} &= \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M}\right] \\ &\geq \mathbf{P}\left[\left\{\widehat{\mathsf{M}} \neq \mathsf{M}, \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right\}\right] \\ &= \mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M} \middle| \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] \mathbf{P}\left[\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right]. \end{aligned} \tag{126}$$

Note that if $\mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2}) = \mathcal{M}$, then T is reached before any of the messages reach a posterior probability of $1 - \delta$. Thus

$$\mathbf{P}\left[\widehat{\mathsf{M}} \neq \mathsf{M} \middle| \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] > \delta. \tag{127}$$

Thus, as a result of (126) and (127), we have

$$\mathbf{P}\left[\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] < \frac{P_{\mathrm{e}}}{\delta}. \tag{128}$$

On the other hand, if $\mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2}) \neq \mathcal{M}$, then the most likely message with a probability at least $(1-\delta)$ is excluded from $\mathcal{A}_2(\mathsf{y}^{\mathsf{t}_2})$. Thus

$$\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \neq \mathcal{M}\right] \leq \delta. \tag{129}$$

[33]The set $\mathcal{A}_2$ is random in the sense that it depends on previous channel outputs.

Using (128) and (129) together with total probability formula, we get

$$\begin{aligned} &\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right] \\ &= \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] \mathbf{P}\left[\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] \\ &\quad + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \neq \mathcal{M}\right] \mathbf{P}\left[\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \neq \mathcal{M}\right] \\ &\leq \mathbf{P}\left[\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) = \mathcal{M}\right] + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathcal{A}_1(\mathsf{Y}^{\mathsf{T}_1}) \neq \mathcal{M}\right] \\ &< \frac{P_{\mathrm{e}}}{\delta} + \delta. \end{aligned} \tag{130}$$

We plug the bounds on $\mathbf{P}_{\{\mathcal{A}_2\}}\left[\widehat{\mathsf{M}} \notin \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]$ and $\mathbf{P}\left[\mathsf{M} \notin \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right]$ given in (125) and (130) in (120) to get

$$\ln P_{\mathbf{e}|\mathsf{m}} \geq \frac{-h(\tilde{\epsilon}_1) - \mathbf{E}[\mathsf{T}_2] J\left(\frac{\mathbf{E}\left[\mathsf{H}(\mathsf{M}) - \mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right]}{\mathbf{E}[\mathsf{T}_2]}\right) - \mathbf{E}[\mathsf{T} - \mathsf{T}_2] D}{1 - \tilde{\epsilon}_1} \tag{131a}$$

$$\ln \frac{P_{\mathbf{e}}}{\lambda\delta} \geq \frac{-h(\tilde{\epsilon}_1) - \mathbf{E}[\mathsf{T} - \mathsf{T}_2] D}{1 - \tilde{\epsilon}_1} \tag{131b}$$

provided that $\tilde{\epsilon}_1 \leq 1/2$ where $\tilde{\epsilon}_1 = P_{\mathbf{e}} + \delta + \frac{P_{\mathbf{e}}}{\delta} + |\mathcal{M}|^{-1}$.

Now, we bound $\mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right]$ from below. Note that $\mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}}$ is a discrete random variable that is either zero or one; its conditional entropy given $\mathsf{Y}^{\mathsf{T}_2}$ is given by

$$\mathsf{H}(\mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}} | \mathsf{Y}^{\mathsf{T}_2}) = h\left(\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathsf{Y}^{\mathsf{T}_2}\right]\right). \tag{132}$$

Furthermore, since $\mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}}$ is a function of $\mathsf{Y}^{\mathsf{T}_2}$ and M, chain rule entropy implies that

$$\begin{aligned} \mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2}) &= \mathsf{H}(\mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}} | \mathsf{Y}^{\mathsf{T}_2}) \\ &\quad + \mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2}, \mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}}) \middle| \mathsf{Y}^{\mathsf{T}_2}\right]. \end{aligned} \tag{133}$$

Since $\mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})$ has at most $|\mathcal{M}|$ elements and its complement, $\mathcal{M} \setminus \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})$, has at most one element, we can bound the conditional entropy of the messages as follows:

$$\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2}, \mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}}) \leq \mathbb{1}_{\left\{\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2})\right\}} \ln |\mathcal{M}|. \tag{134}$$

Thus, using (132)–(134), we get

$$\begin{aligned} \mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2}) &\leq h\left(\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathsf{Y}^{\mathsf{T}_2}\right]\right) \\ &\quad + \mathbf{P}\left[\mathsf{M} \in \mathcal{A}_2(\mathsf{Y}^{\mathsf{T}_2}) \middle| \mathsf{Y}^{\mathsf{T}_2}\right] \ln |\mathcal{M}|. \end{aligned} \tag{135}$$

Then, using concavity of the binary entropy function $h(\cdot)$ together with (130) and (135), we get

$$\mathbf{E}\left[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_2})\right] < h\left(\delta + \tfrac{P_{\mathrm{e}}}{\delta}\right) + \left(\delta + \tfrac{P_{\mathrm{e}}}{\delta}\right) \ln |\mathcal{M}| \tag{136}$$

provided that $\delta + \frac{P_{\mathrm{e}}}{\delta} \leq 1/2$.

If we plug in (136) and the identity $\mathsf{H}(\mathsf{M}) = \ln |\mathcal{M}|$ in (131), we get

$$(1 - \tilde{\epsilon}_1) \frac{\ln P_{\mathbf{e}|\mathsf{m}}}{\mathbf{E}[\mathsf{T}]} \geq -\frac{h(\tilde{\epsilon}_1)}{\mathbf{E}[\mathsf{T}]} - \eta J\left(\frac{(1-\tilde{\epsilon}_1)\mathrm{R} - h(\tilde{\epsilon}_1)/\mathbf{E}[\mathsf{T}]}{\eta}\right) - (1 - \eta) D \tag{137a}$$

$$-(1 - \tilde{\epsilon}_1)\mathrm{E} \geq \frac{-h(\tilde{\epsilon}_1) + \ln \lambda\delta}{\mathbf{E}[\mathsf{T}]} - (1 - \eta) D \tag{137b}$$

provided that $\tilde{\epsilon}_1 \leq 1/2$ where $\eta = \frac{\mathbf{E}[\mathsf{T}_2]}{\mathbf{E}[\mathsf{T}]}$, $\tilde{\epsilon}_1 = P_{\mathrm{e}} + \delta + \frac{P_{\mathrm{e}}}{\delta} + |\mathcal{M}|^{-1}$, $\mathrm{R} = \frac{|\mathcal{M}|}{\mathbf{E}[\mathsf{T}]}$, and $\mathrm{E} = \frac{-\ln P_{\mathrm{e}}}{\mathbf{E}[\mathsf{T}]}$.

Note that the inequality given in (137b) bounds the value of η from above

$$\eta \leq 1 - \tfrac{(1-\tilde{\epsilon}_1)\mathrm{E} - \tilde{\epsilon}_2}{D} \tag{138}$$

where $\tilde{\epsilon}_2 = \frac{h(\tilde{\epsilon}_1) - \ln \lambda\delta}{\mathbf{E}[\mathsf{T}]}$.

Furthermore, for any $\eta_1 \leq \eta_2 \leq \frac{\widetilde{\mathrm{R}}}{C}$, as a result of concavity of $J(\cdot)$, we have

$$\begin{aligned}\eta_1 J\Big(\tfrac{\widetilde{\mathrm{R}}}{\eta_1}\Big) + (1-\eta_1)D &= \eta_1 J\Big(\tfrac{\widetilde{\mathrm{R}}}{\eta_1}\Big) + (\eta_2 - \eta_1)J(0) \\ &\quad + (1-\eta_2)D \\ &\leq \eta_2 J\Big(\tfrac{\widetilde{\mathrm{R}}}{\eta_2}\Big) + (1-\eta_2)D. \end{aligned} \tag{139}$$

Using (138) and (139) we see that the bound in (137a) is lower bounded by its value at $\eta = 1 - \frac{(1-\tilde{\epsilon}_1)\mathrm{E} - \tilde{\epsilon}_2}{D}$ if $\mathrm{E} \geq \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$ and by its value at $\eta = 1$ otherwise, i.e.,

$$\tfrac{\ln P_{\mathrm{e}|\mathsf{m}}}{\mathbf{E}[\mathsf{T}]} \geq \left\{ \begin{array}{ll} -\mathrm{E} - \left(1 - \frac{\mathrm{E}-\tilde{\epsilon}}{D}\right) J\left(\frac{\mathrm{R} - \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}}{1 - \frac{\mathrm{E}-\tilde{\epsilon}}{D}}\right) & \text{if } \mathrm{E} \geq \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1} \\ -\frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1} - \frac{1}{1-\tilde{\epsilon}_1} J((1-\tilde{\epsilon}_1)\mathrm{R} - \tilde{\epsilon}_2) & \text{if } \mathrm{E} < \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1} \end{array} \right\}$$

where $\tilde{\epsilon} = \frac{\tilde{\epsilon}_1 D + \tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$.

Then, for the case $\mathrm{E} \geq \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$, Lemma 6 follows from the fact that $J(\cdot)$ is a nonnegative decreasing function. For the case $\mathrm{E} < \frac{\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$, Lemma 6 follows from the fact that $J(\cdot)$ is a concave nonnegative decreasing function. ■

G. Proof of Lemma 7 for the Case $\mathbf{E}[\mathsf{T}] < \infty$

Proof: We start with proving the bounds given in (57) and (58).

1) Let us start with the bound on $\mathbf{P}_{\{\mathcal{A}_i\}}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big]$ given in (57). Since $\min_{\mathsf{x}\in\mathcal{X},\mathsf{y}\in\mathcal{Y}} W_{\mathsf{x}}(\mathsf{y}) = \lambda$, the posterior probability of a $\mathsf{m}^i \in \mathcal{M}^i$ at time $\tau+1$ cannot be smaller than λ times its value at time τ. Hence, for $\delta < 1/2$, as a result of definitions of T_i and $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$ given in (55) and (56), we have

$$\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \big| \mathsf{Y}^{\mathsf{T}_i} = \mathsf{y}^{\mathsf{t}_i}\big] > \lambda\delta \qquad \forall \mathsf{y}^{\mathsf{t}_i} \in \mathcal{Y}^{\mathsf{T}_i *}, i \in \{1,2,\ldots,\ell\}.$$

Then, as a result of the definition of $\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{m},\mathsf{y}^{\mathsf{t}})$ given in (39), we have

$$\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{m},\mathsf{y}^{\mathsf{t}}) < \mathsf{P}(\mathsf{m},\mathsf{y}^{\mathsf{t}}) \tfrac{\mathbb{1}_{\{\mathsf{m}\in\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})\}}}{\lambda\delta} \qquad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}, i \in \{1,2,\ldots,\ell\}. \tag{140}$$

For $\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})$ given in (56), if the decoded message $\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}})$ is not in $\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})$ but m is in $\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})$, then $\widehat{\mathsf{M}}^i(\mathsf{y}^{\mathsf{t}}) \neq \mathsf{m}^i$

$$\mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}})\notin\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})\right\}} \mathbb{1}_{\{\mathsf{m}\in\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})\}} \leq \mathbb{1}_{\left\{\widehat{\mathsf{M}}^i \neq \mathsf{m}^i\right\}} \qquad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}}, i \in \{1,2,\ldots,\ell\}. \tag{141}$$

Using (140) and (141), we get

$$\mathsf{P}_{\{\mathcal{A}_i\}}(\mathsf{m},\mathsf{y}^{\mathsf{t}}) \mathbb{1}_{\left\{\widehat{\mathsf{M}}(\mathsf{y}^{\mathsf{t}})\notin\mathcal{A}_i(\mathsf{y}^{\mathsf{t}_i})\right\}} < \mathsf{P}(\mathsf{m},\mathsf{y}^{\mathsf{t}}) \tfrac{\mathbb{1}_{\left\{\widehat{\mathsf{M}}^i\neq\mathsf{m}^i\right\}}}{\lambda\delta} \qquad \forall \mathsf{m} \in \mathcal{M}, \mathsf{y}^{\mathsf{t}} \in \mathcal{Y}^{\mathsf{T}*}, i \in \{1,2,\ldots,\ell\}.$$

If we sum over all $(\mathsf{m},\mathsf{y}^{\mathsf{t}})$'s in $\mathcal{M}\times\mathcal{Y}^{\mathsf{T}*}$ and use (37) and (41), we get

$$\begin{aligned}\mathbf{P}_{\{\mathcal{A}_i\}}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big] &< \tfrac{\mathbf{P}\left[\widehat{\mathsf{M}}^i \neq \mathsf{M}^i\right]}{\lambda\delta} \\ &= \tfrac{P_{\mathrm{e}}(i)}{\lambda\delta} \quad \forall i \in \{1,2,\ldots,\ell\}. \end{aligned} \tag{142}$$

2) Let us now prove the bound on $\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big]$ given in (58).

a) If $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}$, then at T_i, there is a m^i with posterior probability $(1-\delta)$ and all the messages m of the form $\mathsf{m} = (\mathsf{m}^i, \mathsf{m}_{i+1}, \ldots, \mathsf{m}_k)$ are excluded from $\mathcal{A}_i$. Consequently, we have

$$\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \big| \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}\big] < \delta. \tag{143}$$

b) If $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}$, then at T_i, there is no m^i with posterior probability $(1-\delta)$ and $\mathsf{T}_i = \mathsf{T}$. Since $\widehat{\mathsf{M}}^i \neq \mathsf{M}^i$ implies that $\widehat{\mathsf{M}} \neq \mathsf{M}$, we have

$$\mathbf{P}\Big[\widehat{\mathsf{M}} \neq \mathsf{M} \Big| \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\Big] \geq \delta. \tag{144}$$

As a result of total probability formula for $\mathbf{P}\Big[\widehat{\mathsf{M}} \neq \mathsf{M}\Big]$, we have

$$\begin{aligned} P_{\mathrm{e}} &= \mathbf{P}\Big[\widehat{\mathsf{M}} \neq \mathsf{M} \Big| \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\Big] \mathbf{P}\big[\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\big] \\ &\quad + \mathbf{P}\Big[\widehat{\mathsf{M}} \neq \mathsf{M} \Big| \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}\Big] \mathbf{P}\big[\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}\big] \\ &\geq \delta \mathbf{P}\big[\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\big]. \end{aligned} \tag{145}$$

If use the total probability formula for $\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big]$ together with (143) and (145), we get

$$\begin{aligned}\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big] &= \mathbf{P}\big[\big\{\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}), \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}\big\}\big] \\ &\quad + \mathbf{P}\big[\big\{\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}), \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\big\}\big] \\ &\leq \mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \big| \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) \neq \mathcal{M}\big] \\ &\quad + \mathbf{P}\big[\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i}) = \mathcal{M}\big] \\ &\leq \delta + \tfrac{P_{\mathrm{e}}}{\delta}. \end{aligned}$$

We apply Lemma 5 for $(\mathsf{T}_1,\mathcal{A}_1),\ldots,(\mathsf{T}_k,\mathcal{A}_k)$ defined in (55) and (56); use the bounds on $\mathbf{P}_{\{\mathcal{A}_i\}}\Big[\widehat{\mathsf{M}} \notin \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\Big]$ and $\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big]$ given in (57) and (58). Then, we can conclude that if $P_{\mathrm{e}} + \delta + P_{\mathrm{e}}/\delta \leq 1/2$, then

$$(1-\tilde{\epsilon}_3)\mathrm{E}_i \leq \tilde{\epsilon}_5 + \sum_{j=i+1}^{\ell+1} \nu_j J(\mathrm{r}_j) \quad i = 1,2,\ldots,\ell \tag{146}$$

where $\mathrm{R}_i, \mathrm{E}_i, \tilde{\epsilon}_3$, and $\tilde{\epsilon}_5$ are defined in Lemma 7, r_j's are defined in (47) of Lemma 5, and ν_j's are defined as follows[34]:

$$\nu_j \triangleq \frac{\mathbf{E}[\mathsf{T}_j] - \mathbf{E}[\mathsf{T}_{j-1}]}{\mathbf{E}[\mathsf{T}]} \quad \forall j \in \{1, 2, \ldots, \ell+1\}. \tag{147}$$

Depending on the values of ν_j and r_j, the bound in (146) takes different values. However, ν_j and r_j are not changing freely. As a result of (118) and the fact that $\mathsf{I}(\mathsf{M}; \mathsf{Y}_{t+1}|\mathsf{Y}^t) \leq C$, we have

$$\mathrm{r}_j \leq C \quad j \in \{1, 2, \ldots, (\ell+1)\}. \tag{148}$$

In addition, ν_j's and r_j's are constrained by the definitions of T_j and $\mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})$ given in (55) and (56). At T_j with high probability, one element of $\mathcal{M}^j$ has a posterior probability $(1-\delta)$. In the following, we use this fact to bound $\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_j})\big]$ from above. Then, we turn this bound into a constraint on the values of ν_j's and r_j's and use that constraint together with (146) and (148) to bound E_i's from above.

For all j in $\{1, 2, \ldots, \ell\}$, $\mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})\}}$ is a discrete random variable that is either zero or one; its conditional entropy is given by

$$\mathsf{H}(\mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})\}}|\mathsf{Y}^{\mathsf{T}_j}) = h\big(\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})\big|\,\mathsf{Y}^{\mathsf{T}_j}\big]\big). \tag{149}$$

Furthermore, since $\mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}}$ is a function of $\mathsf{Y}^{\mathsf{T}_i}$ and M, the chain rule entropy implies that

$$\begin{aligned}\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i}) = \mathsf{H}(\mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}}|\mathsf{Y}^{\mathsf{T}_i}) \\ +\mathbf{E}\Big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i}, \mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}})\Big|\,\mathsf{Y}^{\mathsf{T}_i}\Big].\end{aligned} \tag{150}$$

Note that $\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$ has at most $|\mathcal{M}|$ elements and its complement, $\mathcal{M} \setminus \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})$, has at most $\frac{|\mathcal{M}|}{|\mathcal{M}^i|}$ elements. We can bound the conditional entropy of the messages $\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i}, \mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}})$ as follows:

$$\begin{aligned}\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i}, \mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}}) &\leq \mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}} \ln|\mathcal{M}| \\ &\quad + \mathbb{1}_{\{\mathsf{M}\notin\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}} \ln\tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \\ &= \ln\tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} + \mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\}} \ln|\mathcal{M}^i|.\end{aligned} \tag{151}$$

Thus, using (149)–(151), we get

$$\begin{aligned}\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i}) \leq h\big(\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})\big|\,\mathsf{Y}^{\mathsf{T}_j}\big]\big) + \ln\tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \\ + \mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big|\,\mathsf{Y}^{\mathsf{T}_i}\big] \ln|\mathcal{M}^i|.\end{aligned} \tag{152}$$

If we take the expectation of both sides of the inequality (152) and use the concavity of the binary entropy function, we get

$$\begin{aligned}\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i})\big] \leq h\big(\mathbf{P}\big[\mathsf{M} \in \mathcal{A}_j(\mathsf{Y}^{\mathsf{T}_j})\big]\big) + \ln\tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \\ + \mathbf{P}\big[\mathsf{M} \in \mathcal{A}_i(\mathsf{Y}^{\mathsf{T}_i})\big] \ln|\mathcal{M}^i|.\end{aligned}$$

[34]We use the convention $\mathsf{T}_0 = 0$ and $\mathsf{T}_{\ell+1} = \mathsf{T}$.

Using the inequality given (58) and the fact that binary entropy function is an increasing function on the interval $[0, 1/2]$, we see that

$$\begin{aligned}\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i})\big] < h\big(P_{\mathrm{e}} + \delta + \tfrac{P_{\mathrm{e}}}{\delta}\big) + \ln\tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \\ + (P_{\mathrm{e}} + \delta + \tfrac{P_{\mathrm{e}}}{\delta}) \ln|\mathcal{M}^i|\end{aligned} \tag{153}$$

provided that $P_{\mathrm{e}} + \delta + \frac{P_{\mathrm{e}}}{\delta} \leq 1/2$.

Note that as a result of Fano's inequality for $\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}})\big]$, we have

$$\mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}})\big] < h(P_{\mathrm{e}}) + P_{\mathrm{e}} \ln|\mathcal{M}|. \tag{154}$$

If we divide both sides of the inequalities (153) and (154) to $\mathbf{E}[\mathsf{T}]$, we see that following bounds holds:

$$\begin{aligned}\tfrac{\mathbf{E}[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i})]}{\mathbf{E}[\mathsf{T}]} &\leq \tilde{\epsilon}_4 + \mathrm{R} - \sum_{j=1}^{i} \mathrm{R}_i + \tilde{\epsilon}_3 \sum_{j=1}^{i} \mathrm{R}_i \\ &\qquad i = 1, 2, \ldots, \ell \qquad (155a)\\ \mathbf{E}\big[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}})\big] &\leq h(P_{\mathrm{e}}) + P_{\mathrm{e}} R. \qquad (155b)\end{aligned}$$

Note that

$$\tfrac{\mathbf{E}[\mathsf{H}(\mathsf{M}|\mathsf{Y}^{\mathsf{T}_i})]}{\mathbf{E}[\mathsf{T}]} = \mathrm{R} - \sum_{j=1}^{i} \nu_j \mathrm{r}_j \quad i = 1, 2, \ldots, (\ell+1). \tag{156}$$

Using (155) and (156), we get

$$\begin{aligned}\sum_{j=1}^{i} \nu_j \mathrm{r}_j &\geq (1-\tilde{\epsilon}_3) \sum_{j=1}^{i} \mathrm{R}_j - \tilde{\epsilon}_4 \quad i = 1, 2, \ldots, \ell \qquad (157a)\\ \sum_{j=1}^{\ell+1} \nu_j \mathrm{r}_j &\geq (1-P_{\mathrm{e}})\mathrm{R} - \tfrac{h(P_{\mathrm{e}})}{\mathbf{E}[\mathsf{T}]} \qquad (157b)\end{aligned}$$

where r_j's and ν_j's are given in (47) and (147) respectively.

Thus, using (47), (146)–(148), and (157), we reach the following conclusion. For any variable-length block code satisfying the hypothesis of Lemma 7 and for any positive δ such that $P_{\mathrm{e}} + \delta + \frac{P_{\mathrm{e}}}{\delta} \leq \frac{1}{2}$

$$\begin{aligned}(1-\tilde{\epsilon}_3)\mathrm{E}_i - \tilde{\epsilon}_5 &\leq \sum_{j=i+1}^{\ell+1} \nu_j J(\mathrm{r}_j) \quad i = 1, 2, \ldots, \ell \qquad (158a)\\ (1-\tilde{\epsilon}_3) \sum_{j=1}^{i} \mathrm{R}_j - \tilde{\epsilon}_4 &\leq \sum_{j=1}^{i} \nu_j \mathrm{r}_j \qquad i = 1, 2, \ldots, \ell \qquad (158b)\\ (1-P_{\mathrm{e}})\mathrm{R} - \tfrac{h(P_{\mathrm{e}})}{\mathbf{E}[\mathsf{T}]} &\leq \sum_{j=1}^{\ell+1} \nu_j \mathrm{r}_j \qquad (158c)\end{aligned}$$

for some $(\nu_1^{\ell+1}, \mathrm{r}_1^{\ell+1})$ such that

$$\begin{aligned}\mathrm{r}_i &\in [0, C] \quad i = 1, 2, \ldots, (\ell+1) \qquad (159a)\\ \nu_i &\geq 0 \qquad i = 1, 2, \ldots, (\ell+1) \qquad (159b)\\ \sum\nolimits_{i=1}^{\ell+1} \nu_i &= 1. \qquad (159c)\end{aligned}$$

We show in the following if the constraints given in (158) is satisfied for some $(\nu_1^{\ell+1},\mathrm{r}_1^{\ell+1})$ satisfying (159), and constraints given in (60) is satisfied for some (η_1^{ℓ}) satisfying (61).

One can confirm numerically that

$$(1-\tilde{\epsilon}_3)\ln 2 > h(\tilde{\epsilon}_3) \quad \forall \tilde{\epsilon}_3 \in \left[0,\tfrac{1}{5}\right].$$

Recall that we have assumed that $P_{\mathrm{e}}+\delta+\frac{P_{\mathrm{e}}}{\delta}\le\frac{1}{5}$, i.e., $\tilde{\epsilon}_3\le\frac{1}{5}$. Thus

$$(1-\tilde{\epsilon}_3)\mathrm{R}_1-\tilde{\epsilon}_4>0. \tag{160}$$

Let η_1, $\widetilde{\mathrm{r}}_1$, $\widetilde{\nu}_2$, and $\widetilde{\mathrm{r}}_2$ be

$$\begin{aligned}\eta_1&=\tfrac{(1-\tilde{\epsilon}_3)\mathrm{R}_1-\tilde{\epsilon}_4}{\mathrm{r}_1}\\ \widetilde{\mathrm{r}}_1&=\mathrm{r}_1\\ \widetilde{\nu}_2&=\nu_2+\nu_1-\eta_1\\ \widetilde{\mathrm{r}}_2&=\tfrac{\mathrm{r}_2\nu_2+(\nu_1-\eta_1)\mathrm{r}_1}{\widetilde{\nu}_2}.\end{aligned}$$

Note that $(\eta_1,\widetilde{\nu}_2,\nu_3^{\ell+1},\widetilde{\mathrm{r}}_1,\widetilde{\mathrm{r}}_2,\mathrm{r}_3^{\ell+1})$ satisfies (158b), (158c), and (159) by construction. Furthermore, as a result of concavity of $J(\cdot)$, we have

$$\nu_1 J(\mathrm{r}_1)+\nu_2 J(\mathrm{r}_2)\le\eta_1 J(\widetilde{\mathrm{r}}_1)+\widetilde{\nu}_2 J(\widetilde{\mathrm{r}}_2).$$

Thus, $(\eta_1,\widetilde{\nu}_2,\nu_3^{\ell+1},\widetilde{\mathrm{r}}_1,\widetilde{\mathrm{r}}_2,\mathrm{r}_3^{\ell+1})$ also satisfies (158a).

For $j\ge 2$, we use $\widetilde{\nu}_j$ and $\widetilde{\mathrm{r}}_j$ to define η_j, $\widetilde{\nu}_{j+1}$ and $\widetilde{\mathrm{r}}_{j+1}$ as follows:

$$\eta_j=\tfrac{(1-\tilde{\epsilon}_3)\mathrm{R}_j}{\widetilde{\mathrm{r}}_j} \tag{161a}$$

$$\widetilde{\nu}_{j+1}=\nu_{j+1}+\widetilde{\nu}_j-\eta_j \tag{161b}$$

$$\widetilde{\mathrm{r}}_{j+1}=\tfrac{\mathrm{r}_{j+1}\nu_{j+1}+(\widetilde{\nu}_j-\eta_j)\widetilde{\mathrm{r}}_j}{\widetilde{\nu}_{j+1}}. \tag{161c}$$

Using the fact that $(\eta_1^{j-1},\widetilde{\nu}_j,\nu_{j+1}^{\ell+1},\widetilde{\mathrm{r}}_1^{\,j},\mathrm{r}_{j+1}^{\ell+1})$ satisfies (158) and (159) and the concavity of $J(\cdot)$, we can show that $(\eta_1^{j},\widetilde{\nu}_{j+1},\nu_{j+2}^{\ell+1},\widetilde{\mathrm{r}}_1^{\,j+1},\mathrm{r}_{j+2}^{\ell+1})$ also satisfies (158) and (159). We repeat the iteration given in (161) until we reach $\widetilde{\nu}_{\ell+1}$ and $\widetilde{\mathrm{r}}_{\ell+1}$ and we let $\eta_{\ell+1}=\widetilde{\nu}_{\ell+1}$.

Then, we conclude that for any variable-length block code satisfying the hypothesis of the Lemma 7 and for any positive δ such that $P_{\mathrm{e}}+\delta+\frac{P_{\mathrm{e}}}{\delta}\le\frac{1}{5}$

$$(1-\tilde{\epsilon}_3)\mathrm{E}_i-\tilde{\epsilon}_5\le\sum_{j=i+1}^{\ell+1}\eta_j J(\widetilde{\mathrm{r}}_j)\quad i=1,2,\ldots,\ell \tag{162a}$$

$$(1-\tilde{\epsilon}_3)\mathrm{R}_i-\tilde{\epsilon}_4\mathbb{1}_{\{i=1\}}=\widetilde{\mathrm{r}}_i\eta_i \qquad i=1,2,\ldots,\ell \tag{162b}$$

$$(\tilde{\epsilon}_3-P_{\mathrm{e}})\mathrm{R}+\tfrac{h(\tilde{\epsilon}_3)-h(P_{\mathrm{e}})}{\mathbf{E}[\mathsf{T}]}\le\widetilde{\mathrm{r}}_{\ell+1}\eta_{\ell+1} \tag{162c}$$

for some $(\eta_1,\ldots,\eta_{\ell+1},\widetilde{\mathrm{r}}_1,\ldots,\widetilde{\mathrm{r}}_{\ell+1})$ such that[35]

$$\widetilde{\mathrm{r}}_i\in[0,C]\quad i=1,2,\ldots,(\ell+1) \tag{163a}$$

$$\eta_i\ge 0\qquad i=1,2,\ldots,(\ell+1) \tag{163b}$$

$$\sum_{i=1}^{\ell+1}\eta_i=1. \tag{163c}$$

Lemma 7 follows from the fact that $J(\cdot)\le D$. ■

[35]One can replace the inequality in (162c) by equality because $J(\cdot)$ is a decreasing function.

H. Codes with Infinite Expected Decoding Time on Channels with Positive Transition Probabilities

In this section, we consider variable-length block codes on DMCs with positive transition probabilities, i.e., $\min_{\mathsf{x}\in\mathcal{X},\mathsf{y}\in\mathcal{Y}}W_{\mathsf{x}}(\mathsf{y})>0$, and derive lower bounds to the probabilities of various error events. These bounds, i.e., (166), (172), and (175), enable us to argue that Lemmas 6 and 7 hold for variable-length block codes with infinite expected decoding time, i.e., $\mathbf{E}[\mathsf{T}]=\infty$.

1) $P_{\mathrm{e}}>0$: On DMC such that $\min_{\mathsf{x}\in\mathcal{X},\mathsf{y}\in\mathcal{Y}}W_{\mathsf{x}}(\mathsf{y})=\lambda$, the posterior probability of any message $\mathsf{m}\in\mathcal{M}$ at time τ is lower bounded as

$$\mathbf{P}[\mathsf{M}=\mathsf{m}|\,\mathsf{Y}^{\tau}]\ge\left(\tfrac{\lambda}{1-\lambda}\right)^{\tau}\tfrac{1}{|\mathcal{M}|}.$$

Then, conditioned on the event $\{\mathsf{T}=\tau\}$, the probability of erroneous decoding is lower bounded as

$$\mathbf{P}\left[\widehat{\mathsf{M}}\ne\mathsf{M}\,\middle|\,\mathsf{T}=\tau\right]\ge\tfrac{|\mathcal{M}|-1}{|\mathcal{M}|}\left(\tfrac{\lambda}{1-\lambda}\right)^{\tau}. \tag{164}$$

Note that since $\mathbf{P}[\mathsf{T}<\infty]=1$, the error probability of any variable-length code satisfies

$$P_{\mathrm{e}}=\sum_{\tau=1}^{\infty}\mathbf{P}\left[\mathsf{M}\ne\widehat{\mathsf{M}}\,\middle|\,\mathsf{T}=\tau\right]\mathbf{P}[\mathsf{T}=\tau]. \tag{165}$$

Using (164) and (165), we get

$$P_{\mathrm{e}}\ge\tfrac{|\mathcal{M}|-1}{|\mathcal{M}|}\mathbf{E}\left[\left(\tfrac{\lambda}{1-\lambda}\right)^{\mathsf{T}}\right]. \tag{166}$$

Note that (166) implies that for a variable-length code with infinite expected decoding time, not only the rate R but also the error exponent E is zero.

2) If $P_{\mathrm{e}}+\frac{1}{|\mathcal{M}|}<1$, then $\min_{\mathsf{m}}P_{\mathrm{e}|\mathsf{m}}>0$: Note that since $\mathbf{P}[\mathsf{T}<\infty]=1$ and $|\mathcal{M}|<\infty$

$$\mathbf{P}[\mathsf{T}<\infty|\,\mathsf{M}=\mathsf{m}]=1\quad\forall\mathsf{m}\in\mathcal{M}.$$

For any variable-length block code such that $P_{\mathrm{e}}+\frac{1}{|\mathcal{M}|}<1$, let τ^* be

$$\tau^*=\min\left\{\tau:\max_{\mathsf{m}\in\mathcal{M}}\mathbf{P}[\mathsf{T}>\tau|\,\mathsf{M}=\mathsf{m}]\le\tfrac{|\mathcal{M}|-1}{|\mathcal{M}|}-P_{\mathrm{e}}\right\}. \tag{167}$$

Since $\mathbf{P}[\mathsf{T}<\infty|\,\mathsf{M}=\mathsf{m}]=1$ for all m in $\mathcal{M}$ and $\mathcal{M}$ is finite, τ^* is finite.

Note that for any τ, m, and $\widetilde{\mathsf{m}}$ we have

$$\mathbf{P}[\mathsf{Y}^{\tau}=\mathsf{y}^{\tau}|\,\mathsf{M}=\mathsf{m}]\ge(\tfrac{\lambda}{1-\lambda})^{\tau}\mathbf{P}[\mathsf{Y}^{\tau}=\mathsf{y}^{\tau}|\,\mathsf{M}=\widetilde{\mathsf{m}}]. \tag{168}$$

Then, using (168), we get

$$\begin{aligned}P_{\mathrm{e}|\mathsf{m}}&\ge\sum_{\widetilde{\mathsf{m}}\ne\mathsf{m}}\mathbf{P}\left[\left\{\widehat{\mathsf{M}}=\widetilde{\mathsf{m}},\mathsf{T}\le\tau^*\right\}\middle|\,\mathsf{M}=\mathsf{m}\right]\\ &=(\tfrac{\lambda}{1-\lambda})^{\tau^*}\sum_{\widetilde{\mathsf{m}}\ne\mathsf{m}}\mathbf{P}\left[\left\{\widehat{\mathsf{M}}=\widetilde{\mathsf{m}},\mathsf{T}\le\tau^*\right\}\middle|\,\mathsf{M}=\widetilde{\mathsf{m}}\right]\\ &\ge(\tfrac{\lambda}{1-\lambda})^{\tau^*}\sum_{\widetilde{\mathsf{m}}\ne\mathsf{m}}\left(\mathbf{P}\left[\widehat{\mathsf{M}}=\widetilde{\mathsf{m}}\,\middle|\,\mathsf{M}=\widetilde{\mathsf{m}}\right]-\mathbf{P}[\mathsf{T}>\tau^*|\,\mathsf{M}=\widetilde{\mathsf{m}}]\right).\end{aligned} \tag{169}$$

Note that as a result of (167), we have

$$\mathbf{P}[\mathsf{T} > \tau^* | \, \mathsf{M} = \widetilde{\mathsf{m}}] \leq \left(\tfrac{|\mathcal{M}|-1}{|\mathcal{M}|} - P_{\mathrm{e}} \right) \quad \forall \widetilde{\mathsf{m}} \in \mathcal{M}. \tag{170}$$

Furthermore

$$\sum_{\widetilde{\mathsf{m}} \neq \mathsf{m}} \mathbf{P}\Big[\widehat{\mathsf{M}} = \widetilde{\mathsf{m}} \Big| \, \mathsf{M} = \widetilde{\mathsf{m}} \Big] \geq |\mathcal{M}|(1 - P_{\mathrm{e}}) - 1. \tag{171}$$

Thus, using (169)–(171), we get

$$\min_{\mathsf{m} \in \mathcal{M}} P_{\mathrm{e}|\mathsf{m}} \geq (\tfrac{\lambda}{1-\lambda})^{\tau^*} \left(1 - \tfrac{1}{|\mathcal{M}|} - P_{\mathrm{e}} \right) \tag{172}$$

where τ^* is a finite integer defined in (167).

3) For all $i \in \{1, 2, \ldots, \ell\}$, $P_{\mathrm{e}}(i) > 0$: For a variable-length block code with message set $\mathcal{M}$ of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k$ on a DMC such that $\min_{\mathsf{x} \in \mathcal{X}, \mathsf{y} \in \mathcal{Y}} W_{\mathsf{x}}(\mathsf{y}) = \lambda$, the posterior probability of any element of $\mathcal{M}_i$ at time τ is lower bounded as

$$\mathbf{P}\left[\mathsf{M}^i = \mathsf{m}^i \big| \, \mathsf{Y}^\tau \right] \geq \left(\tfrac{\lambda}{1-\lambda} \right)^{\tau} \tfrac{1}{|\mathcal{M}_i|} \quad \forall \mathsf{m}^i \in \mathcal{M}^i, \ \forall i \, \{1, 2, \ldots, \ell\}.$$

Then, conditioned on the event $\{\mathsf{T} = \tau\}$, the probability of decoding any one of the first i sub-messages erroneously is lower bounded as

$$\mathbf{P}\Big[\widehat{\mathsf{M}}^i \neq \mathsf{M}^i \Big| \, \mathsf{T} = \tau \Big] \geq \tfrac{|\mathcal{M}^i|-1}{|\mathcal{M}^i|} \left(\tfrac{\lambda}{1-\lambda} \right)^{\tau}. \tag{173}$$

Since $\mathbf{P}[\mathsf{T} < \infty] = 1$, $P_{\mathrm{e}}(i)$ satisfies

$$P_{\mathrm{e}} = \sum_{\tau=1}^{\infty} \mathbf{P}\Big[\mathsf{M}^i \neq \widehat{\mathsf{M}}^i \Big| \, \mathsf{T} = \tau \Big] \mathbf{P}[\mathsf{T} = \tau]. \tag{174}$$

Using (173) and (174), we get

$$P_{\mathrm{e}}(i) \geq \tfrac{|\mathcal{M}^i|-1}{|\mathcal{M}^i|} \mathbf{E}\left[\left(\tfrac{\lambda}{1-\lambda} \right)^{\mathsf{T}} \right] \quad \forall i \, \{1, 2, \ldots, \ell\}. \tag{175}$$

Equation (175) implies that for a variable-length code with infinite expected decoding time, not only the rates but also the error exponents of submessages are zero.

I. Proof of Theorem 1

Proof: In Section IV-C, it is shown that for any rate $\mathrm{R} \in [0, C]$, error exponent $\mathrm{E} \in [0, (1 - \frac{\mathrm{R}}{C})D]$, there exists a reliable sequence $\mathbb{Q}$ such that $\mathrm{R}_{\mathbb{Q}} = \mathrm{R}$, $\mathrm{E}_{\mathbb{Q}} = \mathrm{E}$, $\mathrm{E}_{\mathrm{md},\mathbb{Q}} = \mathrm{E} + (1 - \frac{\mathrm{E}}{D}) J\Big(\frac{\mathrm{R}}{1 - \mathrm{E}/D} \Big)$. Thus, as a result of the definition of $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ given in (13), we have

$$\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E}) \geq \mathrm{E} + \left(1 - \tfrac{\mathrm{E}}{D}\right) J\Big(\tfrac{\mathrm{R}}{1 - \mathrm{E}/D} \Big). \tag{176}$$

In Section V-C, we have shown that any reliable sequence of codes $\mathbb{Q}$ with rate $\mathrm{R}_{\mathbb{Q}}$ and error exponent $\mathrm{E}_{\mathbb{Q}}$ satisfies

$$\mathrm{E}_{\mathrm{md},\mathbb{Q}} \leq \mathrm{E}_{\mathbb{Q}} + (1 - \tfrac{\mathrm{E}_{\mathbb{Q}}}{D}) J\Big(\tfrac{\mathrm{R}_{\mathbb{Q}}}{1 - \mathrm{E}_{\mathbb{Q}}/D} \Big).$$

Thus, using the fact that $J(\cdot)$ is a decreasing concave function, we can conclude that

$$\max_{\substack{\mathbb{Q}: \mathrm{R}_{\mathbb{Q}} \geq \mathrm{R} \\ \mathrm{E}_{\mathbb{Q}} \geq \mathrm{E}}} \mathrm{E}_{\mathrm{md},\mathbb{Q}} \leq \mathrm{E} + (1 - \tfrac{\mathrm{E}}{D}) J\Big(\tfrac{\mathrm{R}}{1 - \mathrm{E}/D} \Big).$$

Consequently, as a result of the definition of $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ given in (13), we have

$$\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E}) \leq \mathrm{E} + \left(1 - \tfrac{\mathrm{E}}{D}\right) J\Big(\tfrac{\mathrm{R}}{1 - \mathrm{E}/D} \Big). \tag{177}$$

Thus, using (176) and (177), we can conclude that

$$\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E}) = \mathrm{E} + \left(1 - \tfrac{\mathrm{E}}{D}\right) J\Big(\tfrac{\mathrm{R}}{1 - \mathrm{E}/D} \Big). \tag{178}$$

In order to prove the concavity of $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ in (R, E) pair, let $(\mathrm{R}_a, \mathrm{E}_a)$ and $(\mathrm{R}_b, \mathrm{E}_b)$ be two pairs such that

$$\mathrm{R}_a \in [0, C] \quad \mathrm{E}_a \leq (1 - \tfrac{\mathrm{R}_a}{C})D \tag{179a}$$

$$\mathrm{R}_b \in [0, C] \quad \mathrm{E}_b \leq (1 - \tfrac{\mathrm{R}_b}{C})D. \tag{179b}$$

Then, for any $\alpha \in [0, 1]$, let R_α and E_α be

$$\mathrm{R}_\alpha = \alpha \mathrm{R}_a + (1 - \alpha)\mathrm{R}_b \tag{180a}$$

$$\mathrm{E}_\alpha = \alpha \mathrm{E}_a + (1 - \alpha)\mathrm{E}_b. \tag{180b}$$

From (179) and (180), we have

$$\mathrm{R}_\alpha \in [0, C] \quad \mathrm{E}_\alpha \leq (1 - \tfrac{\mathrm{R}_\alpha}{C}). \tag{181}$$

Furthermore, using the concavity of $J(\cdot)$, we get

$$\begin{aligned}
&\alpha \mathrm{E}_{\mathrm{md}}(\mathrm{R}_a, \mathrm{E}_a) + (1 - \alpha)\mathrm{E}_{\mathrm{md}}(\mathrm{R}_b, \mathrm{E}_b) \\
&\quad = \alpha \left(\mathrm{E}_a + \left(1 - \tfrac{\mathrm{E}_a}{D}\right) J\Big(\tfrac{\mathrm{R}_a}{1 - \mathrm{E}_a/D} \Big) \right) \\
&\qquad + (1 - \alpha) \left(\mathrm{E}_b + \left(1 - \tfrac{\mathrm{E}_b}{D}\right) J\Big(\tfrac{\mathrm{R}_b}{1 - \mathrm{E}_b/D} \Big) \right) \\
&\quad = \mathrm{E}_\alpha + \alpha \left(1 - \tfrac{\mathrm{E}_a}{D}\right) J\Big(\tfrac{\mathrm{R}_a}{1 - \mathrm{E}_a/D} \Big) \\
&\qquad + (1 - \alpha) \left(1 - \tfrac{\mathrm{E}_b}{D}\right) J\Big(\tfrac{\mathrm{R}_b}{1 - \mathrm{E}_b/D} \Big) \\
&\quad \leq \mathrm{E}_\alpha + \left(1 - \tfrac{\mathrm{E}_\alpha}{D}\right) J\Big(\tfrac{\alpha \mathrm{R}_a + (1-\alpha)\mathrm{R}_b}{1 - \mathrm{E}_\alpha/D} \Big) \\
&\quad = \mathrm{E}_{\mathrm{md}}(\mathrm{R}_\alpha, \mathrm{E}_\alpha).
\end{aligned} \tag{182}$$

Thus, $\mathrm{E}_{\mathrm{md}}(\mathrm{R}, \mathrm{E})$ is jointly concave in rate exponent pairs. ■

J. Proof of Theorem 2

Proof: In Section IV-D, it is shown that for any positive integer ℓ, a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ is achievable if there exists a time-sharing vector $\vec{\eta}$ such that

$$\mathrm{E}_i \leq (1 - \sum_{j=1}^{\ell} \eta_j)D + \sum_{j=i+1}^{\ell} \eta_j J\Big(\tfrac{\mathrm{R}_j}{\eta_j} \Big) \qquad \forall i \in \{1, 2, \ldots, \ell\} \tag{183a}$$

$$\mathrm{R}_i \leq C\eta_i \qquad \forall i \in \{1, 2, \ldots, \ell\} \tag{183b}$$

$$\eta_i \geq 0 \qquad \forall i \in \{1, 2, \ldots, \ell\} \tag{183c}$$

$$\sum_{j=1}^{\ell} \eta_j \leq 1. \tag{183d}$$

Thus, the existence of a time-sharing vector $\vec{\eta}$ satisfying (183) is a sufficient condition for the achievability of a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$.

For any reliable code sequence $\mathbb{Q}$ whose message sets are of the form $\mathcal{M}^{(\kappa)} = \mathcal{M}_1^{(\kappa)} \times \mathcal{M}_2^{(\kappa)} \times \cdots \times \mathcal{M}_\ell^{(\kappa)}$, Lemma 7 with $\delta = \frac{-1}{\ln P_{\mathbf{e}}}$ implies that there exists a sequence $\vec{\eta}_\kappa$ such that

$$(1-\tilde{\epsilon}_{3,\kappa})\mathrm{E}_{i,\kappa} - \tilde{\epsilon}_{5,\kappa} \le (1-\sum_{j=1}^{\ell}\eta_{j,\kappa})D + \sum_{j=i+1}^{\ell}\eta_{j,\kappa} J\left(\frac{(1-\tilde{\epsilon}_{3,\kappa})\mathrm{R}_{j,\kappa}}{\eta_{j,\kappa}}\right) \quad i = 1,2,\ldots,\ell \qquad (184a)$$

$$(1-\tilde{\epsilon}_{3,\kappa})\mathrm{R}_{i,\kappa} - \tilde{\epsilon}_{4,\kappa}\mathbb{1}_{\{i=1\}} \le C\eta_{i,\kappa} \quad i = 1,2,\ldots,\ell \qquad (184b)$$

$$\eta_{i,\kappa} \ge 0 \qquad i = 1,2,\ldots,\ell \qquad (184c)$$

$$\sum_{j=1}^{\ell}\eta_{,\kappa} \le 1 \qquad (184d)$$

where $\mathrm{R}_{i,\kappa} = \frac{|\ln \mathcal{M}_i^{(\kappa)}|}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}$, $\mathrm{E}_{i,\kappa} = \frac{-\ln P_{\mathbf{e}}(i)^{(\kappa)}}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}$, $\tilde{\epsilon}_{3,\kappa} = \frac{P_{\mathbf{e}}{}^{(\kappa)} + 1 - P_{\mathbf{e}}{}^{(\kappa)}}{-\ln P_{\mathbf{e}}{}^{(\kappa)}} \ln P_{\mathbf{e}}{}^{(\kappa)}$, $\tilde{\epsilon}_{4,\kappa} = \frac{h(\tilde{\epsilon}_{3,\kappa})}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}$, and $\tilde{\epsilon}_{5,\kappa} = \frac{h(\tilde{\epsilon}_{3,\kappa}) - \ln \lambda\delta}{\mathbf{E}^{(\kappa)}\left[\mathsf{T}^{(\kappa)}\right]}$.

Note that as a result of (184), all members of the sequence $\vec{\eta}_\kappa$ are from a compact metric space.[36] Thus, there exists a convergent subsequence, converging to a $\vec{\eta}$. Using (184), with definitions of $\mathrm{R}_{\mathbb{Q},i}$ and $\mathrm{E}_{\mathbb{Q},i}$ given in Definition 11, we can conclude that $\vec{\eta}$ satisfies

$$\mathrm{E}_{\mathbb{Q},i} \le (1-\sum_{j=1}^{\ell}\eta_j)\, D + \sum_{j=i+1}^{\ell}\eta_j J\left(\frac{\mathrm{R}_{\mathbb{Q},j}}{\eta_j}\right) \quad \forall i \in \{1,2,\ldots,\ell\} \qquad (185a)$$

$$\mathrm{R}_{\mathbb{Q},i} \le C\eta_i \quad \forall i \in \{1,2,\ldots,\ell\} \qquad (185b)$$

$$\eta_i \ge 0 \quad \forall i \in \{1,2,\ldots,\ell\} \qquad (185c)$$

$$\sum_{j=1}^{\ell}\eta_j \le 1. \qquad (185d)$$

According to Definition 11 describing the *bit-wise* UEP problem, a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ is achievable only if there exists a reliable code sequence $\mathbb{Q}$ such that $(\vec{\mathrm{R}}_{\mathbb{Q}}, \vec{\mathrm{E}}_{\mathbb{Q}}) = (\vec{\mathrm{R}}, \vec{\mathrm{E}})$. Consequently, the existence of a time-sharing vector satisfying (183) is also a necessary condition for the achievability of a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$

Thus, we can conclude that a rate-exponent vector $(\vec{\mathrm{R}}, \vec{\mathrm{E}})$ is achievable if and only if there exists a $\vec{\eta}$ satisfying (183).

In order to prove the convexity of region of achievable rate-exponent vectors, let $(\vec{\mathrm{R}}_a, \vec{\mathrm{E}}_a)$ and $(\vec{\mathrm{R}}_b, \vec{\mathrm{E}}_b)$ be two achievable rate-exponent vectors. Then, there exist triples $(\vec{\mathrm{R}}_a, \vec{\mathrm{E}}_a, \vec{\eta}_a)$ and $(\vec{\mathrm{R}}_b, \vec{\mathrm{E}}_b, \vec{\eta}_b)$ satisfying (183).

For any $\alpha \in [0,1]$, let $\vec{\mathrm{R}}_\alpha$, $\vec{\mathrm{E}}_\alpha$, and $\vec{\eta}_\alpha$ be

$$\vec{\mathrm{R}}_\alpha = \alpha\vec{\mathrm{R}}_a + (1-\alpha)\vec{\mathrm{R}}_b$$
$$\vec{\mathrm{E}}_\alpha = \alpha\vec{\mathrm{E}}_a + (1-\alpha)\vec{\mathrm{E}}_b$$
$$\vec{\eta}_\alpha = \alpha\vec{\eta}_a + (1-\alpha)\vec{\eta}_b.$$

[36]Let the metric be $\|\vec{\eta} - \vec{\nu}\| = \max_j |\eta_j - \nu_j|$.

As $J(\cdot)$ is concave and the triples $(\vec{\mathrm{R}}_a, \vec{\mathrm{E}}_a, \vec{\eta}_a)$ and $(\vec{\mathrm{R}}_b, \vec{\mathrm{E}}_b, \vec{\eta}_b)$ satisfy the constraints given in (183), the triple $(\vec{\mathrm{R}}_\alpha, \vec{\mathrm{E}}_\alpha, \vec{\eta}_\alpha)$ also satisfies the constraints given in (183). Consequently, the rate-exponent vector $(\vec{\mathrm{R}}_\alpha, \vec{\mathrm{E}}_\alpha)$ is achievable and the region of achievable rate-exponent vectors is convex. ■

ACKNOWLEDGMENT

Authors would like to thank Gerhard Kramer and Reviewer II for pointing out the importance of the case of varying the number of groups of bits with expected block length. Reviewer II has also coined the term rates-exponents vector which replaced authors' previous attempt: rate vector error exponent vector pair. Authors would also like to thank all of the reviewers for their meticulous reviews and penetrating questions.

REFERENCES

[1] P. Berlin, B. Nakiboğlu, 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.

[2] 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, Dec. 2009.

[3] M. V. Burnashev, "Data transmission over a discrete channel with feedback, random transmission time," *Problemy Perdachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1976.

[4] F. Chung and L. Lu, "Concentration inequalities and martingale inequalities: A survey," *Internet Math.*, vol. 3, no. 1, pp. 79–127, 2006.

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

[6] S. K. Gorantla, B. Nakiboğlu, T. P. Coleman, and L. Zheng, "Bit-wise unequal error protection for variable length blockcodes with feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 241–245.

[7] B. D. Kudryashov, "On message transmission over a discrete channel with noiseless feedback," *Problemy Perdachi Informatsii*, vol. 15, no. 1, pp. 3–13, 1973.

[8] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 24–49, Jan. 2012.

[9] B. Nazer, Y. Shkel, and S. C. Draper, The AWGN red alert problem 2012, arXiv:1102.4411[cs.IT]. DOI:10.1109/TIT.2012.2235120..

[10] A. N. Shiriaev, *Probability*. New York: Springer-Verlag, 1996.

[11] D. Wang, V. Chandar, S. Y. Chung, and G. W. Wornell, "On reliability functions for single-message unequal error protection," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2934–2938.

[12] H. Yamamoto and K. Itoh, "Asymptotic performance of a modified Schalkwijk–Barron scheme for channels with noiseless feedback," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 729–733, Nov. 1979.

Barış Nakiboğlu received the B.S. degrees in electrical and electronics engineering and in physics from Middle East Technical University (METU), Ankara, Turkey, in 2002 and the M.S. degree in electrical engineering and computer science and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, in 2005 and 2011, respectively. He is now a postdoctoral researcher at University of California Berkeley, working with Anant Sahai. He is also a research affiliate at Laboratory for Information and Decision Systems (LIDS) at MIT working with Sanjoy Mitter.

Siva K. Gorantla received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, in 2007. He received the M.S. and Ph.D. degrees in electrical and computer engineering and M.S. in Statistics degree from the University of Illinois at Urbana-Champaign, Urbana, in 2009, 2012 and 2012 respectively. Since June 2012, he has been working as a Research Scientist at Adchemy Inc, CA. His research interests include stochastic control, information theory, statistical learning and prediction.

Mr. Gorantla has been awarded the James Henderson fellowship at UIUC and S Subramanian Award (Institute Merit Prize) from IIT Madras in the years 2008 and 2004.

Lizhong Zheng 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, where he is currently an associate professor. His research interests include information theory, wireless communications and wireless networks. He received Eli Jury award from UC Berkeley in 2002, IEEE Information Theory Society Paper Award in 2003, and NSF CAREER award in 2004, and the AFOSR Young Investigator Award in 2007.

Todd P. Coleman (S'01–M'05–SM'11) received the B.S. degrees in electrical engineering (summa cum laude) as well as computer engineering (summa cum laude) from the University of Michigan, Ann Arbor, in 2000. He received the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002, and 2005. He was a postdoctoral scholar in neuroscience at MIT during the 2005–2006 academic year. He was an Assistant Professor in ECE and Neuroscience at the University of Illinois from 2006–2011.

Dr. Coleman is currently an Associate Professor in Bioengineering and director of the Neural Interaction Laboratory at UCSD, where his group builds flexible bio-electronics for neurological monitoring applications. His research is highly inter-disciplinary and lies at the intersection of bio-electronics, neuroscience, medicine, and applied mathematics. Dr. Coleman is a science advisor for the Science & Entertainment Exchange (National Academy of Sciences).