Errors-and-Erasures Decoding for Block Codes With Feedback

Barış Nakiboğlu and Lizhong Zheng

***Abstract*—Inner and outer bounds are derived on the optimal performance of fixed-length block codes on discrete memoryless channels with feedback and errors-and-erasures decoding. First, an inner bound is derived using a two-phase encoding scheme with communication and control phases together with the optimal decoding rule for the given encoding scheme, among decoding rules that can be represented in terms of pairwise comparisons between the messages. Then, an outer bound is derived using a generalization of the straight-line bound to errors-and-erasures decoders and the optimal error-exponent tradeoff of a feedback encoder with two messages. In addition, upper and lower bounds are derived, for the optimal erasure exponent of error-free block codes in terms of the rate. Finally, a proof is provided for the fact that the optimal tradeoff between error exponents of a two-message code does not improve with feedback on discrete memoryless channels (DMCs).**

***Index Terms*—Decision feedback, discrete memoryless channels (DMCs), error exponent, errors-and-erasures decoding, feedback, feedback encoding schemes, soft decoding, two-phase encoding schemes, variable-length coding.**

I. Introduction

SHANNON showed in [29] that the capacity of discrete memoryless channels (DMCs) does not increase even when a noiseless and delay-free feedback link is available from the receiver to the transmitter. On symmetric DMCs, the sphere packing exponent bounds the error exponent of fixed-length block codes from above, as shown by Dobrushin[1] in [11]. Thus, relaxations like errors-and-erasures decoding or variable-length coding are needed for feedback to increase the error exponent of block codes at rates larger than the critical rate on symmetric DMCs. In this work, we investigate one such relaxation, namely, errors-and-erasures decoding, and find inner and outer bounds to the optimal error-exponent/erasure-exponent tradeoff.

Finding the optimal encoding and decoding schemes, and hence finding optimal performance by characterizing the surface of achievable error-exponent/erasure-exponent pairs, is an important motivation for the investigation of errors-and-erasures decoding. Note, however, that finding the optimal performance with erasures will implicitly solve the problem of finding the optimal feedback encoder and determining the error exponent for the erasure-free fixed-length block codes with feedback which is a long standing open problem. Finding the optimal performance, however, is far from being the only important aspect of the problem. Determining the performance of feedback encoding schemes that are easier to implement, more robust to the degradations of the feedback link, and bounding the loss in the performance compared to the more complicated encoding schemes are both important tasks practically and interesting ones intellectually. This will be our aim in this paper. We will first analyze the performance of a two-phase encoding scheme inspired by the optimal encoding schemes for variable-length block codes and derive inner bounds to the optimal performance. Then, we will derive outer bounds to the performance of general feedback encoding schemes with erasures and quantify the loss of performance by restricting ourselves to the above mentioned two-phase schemes. This analysis complements the research on two related block coding schemes: variable-length block coding and errors-and-erasures decoding for block codes without feedback. We start with a very brief overview of the previous work on these problems to motivate our investigation further.

Burnashev [3]–[5] was the first one to consider variable-length block codes with feedback, instead of fixed-length ones. He obtained the exact expression for the error exponent at all rates. Later, Yamamoto and Itoh [33] suggested a coding scheme which achieves the best error exponent for variable-length block codes with feedback by using a fixed-length block code with an errors-and-erasures decoding and repeating the same codeword until a nonerasure decoding occurs.[2] In fact, any fixed-length block code with erasures can be used in this repetitive fashion, like it was done in [33], to get a variable-length block code with essentially the same error exponent as the original fixed-length block code. Thus, [3] can be reinterpreted to give an upper bound to the error exponent achievable by fixed-length block codes with erasures. Furthermore, this upper bound is achieved by the fixed-length block codes with erasures described in [33], when erasure probability is decaying to zero subexponentially with block length. However, the techniques used in this line of work are insufficient for deriving proper inner or outer bounds for the situation when erasure probability is decaying exponentially with block length. As explained in the following paragraph, the

Manuscript received March 25, 2009; revised September 01, 2011; accepted September 02, 2011. Date of current version January 06, 2012.

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

Communicated by H.-A. Loeliger, Associate Editor for Coding Techniques.

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.2011.2169529

[1]Later, Haroutunian [17] established an upper bound on the error exponent of block codes with feedback. This upper bound is equal to the sphere packing exponent for symmetric channels but it is strictly larger than the sphere packing exponent for nonsymmetric channels.

[2]Including erasures will not increase the exponent for variable-length block codes with feedback.

case with strictly positive erasure exponent is important both for engineering applications and for a better understanding of soft decoding with feedback. Our investigation provides proper tools for such an analysis, results in inner and outer bounds to the tradeoff between error-and-erasure exponents, and recovers all previously known results for the zero erasure-exponent case.

When considered together with higher layers, the codes in the physical layer are part of a variable-length/delay communication scheme with feedback. However, in the physical layer itself, fixed-length block codes are used instead of variable-length ones because of their amenability to modular design and robustness against the noise in the feedback link. In such an architecture, retransmissions affect the performance of higher layers. The average transmission time is only a first-order measure of this effect: as long as the erasure probability is vanishing with increasing block length, average transmission time will essentially be equal to the block length of the fixed-length block code. Thus, with an analysis like the one in [33], the cost of retransmissions is ignored as long as the erasure probability goes to zero with increasing block length. In a communication system with multiple layers, however, retransmissions usually have costs beyond their effect on average transmission time, which are described by constraints on the probability distribution of the decoding time. Knowledge of error erasure-exponent tradeoff is useful in coming up with designs to meet those constraints. An example of this phenomenon is variable-length block coding schemes with hard deadlines for decoding time, which has already been investigated by Gopala *et al.* [16] for block codes without feedback. They have used a block coding scheme with erasures and resent the message whenever an erasure occurred. But because of the hard deadline, they employed this scheme only for some fixed number of trials. If all those trials failed, i.e., led to an erasure, they used a nonerasure block code. Using the error-exponent/erasure-exponent tradeoff they were able to obtain the best over all error performance for the given architecture.

This brings us to the second line of research we complement with our investigation: errors-and-erasures decoding for block codes without feedback. Forney [14] was the first one to consider errors-and-erasures decoding without feedback. He obtained an achievable tradeoff between the exponents of error-and-erasure probabilities. Then, Csiszár and Körner [10] achieved the same performance using universal coding and decoding algorithms. Later, Telatar and Gallager [32] introduced a strict improvement on certain channels over the results presented in [14] and [10]. Recently, there has been a revived interest in the errors-and-erasures decoding for universally achievable performances [21], [22], for alternative methods of analysis [20], for extensions to the channels with side information [26], and implementation with linear block codes [18]. The encoding schemes in these codes do not have access to any feedback. However, if the transmitter can learn whether the decoded message was an erasure, it can resend the message whenever it is erased. Because of this block retransmission variant, these problems are sometimes called decision feedback problems.

We complement the results on the error-exponent/erasure-exponent tradeoff without feedback and the results about error exponent of variable-length block codes with feedback, by finding inner and outer bounds to the error-exponent/erasure-exponent tradeoff of fixed-length block codes with feedback. We first introduce our model and notation in Section II. Then, in Section III, we derive a lower bound using a two-phase coding algorithm similar to the one described by Yamamoto and Ito [33] and decoding rule and analysis techniques, inspired by Telatar [31] for the nonfeedback case. Note that the analysis and the decoding rule in [31] are tailored for a single-phase scheme and without feedback and the two-phase scheme of [33] is tuned specifically to zero-erasure exponent; coming up with framework in which both of the ideas can be used efficiently is the main technical challenge here. In Section IV, we first extend the straight-line bound idea introduced by Shannon *et al.* [30] to block codes with erasures. Then, we use it together with the outer bound on the error-exponent tradeoff between two codewords with feedback to establish an outer bound for the error exponent of fixed-length block codes with feedback and erasures. In Section V, we first introduce error-free block codes with erasures and discuss their relation to the fixed-length block codes with errors-and-erasures decoding, and then we present inner and outer bounds to the erasure exponent of error-free block codes and point out its relation to the error-exponent/erasure-exponent tradeoff.

Before presenting our analysis, let us make a brief digression and discuss two channel models in which the use of feedback had been investigated for block codes without erasures. First channel model is the well-known additive white Gaussian noise channel (AWGNC) model. In AWGNCs, if the power constraint $\mathcal{P}$ is on the expected value of the energy spent on a block $\mathbf{E}[\mathcal{S}_{\mathsf{n}}]$, i.e., power constraint is of the form $\mathbf{E}[\mathcal{S}_{\mathsf{n}}] \leq \mathcal{P}\mathsf{n}$, the error probability can be made to decay faster than any exponential function with block-length n. Schalkwijk and Kailath suggested a coding algorithm [28], which achieves a doubly exponential decay in error probability for continuous time AWGNCs, i.e., infinite bandwidth case. Later, Schalkwijk [27] modified that scheme to achieve the same performance in discrete time AWGNCs, i.e., finite bandwidth case. Concatenating Schalkwijk and Kailath scheme with pulse amplitude modulation stages gives a multifold exponential decrease in the error probability [15], [25], [34]. However, this behavior relies on the absence of any amplitude limit, the particular form of the power constraint, and the noise-free nature of the feedback link. First, as observed in [5] and [24], when there is an amplitude limit, error probability decays only exponentially with block length. More importantly, if the power constraint restricts the energy spent in transmission of each message for all noise realizations, i.e., if the power constraint is an almost sure power constraint[3] of the form $\mathcal{S}_{\mathsf{n}} \leq \mathcal{P}\mathsf{n}$, then sphere packing exponent is still an upper bound to the error exponent for AWGNCs as shown by Pinsker [25]. Furthermore, if the feedback link is also an AWGNC and if there is a power constraint[4] on the feedback transmissions, then even in the case when there are only two messages, error probability decays only exponentially as it has been recently shown by Kim *et al.* [19].

[3]As Kim *et al.* [19] call it.

[4]This constraint can be an expected or almost sure constraint.

The second channel model is the DMC model. Although feedback cannot increase the error exponent for rates over the critical rate, it can simplify the encoding scheme [13], [34]. Furthermore, for rates below the critical rate, it is possible to improve the error exponent using feedback. Zigangirov [34] has established lower bounds to the error exponent for BSCs using a simple encoding scheme. Zigangirov's lower bound is equal to the sphere packing exponent for all rates in the interval $[R'_{\text{crit}}, \mathcal{C}]$ where $R'_{\text{crit}} < R_{\text{crit}}$ and Zigangirov's lower bound is strictly larger than the corresponding nonfeedback exponent for rates below R'_{crit}. Later, Burnashev [6] introduced an improvement to Zigangirov's bound for all positive rates less than R'_{crit}. D'yachkov [13] generalized Zigangirov's encoding scheme for general DMCs and established a lower bound to the error exponent for general binary input channels and k-ary symmetric channels. However, it is still an open problem to find a constructive technique that can be used for all DMCs which outperforms the random coding bound. Like AWGNCs, there has been a revived interest in the effect of a noisy feedback link and achievable performances with noisy feedback on DMCs. Burnashev and Yamamoto recently showed that error exponent of BSC channel increases even with a noisy feedback link [7], [8]. Furthermore, Draper and Sahai [12] investigated the use of noisy feedback link in variable-length schemes.

II. Model and Notation

The input and output alphabets of the forward channel are $\mathcal{X}$ and $\mathcal{Y}$, respectively. The channel input and output symbols at time t will be denoted by X_t and Y_t, respectively. Furthermore, the sequences of input and output symbols from time t_1 to time t_2 are denoted by $\mathsf{X}_{t_1}^{t_2}$ and $\mathsf{Y}_{t_1}^{t_2}$. When $t_1 = 1$, we omit t_1 and simply write X^{t_2} and Y^{t_2} instead of $\mathsf{X}_1^{t_2}$ and $\mathsf{Y}_1^{t_2}$. The forward channel is a stationary memoryless channel characterized by an $|\mathcal{X}|$-by-$|\mathcal{Y}|$ transition probability matrix W

$$\mathbf{P}\left[\mathsf{Y}_t | \mathsf{X}^t, \mathsf{Y}^{t-1}\right] = \mathbf{P}\left[\mathsf{Y}_t | \mathsf{X}_t\right] = W(\mathsf{Y}_t|\mathsf{X}_t) \qquad \forall t. \tag{1}$$

The feedback channel is noiseless and delay free, i.e., the input of the feedback channel Z_{t-1}, chosen at the receiver, is observed at the transmitter before transmission of X_t. In addition, we assume that feedback channel is of infinite capacity thus Z_{t-1} includes all of the observation of the receiver at time $t-1$, i.e.,[5] $\mathsf{Z}_{t-1} = (\mathsf{Y}_{t-1}, \mathsf{A}_{t-1})$. The random variables $\mathsf{A}_0, \mathsf{A}_1, \ldots, \mathsf{A}_\mathsf{n}$ are there to enable randomized encoding and decoding schemes as we will see shortly. It is assumed that the choice A's does not affect the forward channels behavior, i.e., in addition to (1), we have[6]

$$\mathbf{P}\left[\mathsf{Y}_t | \mathsf{X}^t, \mathsf{Z}^{t-1}\right] = W(\mathsf{Y}_t|\mathsf{X}_t) \qquad \forall t. \tag{2}$$

The message M is drawn from the message set $\mathcal{M}$ with a uniform probability distribution and is given to the transmitter at time zero. At each time $t \in [1, \mathsf{n}]$ the input symbol $X_t(\mathsf{M}, \mathsf{Z}^{t-1})$ is sent. The sequence of functions $X_t(\cdot) : \mathcal{M} \times \mathcal{Z}^{t-1}$ which assigns an input symbol for each $m \in \mathcal{M}$ and $z^{t-1} \in \mathcal{Z}^{t-1}$ is called the encoding function. Note that the random variables $\mathsf{A}_0, \mathsf{A}_1, \ldots, \mathsf{A}_{\mathsf{n}-1}$ enable randomized encoding schemes. After receiving Y^n the receiver draws the final A, i.e., A_n, and decodes to the message $\hat{\mathsf{M}}(\mathsf{Z}^\mathsf{n}) \in \{\mathbf{x}\} \cup \mathcal{M}$ where $\mathbf{x}$ is the erasure symbol. The random variable A_n does not have any effect on the encoding; it is used only to enable randomized decoding schemes.

The conditional error-and-erasure probabilities $P_{\mathrm{e}|\mathsf{M}}$ and $P_{\mathbf{x}|\mathsf{M}}$ and average error-and-erasure probabilities P_e and $P_\mathbf{x}$ are defined as

$$P_{\mathrm{e}|\mathsf{M}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathsf{M}\right] - P_{\mathbf{x}|\mathsf{M}} \qquad P_{\mathbf{x}|\mathsf{M}} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x} \middle| \mathsf{M}\right]$$
$$P_\mathrm{e} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] - P_\mathbf{x} \qquad P_\mathbf{x} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x}\right].$$

Since all the messages are equally likely, we have

$$P_\mathrm{e} = \frac{1}{|\mathcal{M}|}\sum_m P_{\mathrm{e}|m} \qquad P_\mathbf{x} = \frac{1}{|\mathcal{M}|}\sum_m P_{\mathbf{x}|m}.$$

We use a somewhat abstract but rigorous approach in defining the rate and achievable exponent pairs. A reliable sequence $\mathcal{Q}$ is a sequence of codes indexed by their block lengths such that

$$\lim_{\mathsf{n}\to\infty}\left(P_\mathrm{e}^{(\mathsf{n})} + P_\mathbf{x}^{(\mathsf{n})} + \frac{1}{|\mathcal{M}^{(\mathsf{n})}|}\right) = 0.$$

In other words, reliable sequences are sequences of codes whose overall error probability, detected and undetected, vanishes and whose size of message set grows to infinity with block length n.

Definition 1: The rate, erasure exponent, and error exponent of a reliable sequence $\mathcal{Q}$ are given by

$$R_\mathcal{Q} \triangleq \liminf_{\mathsf{n}\to\infty}\frac{\ln|\mathcal{M}^{(\mathsf{n})}|}{\mathsf{n}}$$
$$E_{\mathbf{x}\mathcal{Q}} \triangleq \liminf_{\mathsf{n}\to\infty}\frac{-\ln P_\mathbf{x}^{(\mathsf{n})}}{\mathsf{n}}$$
$$E_{\mathrm{e}\mathcal{Q}} \triangleq \liminf_{\mathsf{n}\to\infty}\frac{-\ln P_\mathrm{e}^{(\mathsf{n})}}{\mathsf{n}}.$$

Haroutunian [17, Th. 2] has already established a strong converse for erasure-free block codes with feedback which in our setting implies that $\lim_{\mathsf{n}\to\infty}(P_\mathrm{e}^{(\mathsf{n})} + P_\mathbf{x}^{(\mathsf{n})}) = 1$ for all codes whose rates are strictly above the capacity, i.e., $R > \mathcal{C}$. Thus, we consider only rates that are less than or equal to the capacity $R \leq \mathcal{C}$. For all rates R below capacity and for all nonnegative erasure exponents $E_\mathbf{x}$, we define the (true) error exponent $\mathcal{E}_\mathrm{e}(R, E_\mathbf{x})$ of fixed-length block codes with feedback to be the best error exponent of the reliable sequences[7] whose rate is at least R and whose erasure exponent is at least $E_\mathbf{x}$.

Definition 2: $\forall R \leq \mathcal{C}$ and $\forall E_\mathbf{x} \geq 0$, the error exponent $\mathcal{E}_\mathrm{e}(R, E_\mathbf{x})$ is

$$\mathcal{E}_\mathrm{e}(R, E_\mathbf{x}) \triangleq \sup_{\mathcal{Q}: R_\mathcal{Q} \geq R, E_{\mathbf{x}\mathcal{Q}} \geq E_\mathbf{x}} E_{\mathrm{e}\mathcal{Q}}. \tag{3}$$

Note that

$$\mathcal{E}_\mathrm{e}(R, E_\mathbf{x}) = \mathcal{E}(R) \quad \forall E_\mathbf{x} > \mathcal{E}(R) \tag{4}$$

[5]For $t = 1$, we have $\mathsf{Z}_0 = \mathsf{A}_0$.

[6]We make a slight abuse of notation and denote Z_0^t by Z^t.

[7]We restrict ourselves to the reliable sequences in order to ensure finite error exponent at zero erasure exponent. Note that a decoder which always declares erasures has zero erasure exponent and infinite error exponent.

where $\mathcal{E}(R)$ is the (true) error exponent of erasure-free block codes on DMCs with feedback.[8] Thus, benefit of the errors-and-erasures decoding is the possible increase in the error exponent as the erasure exponent goes below $\mathcal{E}(R)$.

Determining $\mathcal{E}(R)$ for all R's and for all channels is still an open problem; only upper and lower bounds to $\mathcal{E}(R)$ are known. Our investigation focuses on quantifying the gains of errors-and-erasures decoding instead of finding $\mathcal{E}(R)$. Consequently, we restrict ourselves to the region where the erasure exponent is lower than the error exponent for the encoding scheme.

For future reference let us recall the expressions for the random coding exponent and the sphere packing exponent

$$E_r(R,P) = \min_{V} \mathsf{D}\left(V\| W|P\right) + |\mathsf{I}\left(P,V\right) - R|^{+} \quad (5)$$

$$E_r(R) = \max_{P} E_r(R,P) \quad (6)$$

$$E_{sp}(R,P) = \min_{V:\mathsf{I}(P,V)\le R} \mathsf{D}\left(V\| W|P\right) \quad (7)$$

$$E_{sp}(R) = \max_{P} E_{sp}(R,P) \quad (8)$$

where $\mathsf{D}\left(V\| W|P\right)$ stands for conditional Kullback–Leibler divergence of V and W under P, and $\mathsf{I}\left(P,V\right)$ stands for mutual information for input distribution P and channel V.

We denote the y marginal of a distribution like $P(x)V(y|x)$ by $(PV)_Y$. The support of a probability distribution P is denoted by $\mathsf{supp}P$.

III. An Achievable Error-Exponent/Erasure-Exponent Tradeoff

In this section, we establish a lower bound to the achievable error exponent as a function of erasure exponent and rate. We use a two-phase encoding scheme similar to the one described by Yamamoto and Ito [33] together with a decoding rule similar to the one described by Telatar [31]. In the first phase, the transmitter uses a fixed-composition code of length $\alpha\mathsf{n}$ and rate $\frac{R}{\alpha}$. At the end of the first phase, the receiver makes a maximum mutual information decoding to obtain a tentative decision $\tilde{\mathsf{M}}$. The transmitter knows $\tilde{\mathsf{M}}$ because of the feedback link. In the remaining $(\mathsf{n}-\mathsf{n}_1)$ time units, i.e., the second phase, the transmitter confirms the tentative decision by sending the accept codeword, if $\tilde{\mathsf{M}}=\mathsf{M}$, and rejects it by sending the reject codeword otherwise. At the end of the second phase, the receiver either declares an erasure or declares the tentative decision as the decoded message. Receiver declares the tentative decision as the decoded message only when the tentative decision "dominates" all other messages. The word "dominate" will be made precise later in Section III-B. Our scheme is inspired by [33] and [31]. However, unlike [33], our decoding rule makes use of outputs of both of the phases instead of output of just second phase while deciding between declaring an erasure or declaring the tentative decision as the final one, and unlike [31], our encoding scheme is a feedback encoding scheme with two phases.

[8]In order to see this consider a reliable sequence with erasures $\mathcal{Q}$ and replace its decoding algorithm by an erasure-free decoding algorithm such that $\hat{\mathsf{M}}'(z^{\mathsf{n}}) = \hat{\mathsf{M}}(z^{\mathsf{n}})$ if $\hat{\mathsf{M}}(z^{\mathsf{n}}) \neq \mathbf{x}$, to obtain a new reliable sequence $\mathcal{Q}'$. Then, $P_{\mathbf{e}_{\mathcal{Q}'}}^{(\mathsf{n})} \le P_{\mathbf{x}_{\mathcal{Q}}}^{(\mathsf{n})} + P_{\mathbf{e}_{\mathcal{Q}}}^{(\mathsf{n})}$; thus, $E_{\mathbf{e}\mathcal{Q}'} = \min\{E_{\mathbf{x}\mathcal{Q}}, E_{\mathbf{e}\mathcal{Q}}\}$ and $R_{\mathcal{Q}'} = R_{\mathcal{Q}}$. This together with the definition of $\mathcal{E}(R)$ leads to (4).

In the rest of this section, we analyze the performance of this coding architecture and derive an achievable error-exponent expression in terms of a given rate R, erasure exponent $E_{\mathbf{x}}$, time-sharing constant α, communication phase type P, control phase type (joint empirical type of the accept codeword and reject codeword) Π, and domination rule $\succ$. Then, we optimize over $\succ$, Π, P, and α to obtain an achievable error-exponent expression as a function of rate R and erasure exponent $E_{\mathbf{x}}$.

A. Fixed-Composition Codes and the Packing Lemma

We start with a very brief overview of certain properties of types. Those readers who are not familiar with method types can use [9] for a concise introduction or [10] for a thorough study. The empirical distribution of an $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ is called the type of x^{n} and the empirical distribution of transitions from a $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ to a $y^{\mathsf{n}} \in \mathcal{Y}^{\mathsf{n}}$ is called the conditional type[9]

$$\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) \triangleq \sum_{t=1}^{\mathsf{n}} \frac{\mathbb{1}_{\{x_t=\tilde{x}\}}}{\mathsf{n}} \qquad \forall \tilde{x} \in \mathcal{X}. \quad (9)$$

$$\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}(\tilde{y}|\tilde{x}) \triangleq \sum_{t=1}^{\mathsf{n}} \frac{\mathbb{1}_{\{x_t=\tilde{x}\}}\mathbb{1}_{\{y_t=\tilde{y}\}}}{\mathsf{n}\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x})} \quad \forall \tilde{y}\in\mathcal{Y}, \forall \tilde{x} : \mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) > 0. \quad (10)$$

For any probability transition matrix $W : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \rightarrow \mathcal{Y}$, we have[10]

$$\prod_{t=1}^{\mathsf{n}} W(y_t|x_t) = e^{-\mathsf{n}\left(\mathsf{D}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}\middle\| W|\mathsf{P}_{x^{\mathsf{n}}}\right) + \mathsf{H}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}\middle|\mathsf{P}_{x^{\mathsf{n}}}\right)\right)}. \quad (11)$$

The set of all y^{n}'s with the same conditional type V with respect to x^{n} is called the V-shell of x^{n} and denoted by $T_V\left(x^{\mathsf{n}}\right)$

$$T_V\left(x^{\mathsf{n}}\right) = \{y^{\mathsf{n}} : \mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} = V\}. \quad (12)$$

Note that for any transition probability matrix from $\mathcal{X}$ to $\mathcal{Y}$ total probability of $T_V\left(x^{\mathsf{n}}\right)$ has to be less than one. Thus, by assuming that transition probabilities are V and using (11), we can conclude that

$$|T_V\left(x^{\mathsf{n}}\right)| \le e^{\mathsf{H}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}\middle|\mathsf{P}_{x^{\mathsf{n}}}\right)}. \quad (13)$$

Codes whose codewords all have the same empirical distribution $\mathsf{P}_{x^{\mathsf{n}}(m)} = P\ \forall m \in \mathcal{M}$ are called fixed-composition codes. In Section III-D, we will describe the error-and-erasure events in terms of the intersections of V-shells of different codewords. For doing that, let us define $F^{(\mathsf{n})}\left(V,\hat{V},m\right)$ as the intersection of V-shell of $x^{\mathsf{n}}(m)$ and the $\hat{V}$-shells of other codewords

$$F^{(\mathsf{n})}\left(V,\hat{V},m\right) \triangleq T_V\left(x^{\mathsf{n}}(m)\right) \bigcap \cup_{\tilde{m}\neq m} T_{\hat{V}}\left(x^{\mathsf{n}}(\tilde{m})\right). \quad (14)$$

The following packing lemma, proved by Csiszár and Körner [10, Lemma 2.5.1], claims the existence of a code with a guaranteed upper bound on the size of $F^{(\mathsf{n})}\left(V,\hat{V},m\right)$.

[9]Note that $\mathsf{P}_{x^{\mathsf{n}}}$ corresponds to a distribution on $\mathcal{X}$ for all $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$, where as $\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}$ determines a channel from the support of $\mathsf{P}_{x^{\mathsf{n}}}$ to $\mathcal{Y}$.

[10]Note that for any $W : \mathcal{X} \rightarrow \mathcal{Y}$, there is unique consistent $W' : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \rightarrow \mathcal{Y}$.

Lemma 1: For every block length $\mathsf{n} \geq 1$, rate $R > 0$, and type P satisfying $H(P) > R$, there exist at least $\lfloor e^{\mathsf{n}(R-\delta_{\mathsf{n}})} \rfloor$ distinct type P sequences in $\mathcal{X}^{\mathsf{n}}$ such that for every pair of stochastic matrices $V : \mathsf{supp}P \to \mathcal{Y}$, $\hat{V} : \mathsf{supp}P \to \mathcal{Y}$ and $\forall m \in \mathcal{M}$

$$\left| F^{(\mathsf{n})}\left(V, \hat{V}, m\right) \right| \leq \left| T_V\left(x^{\mathsf{n}}(m)\right) \right| e^{-\mathsf{n}|I(P,\hat{V})-R|^+}$$

where $\delta_{\mathsf{n}} = \frac{\ln 4 + (4|\mathcal{X}|+6|\mathcal{X}||\mathcal{Y}|)\ln(\mathsf{n}+1)}{\mathsf{n}}$.

The above lemma is stated in a slightly different way by Csiszár and Körner [10], for a fixed δ and large enough n. However, this form follows immediately from their proof.

If we use Lemma 1 together with (11) and (13) we can bound the conditional probability of observing a $y^{\mathsf{n}} \in F^{(\mathsf{n})}\left(V, \hat{V}, m\right)$ when $\mathsf{M} = m$ as follows.

Corollary 1: In a code satisfying Lemma 1, when message $m \in \mathcal{M}$ is sent, the probability of receiving a $y^{\mathsf{n}} \in T_V\left(x^{\mathsf{n}}(m)\right)$ which is also in $T_{\hat{V}}\left(x^{\mathsf{n}}(\tilde{m})\right)$, for some $\tilde{m} \in \mathcal{M}$ such that $\tilde{m} \neq m$ is bounded as follows:

$$\mathbf{P}\left[F^{(\mathsf{n})}\left(V, \hat{V}, \mathsf{M}\right) \middle| \mathsf{M} \right] \leq e^{-\mathsf{n}\eta\left(R,P,V,\hat{V}\right)} \tag{15}$$

where

$$\eta\left(R, P, V, \hat{V}\right) \triangleq \mathsf{D}\left(V \| W | P\right) + \left|\mathsf{I}\left(P, \hat{V}\right) - R\right|^+. \tag{16}$$

B. Coding Algorithm

In the first phase, the communication phase, we use a length $\mathsf{n}_1 = \lceil \alpha \mathsf{n} \rceil$ type P fixed-composition code with $\lfloor e^{\mathsf{n}_1(\frac{R}{\alpha}-\delta_{\mathsf{n}_1})} \rfloor$ codewords which satisfies the property described in Lemma 1. At the end of the first phase, the receiver makes a tentative decision by choosing the codeword that has the maximum empirical mutual information with the output sequence $\mathsf{Y}^{\mathsf{n}_1}$. If there is a tie, i.e., if there are more than one codewords which have the maximum empirical mutual information, the receiver chooses the codeword which has the lowest index

$$\tilde{\mathsf{M}} = \left\{ m : \begin{array}{l} \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) > \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) \ \forall \tilde{m} < m \\ \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) \geq \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) \ \forall \tilde{m} > m \end{array} \right\}. \tag{17}$$

In the remaining $(\mathsf{n} - \mathsf{n}_1)$ time units, the transmitter sends the accept codeword $x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)$ if $\tilde{\mathsf{M}} = \mathsf{M}$ and sends the reject codeword $x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)$ otherwise.

Note that our encoding scheme uses the feedback link actively for the encoding neither within the first phase nor within the second phase. It does not even change the codewords it uses for accepting or rejecting the tentative decision depending on the observation in the first phase. Feedback is only used to reveal the tentative decision to the transmitter.

Accept and reject codewords have joint type $\Pi(\tilde{x}, \tilde{\tilde{x}})$, i.e., the ratio of the number of instances in which accept codeword has an $\tilde{x} \in \mathcal{X}$ and reject codeword has a $\tilde{\tilde{x}} \in \mathcal{X}$ to the length of the codewords $(\mathsf{n} - \mathsf{n}_1)$ is $\Pi(\tilde{x}, \tilde{\tilde{x}})$. The joint conditional type of the output sequence in the second phase $\mathsf{U}_{y^{\mathsf{n}}_{\mathsf{n}_1+1}}$ is the empirical conditional distribution of $y^{\mathsf{n}}_{\mathsf{n}_1+1}$. We call the set of all output sequences $y^{\mathsf{n}}_{\mathsf{n}_1+1}$ whose joint conditional type is U the U-shell and denote it by T_U.

Like we did in the Corollary 1, we can upper bound the probability of U-shells. Note that if $\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1} \in T_U$, then

$$\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1} \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)\right] = e^{-(\mathsf{n}-\mathsf{n}_1)(\mathsf{D}(U\|W_a|\Pi)+\mathsf{H}(U|\Pi))}$$
$$\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1} \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right] = e^{-(\mathsf{n}-\mathsf{n}_1)(\mathsf{D}(U\|W_r|\Pi)+\mathsf{H}(U|\Pi))}$$

where $x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)$ is the accept codeword, $x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)$ is the reject codeword, $W_a(y|\tilde{x}, \tilde{\tilde{x}}) = W(y|\tilde{x})$, and $W_r(y|\tilde{x}, \tilde{\tilde{x}}) = W(y|\tilde{\tilde{x}})$. Noting that $|T_U| \leq e^{(\mathsf{n}-\mathsf{n}_1)\mathsf{H}(U|\Pi)}$, we get

$$\mathbf{P}\left[T_U \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)\right] \leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_a|\Pi)} \tag{18a}$$
$$\mathbf{P}\left[T_U \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right] \leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_r|\Pi)}. \tag{18b}$$

C. Decoding Rule

For an encoder like the one in Section III-B, a decoder that depends only on the conditional type of $\mathsf{Y}^{\mathsf{n}_1}$ for different codewords in the communication phase, i.e., $\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(m)}$'s for $m \in \mathcal{M}$, the conditional type of the channel output in the control phase, i.e., $\mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}$, and the indices of the codewords can achieve the minimum error probability for a given erasure probability. However, finding that decoder becomes analytically intractable. Instead, we restrict ourselves to the decoders that can be written in terms of pairwise comparisons between messages given Y^{n}. Furthermore, we assume that these pairwise comparisons depend only on the conditional type of $\mathsf{Y}^{\mathsf{n}_1}$ for the messages compared, the conditional output type in the control phase, and the indices of the messages. Thus, if the triplet corresponding to the tentative decision $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(\tilde{\mathsf{M}})}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}})$ dominates all other triplets of the form $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(m)}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m)$ for $m \neq \tilde{\mathsf{M}}$, the tentative decision becomes final; else an erasure is declared.[11] Hence, the decoder is of the form given in (19), shown at the bottom of the page.

The binary relation $\succ$ used in (19) is such that if (V, U, m) dominates $(\hat{V}, U, \tilde{m})$ then $(\hat{V}, U, \tilde{m})$ does not dominate (V, U, m), i.e.,

$$(V, U, m) \succ (\hat{V}, U, \tilde{m}) \Rightarrow (\hat{V}, U, \tilde{m}) \not\succ (V, U, m).$$

[11]Note that conditional probability $\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}} \middle| \mathsf{M} = m\right]$ is only a function of corresponding $\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}$ and $\mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}$. Thus, all decoding rules that accept or reject the tentative decision $\tilde{\mathsf{M}}$, based on a threshold test on likelihood ratios, $\frac{\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}} | \mathsf{M}=\tilde{\mathsf{M}}\right]}{\mathbf{P}\left[\mathsf{Y}^{\mathsf{n}} | \mathsf{M}=m\right]}$, for $m \neq \tilde{\mathsf{M}}$, are in this family of decoding rules.

$$\hat{\mathsf{M}} = \begin{cases} \tilde{\mathsf{M}}, & \text{if } (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\tilde{\mathsf{M}}}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m) \qquad \forall m \neq \tilde{\mathsf{M}} \\ \mathbf{x}, & \text{if } \exists m \neq \tilde{\mathsf{M}} \text{ such that } (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\tilde{\mathsf{M}}}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \not\succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m). \end{cases} \tag{19}$$

This property is a necessary and sufficient condition for a binary relation to be a domination rule. Decoder given by (19), however, either accepts or rejects the tentative decision $\tilde{\mathsf{M}}$ given in (17). Consequently, its domination rule also satisfies the following two properties.

1) If the empirical mutual informations of the messages in the communication phase are not equal, only the message with larger mutual information can dominate the other one.
2) If the empirical mutual informations of the messages in the communication phase are equal, only the message with lower index can dominate the other one.

For any such binary relation there is a corresponding decoder of the form given in (19). In our scheme, we either use the trivial domination rule leading to the trivial decoder $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$ or the domination rule given in (20), shown at the bottom of the page, both of which satisfy these conditions.

Among the family of decoders we are considering, i.e., among the decoders that only depend on the pairwise comparisons between conditional types and indices of the messages compared, the decoder given in (19) and (20) is optimal in terms of error-exponent/erasure-exponent tradeoff. Furthermore, in order to employ this decoding rule, the receiver needs to determine only the two messages with the highest empirical mutual information in the first phase. Then, the receiver needs to check whether the triplet corresponding to the tentative decision dominates the triplet corresponding to the message with the second highest empirical mutual information. If it does, then, for the rule given in (20), it is guaranteed to dominate the rest of the triplets as well.

D. Error Analysis

Using the encoder like the one described in Section III-B and the decoder like the one in (19), we achieve the performance given below. If $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$, then the domination rule given in (20) is used in the decoder; else a trivial domination rule that leads to an erasure-free decoding $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$ is used in the decoder.

Theorem 1: For any block length $\mathsf{n} \geq 1$, rate R, erasure exponent $E_{\mathbf{x}}$, time-sharing constant α, communication phase type P, and control phase type Π, there exists a length n block code with feedback such that

$$\begin{aligned} \ln|\mathcal{M}| &\geq e^{\mathsf{n}(R-\delta_{\mathsf{n}})} \\ P_{\mathbf{x}} &\leq e^{-\mathsf{n}(E_{\mathbf{x}}-\delta'_{\mathsf{n}})} \\ P_{\mathbf{e}} &\leq e^{-\mathsf{n}(E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)-\delta'_{\mathsf{n}})} \end{aligned}$$

where $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is given by (21), shown at the bottom of the page. The optimization problem given in (21) is a convex optimization problem: it is minimization of a convex function over a convex set. Thus, the value of the exponent $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ can numerically be calculated relatively easily. Furthermore, $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ can be written in terms of solutions of lower dimensional optimization problems; see (42). However, problem of finding the optimal (α, P, Π) triple for a given $(R, E_{\mathbf{x}})$ pair is not that easy in general, as we will discuss in more detail in Section III-E.

For all control phase types Π and control phase output types U, $\mathsf{D}(U\| W_a|\Pi) \geq 0$, $\mathsf{D}(U\| W_r|\Pi) \geq 0$. Using this fact together with the definitions of $E_r(R,P)$, $\eta\left(R,P,\hat{V},V\right)$, and $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ given in (6), (16), and (21), we can conclude that $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) \geq \alpha E_r(\frac{R}{\alpha}, P)$ for all $(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ such that $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$.

We are interested in quantifying the gains of errors-and-erasures decoding over the decoding schemes without erasures, thus we are ultimately interested only in the region where $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ holds. However, (21) gives us the whole achievable region for the family of codes we are considering.

$$(V,U,m) \succ (\hat{V},U,\tilde{m}) \Leftrightarrow \begin{cases} \mathsf{I}(P,V) > \mathsf{I}\left(P,\hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right) + (1-\alpha)\mathsf{D}(U\|W_a|\Pi) \leq E_{\mathbf{x}}, & \text{for } m \geq \tilde{m} \\ \mathsf{I}(P,V) \geq \mathsf{I}\left(P,\hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right) + (1-\alpha)\mathsf{D}(U\|W_a|\Pi) \leq E_{\mathbf{x}}, & \text{for } m < \tilde{m}. \end{cases} \tag{20}$$

$$E_{\mathbf{e}} = \left\{ \begin{array}{ll} \alpha E_r\left(\frac{R}{\alpha},P\right), & \text{if } E_{\mathbf{x}} > \alpha E_r\left(\frac{R}{\alpha},P\right) \\ \min\limits_{\substack{(V,\hat{V},U):(V,\hat{V},U)\in\mathcal{V} \\ \alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_a|\Pi)\leq E_{\mathbf{x}}}} \alpha\eta\left(\frac{R}{\alpha},P,\hat{V},V\right) + (1-\alpha)\mathsf{D}(U\|W_r|\Pi), & \text{if } E_{\mathbf{x}} \leq \alpha E_r\left(\frac{R}{\alpha},P\right) \end{array} \right\} \tag{21a}$$

$$\mathcal{V} = \{(V_1,V_2,U) : \mathsf{I}(P,V_1) \geq \mathsf{I}(P,V_2) \text{ and } (PV_1)_Y = (PV_2)_Y\} \tag{21b}$$

$$\delta'_{\mathsf{n}} = \frac{(|\mathcal{X}|+1)^2|\mathcal{Y}|\log(\mathsf{n}+1)}{\mathsf{n}}. \tag{21c}$$

Proof: A decoder of the form given in (19) decodes correctly when $\tilde{\mathsf{M}} = \mathsf{M}$ and $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \succ (\mathsf{Y}^{\mathsf{n}}, m)$ for all[12] $m \neq \mathsf{M}$. Thus, an error or an erasure occurs only when the correct message does not dominate all other messages, i.e., when $\exists m \neq \mathsf{M}$ such that $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \not\succ (\mathsf{Y}^{\mathsf{n}}, m)$. This can happen in two ways: either there is an error in the first phase, i.e., $\tilde{\mathsf{M}} \neq m$ or first phase tentative decision is correct, i.e., $\tilde{\mathsf{M}} = m$, but the second phase observation $y_{\mathsf{n}_1+1}^{\mathsf{n}}$ leads to an erasure, i.e., $\hat{\mathsf{M}} = \mathbf{x}$. For a domination rule satisfying constraints described in Section III-C, the total probability of the above mentioned two events, hence the sum of error-and-erasure probabilities, can be bounded as[13]

$$\begin{aligned} &P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \\ &= \mathbf{P}\left[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t. } (y^{\mathsf{n}}, m) \not\succ (y^{\mathsf{n}}, \tilde{m})\} \,|\, \mathsf{M} = m\right] \\ &\leq \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \geq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}\left[y^{\mathsf{n}_1} | m\right] \\ &\quad + \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \leq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}\left[y^{\mathsf{n}_1} | m\right] \\ &\quad \times \sum_{U:(V,U,m) \not\succ (\hat{V},U,m+1)} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right]. \end{aligned} \tag{22}$$

where $F^{(\mathsf{n}_1)}(V, \hat{V}, m)$ is the intersection of V-shell of message $m \in \mathcal{M}$ with the $\hat{V}$-shells of other messages, defined in (14).

Now we bound the summands in (22).

- As a result of Corollary 1, we have

$$\begin{aligned} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}\left[y^{\mathsf{n}_1} | m\right] &= \mathbf{P}\left[F^{(\mathsf{n}_1)}\left(V, \hat{V}, m\right) \middle| \mathsf{M} = m\right] \\ &\leq e^{-\mathsf{n}_1 \eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right)}. \end{aligned}$$

- Furthermore, because of (18a)

$$\begin{aligned} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right] &= \mathbf{P}\left[T_U \middle| \mathsf{X}_{\mathsf{n}_1+1}^{\mathsf{n}} = x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right] \\ &\leq e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_a|\Pi)}. \end{aligned}$$

- In addition, the number of different nonempty V-shells in the communication phase is less than $(\mathsf{n}_1 + 1)^{|\mathcal{X}||\mathcal{Y}|}$ and the number of nonempty U-shells in the control phase is less than $(\mathsf{n} - \mathsf{n}_1 + 1)^{|\mathcal{X}|^2|\mathcal{Y}|}$.

Thus, we can bound $P_{\mathbf{e}|m} + P_{\mathbf{x}|m}$ like

$$\begin{aligned} &P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \\ &\leq (\mathsf{n}_1 + 1)^{2|\mathcal{X}||\mathcal{Y}|} \max_{V,\hat{V}: \mathsf{I}(P,V) \leq \mathsf{I}(P,\hat{V})} e^{-\mathsf{n}_1 \eta\left(R/\alpha, P, V, \hat{V}\right)} \\ &\quad + (\mathsf{n}_1 + 1)^{2|\mathcal{X}||\mathcal{Y}|} (\mathsf{n} - \mathsf{n}_1 + 1)^{|\mathcal{X}|^2|\mathcal{Y}|} \\ &\quad \max_{(V,\hat{V},U) \in \mathcal{V}_{\mathbf{x}}} e^{-\mathsf{n}_1\left(\eta\left(R/\alpha, P, V, \hat{V}\right)\right) - (\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_a|\Pi)} \end{aligned} \tag{23}$$

where $\mathcal{V}_{\mathbf{x}}$ is the set of $(V, \hat{V}, U)$ triples corresponding to erasures with a correct tentative decision, defined in[14]

$$\begin{aligned} \mathcal{V}_{\mathbf{x}} \triangleq \Big\{(V, \hat{V}, U) : \mathsf{I}(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \\ \text{and } (PV)_Y = (P\hat{V})_Y \\ \text{and } (V, U, m) \not\succ (\hat{V}, U, m+1)\Big\}. \end{aligned} \tag{24}$$

Using the definition of $E_r(\frac{R}{\alpha}, P)$ given in (5) together with the inequality (23), we bound $P_{\mathbf{e}|m} + P_{\mathbf{x}|m}$ by the inequality (25), shown at the bottom of the page.

On the other hand, an error occurs only when an incorrect message dominates all other messages, i.e., when $\exists \tilde{m} \neq m$ such that $(\mathsf{Y}^{\mathsf{n}}, \tilde{m}) \succ (\mathsf{Y}^{\mathsf{n}}, \tilde{\tilde{m}})$ for all $\tilde{m} \neq \tilde{\tilde{m}}$, thus $P_{\mathbf{e}}$ is given by

$$\begin{aligned} &P_{\mathbf{e}|m} \\ &= \mathbf{P}\left[\left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t. } (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, \tilde{\tilde{m}})\ \forall \tilde{\tilde{m}} \neq \tilde{m}\right\} \middle| \mathsf{M} = m\right]. \end{aligned} \tag{26}$$

When a $\tilde{m} \in \mathcal{M}$ dominates all other $\tilde{\tilde{m}} \neq \tilde{m}$, it also dominates m, i.e., we have the relation given in

$$\begin{aligned} &\left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t. } (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, \tilde{\tilde{m}})\ \forall \tilde{\tilde{m}} \neq \tilde{m}\right\} \\ &\qquad \subset \left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t. } (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, m)\right\}. \end{aligned} \tag{27}$$

Consequently, we can bound the error probability by

$$\begin{aligned} P_{\mathbf{e}|m} &\leq \mathbf{P}\left[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.} (y^{\mathsf{n}}, \tilde{m}) \succ (y^{\mathsf{n}}, m)\} \,|\, \mathsf{M} = m\right] \\ &= \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \geq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \mathsf{M} = m] \\ &\quad \times \sum_{U:(\hat{V},U,m-1) \succ (V,U,m)} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| x_{\mathsf{n}_1+1}^{\mathsf{n}}(r)\right]. \end{aligned} \tag{28}$$

The tentative decision is not equal to m only if there is a message with a strictly higher empirical mutual information or if there is a messages which has an equal mutual information but smaller index. This is the reason why we sum over $(\hat{V}, U, m-1) \succ (V, U, m)$ in (28). Using inequality (18b) in the inner most two sums and then applying inequality (15), we get (29), shown

[12]We use the short hand $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \succ (\mathsf{Y}^{\mathsf{n}}, m)$ for $\left(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|\mathsf{M}}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, \mathsf{M}\right) \succ \left(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|m}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, m\right)$ in the rest of this section.

[13]Note that for the case when $m = |\mathcal{M}|$, we need to replace $(V, U, m) \not\succ (\hat{V}, U, m+1)$ with $(V, U, m-1) \not\succ (\hat{V}, U, m)$.

[14]Note that m in (24) is a dummy variable and $\mathcal{V}_{\mathbf{x}}$ is the same set for all $m \in \mathcal{M}$.

$$P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq e^{\mathsf{n}\delta'_{\mathsf{n}}} \max\left\{e^{-\mathsf{n}\alpha E_r(R/\alpha, P)}, e^{-\mathsf{n} \min_{(V,\hat{V},U) \in \mathcal{V}_{\mathbf{x}}} \alpha\eta\left(R/\alpha, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}(U\|W_a|\Pi)}\right\}. \tag{25}$$

at the bottom of the page, where[15] $\mathcal{V}_{\mathbf{e}}$ is the complement of $\mathcal{V}_{\mathbf{x}}$ in $\mathcal{V}$ given by

$$\mathcal{V}_{\mathbf{e}} \triangleq \Big\{(V,\hat{V},U) : \mathsf{I}(P,V) \geq \mathsf{I}\left(P,\hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and } (V,U,m) \succ (\hat{V},U,m+1)\Big\}. \tag{30}$$

The domination rule $\succ$ divides the set $\mathcal{V}$ into two subsets: the erasure subset $\mathcal{V}_{\mathbf{x}}$ and the error subset $\mathcal{V}_{\mathbf{e}}$. Choosing the domination rule is equivalent to choosing the $\mathcal{V}_{\mathbf{e}}$. Depending on the value of $\alpha E_r(\frac{R}{\alpha},P)$ and $E_{\mathbf{x}}$, we chose different $\mathcal{V}_{\mathbf{e}}$'s as follows.

i) $E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha},P)$: $\mathcal{V}_{\mathbf{e}} = \mathcal{V}$. Then, $\mathcal{V}_{\mathbf{x}} = \emptyset$ and Theorem 1 follows from (25).

ii) $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha},P)$: $\mathcal{V}_{\mathbf{e}}$ is given by

$$\mathcal{V}_{\mathbf{e}} = \Big\{(V,\hat{V},U) : \mathsf{I}(P,V) \geq \mathsf{I}\left(P,\hat{V}\right), (PV)_Y = (P\hat{V})_Y \text{ and } \alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right) + (1-\alpha)\mathsf{D}\left(U\|\, W_a|\Pi\right) \leq E_{\mathbf{x}}\Big\}. \tag{31}$$

Then, all the $(V,\hat{V},U)$ triples satisfying

$$\alpha\eta\left(\frac{R}{\alpha},P,V,\hat{V}\right) + (1-\alpha)\mathsf{D}\left(U\|\, W_a|\Pi\right) \leq E_{\mathbf{x}}$$

are in the error subset. Thus, as a result of (25), erasure probability is bounded as $P_{\mathbf{x}} \leq e^{-\mathsf{n}(E_{\mathbf{x}}-\delta'_{\mathsf{n}})}$ and Theorem 1 follows from (29). ☐

E. Lower Bound to $\mathcal{E}_{\mathbf{E}}(R,E_{\mathbf{X}})$

In this section, we use Theorem 1 to derive a lower bound to the optimal error exponent $\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}})$. We do that by optimizing the achievable performance $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ over α, P, and Π.

High Erasure-Exponent Region [i.e., $E_{\mathbf{X}} > E_r(R)$]: As a result of (21), $\forall R \geq 0$, and $\forall E_{\mathbf{x}} > E_r(R)$, we have:

- $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \alpha E_r(\frac{R}{\alpha},P) \leq E_r(R)$, for all P, Π and $\alpha \in [0,1]$;
- $E_{\mathbf{e}}(R,E_{\mathbf{x}},\tilde{\alpha},\tilde{P},\Pi) = E_r(R)$ for $\tilde{\alpha} = 1$, for all Π and for $\tilde{P} = \arg\max_P E_r(R,P)$.

Thus, for all $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} > E_r(R)$: optimal time sharing constant is 1, optimal input distribution is the optimal input distribution for random coding exponent at rate R, we use maximum mutual information decoding and never declare erasures. Furthermore, since $\alpha = 1$, we have only a single phase in our scheme. Thus, $\forall R \geq 0$ and $\forall E_{\mathbf{x}} > E_r(R)$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}}) = E_{\mathbf{e}}(R,E_{\mathbf{x}},1,P_{r(R)},\Pi) = E_r(R) \tag{32}$$

where $P_{r(R)}$ satisfies $E_r(R,P_{r(R)}) = E_r(R)$ and Π can be any control phase type. Evidently benefits of errors-and-erasures decoding are not observed in this region.

Low Erasure-Exponent Region [i.e., $E_{\mathbf{X}} \leq E_r(R)$]: We observe and quantify the benefits of errors-and-erasures decoding for $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$. Since $E_r(R)$ is a nonnegative nonincreasing and convex function of R, $\forall R \geq 0$, $\forall 0 < E_{\mathbf{x}} \leq E_r(R)$, we have

$$\alpha \in [\alpha^*(R,E_{\mathbf{x}}),1] \Leftrightarrow E_{\mathbf{x}} \leq \alpha E_r\left(\frac{R}{\alpha}\right)$$

where $\alpha^*(R,E_{\mathbf{x}})$ is the unique solution of the equation $\alpha E_r(\frac{R}{\alpha}) = E_{\mathbf{x}}$.

For the case $E_{\mathbf{x}} = 0$, however, $\alpha E_r(\frac{R}{\alpha}) = 0$ has multiple solutions and Theorem 1 holds but resulting error exponent $E_{\mathbf{e}}(R,0,\alpha,P,\Pi)$ does not correspond to the error exponent of a reliable sequence. Convention introduced in (33) addresses both issues at once, by choosing the minimum of those solutions as $\alpha^*(R,0)$. In addition, by this convention, $\alpha^*(R,E_{\mathbf{x}})$ is also continuous at $E_{\mathbf{x}} = 0$: $\lim_{E_{\mathbf{x}} \to 0} \alpha^*(R,E_{\mathbf{x}}) = \alpha^*(R,0)$

$$\alpha^*(R,E_{\mathbf{x}}) \triangleq \begin{cases} \dfrac{R}{g^{-1}(E_{\mathbf{x}}/R)} & E_{\mathbf{x}} \in (0,E_r(R)] \\ R/\mathcal{C} & E_{\mathbf{x}} = 0 \end{cases} \tag{33}$$

where $g^{-1}(\cdot)$ is the inverse of the function $g(r) = \frac{E_r(r)}{r}$.

As a result of (21) and (33), $\forall R \geq 0$ and $\forall 0 < E_{\mathbf{x}} \leq E_r(R)$, we have:

- $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \alpha E_r(\frac{R}{\alpha},P) \leq E_r(R)$ for all P, Π and $\alpha \in [0,\alpha^*(R,E_{\mathbf{x}}))$;
- $E_{\mathbf{e}}(R,E_{\mathbf{x}},\tilde{\alpha},\tilde{P},\Pi) = E_r(R)$ for $\tilde{\alpha} = 1$, for all Π and for $\tilde{P} = \arg\max_P E_r(R,P)$.

Thus, for all $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$, optimal time sharing constant is in the interval $[\alpha^*(R,E_{\mathbf{x}}),1]$.

For an $(R,E_{\mathbf{x}},\alpha)$ triple such that $R \geq 0$, $E_{\mathbf{x}} \leq E_r(R)$, and $\alpha \in [\alpha^*(R,E_{\mathbf{x}}),1]$, let $\mathcal{P}(R,E_{\mathbf{x}},\alpha)$ be

$$\mathcal{P}(R,E_{\mathbf{x}},\alpha) \triangleq \left\{P : \alpha E_r(\frac{R}{\alpha},P) \geq E_{\mathbf{x}}, \mathsf{I}(P,W) \geq \frac{R}{\alpha}\right\}. \tag{34}$$

The constraint on mutual information is there to ensure that $E_{\mathbf{e}}(R,0,\alpha,P,\Pi)$'s are corresponding to error exponent of

[15]Note that m in the definition of $\mathcal{V}_{\mathbf{e}}$ is also a dummy variable.

$$\begin{aligned} P_{\mathbf{e}|m} &\leq (\mathsf{n}+1)^{(|\mathcal{X}|^2+2|\mathcal{X}|)|\mathcal{Y}|} \max_{\substack{(V,\hat{V},U):\, \mathsf{I}\left(P,\hat{V}\right)\geq \mathsf{I}(P,V) \\ (\hat{V},U,m-1)\succ(V,U,m)}} e^{-\mathsf{n}\left(\alpha\eta\left(R/\alpha,P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi)\right)} \\ &\leq e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n}\min_{(\hat{V},V,U)\in\mathcal{V}_{\mathbf{e}}}\left(\alpha\eta\left(R/\alpha,P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi)\right)} \\ &= e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n}\min_{(V,\hat{V},U)\in\mathcal{V}_{\mathbf{e}}}\left(\alpha\eta\left(R/\alpha,P,\hat{V},V\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi)\right)} \end{aligned} \tag{29}$$

reliable sequences. The set $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ is convex because $E_r(R,P)$ and $\mathsf{I}(P,W)$ are concave in P.

Note the following.

- If $R \geq 0$ and $E_{\mathbf{x}} \in (0, E_r(R)]$, $E_{\mathrm{e}}(R, 0, \alpha, P, \Pi)$ satisfies the constraint given in (35a)–(35b), shown at the bottom of the page. Thus, for such $(R, E_{\mathbf{x}})$ pairs, we can restrict the optimization over P to the set $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$.
- If $R \geq 0$ and $E_{\mathbf{x}} = 0$, $E_{\mathrm{e}}(R, 0, \alpha, P, \Pi)$ corresponds to the error exponent of a reliable sequence only if it satisfies constraints given in (35a)–(35b).

Thus, using Theorem 1, we conclude that $E_{\mathrm{e}}(R, E_{\mathbf{x}})$ given in (36), shown at the bottom of the page, is an achievable error exponent at rate R and erasure exponent $E_{\mathbf{x}}$, where $E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$, $\alpha^*(R, E_{\mathbf{x}})$ and $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ are given in (21), (33), and (34), receptively.

Note that unlike $E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ itself, $E_{\mathrm{e}}(R, E_{\mathbf{x}})$ as defined in (36) corresponds to the error exponent of reliable code sequences even at $E_{\mathbf{x}} = 0$.

If the maximizing P, for the inner maximization in (36), is the same for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]$, the optimal value of α is $\alpha^*(R, E_{\mathbf{x}})$. In order to see that, we first observe that in any fixed $(R, E_{\mathbf{x}}, P, \Pi)$ such that $E_r(R,P) \geq E_{\mathbf{x}}$, function $E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is convex in α for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}, P), 1]$, where $\alpha^*(R, E_{\mathbf{x}}, P)$ is the unique solution of the equation[16] $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$ as is shown in Lemma 10 in Appendix B. Since the maximization preserves the convexity, $\max_{\Pi} E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is also convex in α for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}, P), 1]$. Thus, for any $(R, E_{\mathbf{x}}, P)$ triple, $\max_{\Pi} E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ takes its maximum value either at the minimum possible value of α, i.e., $\alpha^*(R, E_{\mathbf{x}}, P) = \alpha^*(R, E_{\mathbf{x}})$, or at the maximum possible value of α, i.e., 1. It is shown in Appendix C that $\max_{\Pi} E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ takes its maximum value at $\alpha = \alpha^*(R, E_{\mathbf{x}})$.

Furthermore, if the maximizing P is not only the same for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]$ for a given $(R, E_{\mathbf{x}})$ pair but also for all $(R, E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$, then we can find the optimal $E_{\mathrm{e}}(R, E_{\mathbf{x}})$ by simply maximizing over Π's. In symmetric channels, for example, uniform distribution is the optimal distribution for all $(R, E_{\mathbf{x}})$ pairs and the error exponent is simply given by (37), shown at the bottom of the page, where P^* is the uniform distribution.

F. Alternative Expression for Exponent

The minimization given in (21) for $E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is over transition probability matrices and control phase output types. In order to get a better grasp of the resulting expression, we simplify the analytical expression in this section. We do that by expressing the minimization in (21) in terms of solutions of lower dimensional optimization problems.

Let $\boldsymbol{\zeta}(R, P, Q)$ be the minimum Kullback–Leibler divergence under P with respect to W among the transition probability matrices whose mutual information under P is less than R and whose output distribution under P is Q. It is shown in Appendix B that for a given P, $\boldsymbol{\zeta}(R, P, Q)$ is convex in (R, Q) pair. Evidently, for a given (P, Q) pair, $\boldsymbol{\zeta}(R, P, Q)$ is nonincreasing in R. Thus, for a given (P, Q) pair, $\boldsymbol{\zeta}(R, P, Q)$ is strictly decreasing on a closed interval and is an extended real-valued function of the form given in (38a)–(38c), shown at the bottom of the next page, where $PV \gg PW$ iff for all (x, y) pairs such that $P(x)W(y|x)$ is zero $P(x)V(y|x)$ is also zero.

Let $\Gamma(T, \Pi)$ be the minimum Kullback–Leibler divergence with respect to W_r under Π, among the U's whose Kullback–Leibler divergence with respect to W_a under Π is less than or equal to T

$$\Gamma(T, \Pi) \triangleq \min_{U: \mathsf{D}(U \| W_a | \Pi) \leq T} \mathsf{D}(U \| W_r | \Pi). \tag{39}$$

[16]Evidently, we need to make a minor modification for $E_{\mathbf{x}} = 0$ case as before to ensure that we consider only the $E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$'s that correspond to the reliable sequences: $\alpha^*(R, 0, P) = \frac{R}{\mathsf{I}(P,W)}$.

$$E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \alpha E_r\left(\frac{R}{\alpha}, P\right) \quad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1] \quad \forall P \notin \mathcal{P}(R, E_{\mathbf{x}}, \alpha) \quad \forall \Pi \tag{35a}$$

$$E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, \tilde{P}, \Pi) \geq \alpha E_r\left(\frac{R}{\alpha}\right) \quad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1] \quad \tilde{P} = \arg\max_P E_r\left(\frac{R}{\alpha}, P\right) \quad \forall \Pi. \tag{35b}$$

$$E_{\mathrm{e}}(R, E_{\mathbf{x}}) = \max_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max_{\Pi} E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) \quad \forall R \geq 0 \quad \forall E_{\mathbf{x}} \leq E_r(R) \tag{36}$$

$$E_{\mathrm{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{ll} E_{\mathrm{e}}(R, E_{\mathbf{x}}, 1, P^*, \Pi), & \text{if } E_{\mathbf{x}} > E_r(R, P^*) \\ \max_{\Pi} E_{\mathrm{e}}(R, E_{\mathbf{x}}, \alpha^*(R, E_{\mathbf{x}}), P^*, \Pi), & \text{if } E_{\mathbf{x}} \leq E_r(R, P^*) \end{array} \right\}. \tag{37}$$

For a given Π, $\Gamma(T,\Pi)$ is nonincreasing and convex in T, thus $\Gamma(T,\Pi)$ is strictly decreasing in T on a closed interval. Equivalent expressions for $\Gamma(T,\Pi)$ and boundaries of this closed interval are derived in Appendix A given in

$$\Gamma(T,\Pi) = \left\{\begin{array}{ll} \infty & \text{if } T < \mathsf{D}(U_0\| W_a|\Pi) \\ \mathsf{D}(U_s\| W_r|\Pi) & \text{if } \exists s\in[0,1]: T=\mathsf{D}(U_s\| W_a|\Pi) \\ \mathsf{D}(U_1\| W_r|\Pi) & \text{if } T > \mathsf{D}(U_1\| W_a|\Pi) \end{array}\right\} \tag{40}$$

where

$$U_s(y|x_1,x_2) = \left\{\begin{array}{ll} \frac{\mathbb{1}_{\{W(y|x_2)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x_2)>0} W(\tilde{y}|x_1)} W(y|x_1) & \text{if } s=0 \\ \frac{W(y|x_1)^{1-s}W(y|x_2)^s}{\sum_{\tilde{y}} W(\tilde{y}|x_1)^{1-s}W(\tilde{y}|x_2)^s} & \text{if } s\in(0,1) \\ \frac{\mathbb{1}_{\{W(y|x_1)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x_1)>0} W(\tilde{y}|x_2)} W(y|x_2) & \text{if } s=1 \end{array}\right\}.$$

For a $(R,E_{\mathbf{x}},\alpha,P,\Pi)$ such that $E_{\mathbf{x}} \le \alpha E_r(\frac{R}{\alpha},P)$, using the definition of $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ in (21) together with the (16), (38a)–(38c), and (40), we get (41), shown at the bottom of the page.

For any $(R,E_{\mathbf{x}},\alpha,P,\Pi)$, the above minimum is also achieved at a (Q,R_1,R_2,T) such that $R_1\ge R_2\ge R$. In order to see this take any minimizing (Q^*,R_1^*,R_2^*,T^*), then there are three possibilities:

a) $R_1^*\ge R_2^*\ge R$ claim holds trivially;
b) $R_1^* \ge R > R_2^*$, since $\boldsymbol{\zeta}(\frac{R_2}{\alpha},P,Q)$ is a nonincreasing function (Q^*,R_1^*,R,T^*), is also minimizing, thus the claim holds;
c) $R > R_1^* > R_2^*$, since $\boldsymbol{\zeta}(\frac{R}{\alpha},P,Q)$ is a nonincreasing function (Q^*,R,R,T^*), is also minimizing, thus the claim holds.

Thus, $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ is given by (42), shown at the bottom of the page.

Equation (42) is simplified further for symmetric channels as follows. Recall that for symmetric channels

$$E_{sp}(R) = \boldsymbol{\zeta}(R,P^*,Q^*) = \min_{Q} \boldsymbol{\zeta}(R,P^*,Q) \tag{43}$$

where P^* is the uniform input distribution and Q^* is the corresponding output distribution under W. Using an alternative expression for $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ given in (42) together with (37) and (43) for symmetric channels, we get (44), shown at the bottom of the next page, where $\alpha^*(R,E_{\mathbf{x}})$ is given in (33).

Although (43) does not hold in general using definition of $\boldsymbol{\zeta}(R,P,Q)$ and $E_{sp}(R,P)$, we can assert that

$$\boldsymbol{\zeta}(R,P,Q) \ge \min_{\tilde{Q}} \boldsymbol{\zeta}(R,P,\tilde{Q}) = E_{sp}(R,P). \tag{45}$$

Note that inequality given in (45) can be used to bound the minimized expression in (42) from below. In addition, recall that if the constraint set of a minimization is enlarged, then the resulting minimum cannot increase. We can use (45) also to

$$\boldsymbol{\zeta}(R,P,Q) = \left\{\begin{array}{ll} \infty & R < R_l^*(P,Q) \\ \min_{V:\substack{\mathsf{I}(P,V)\le R \\ (PV)_Y=Q}} \mathsf{D}(V\| W|P) & R\in[R_l^*(P,Q),R_h^*(P,Q)] \\ \min_{V:(PV)_Y=Q} \mathsf{D}(V\| W|P) & R > R_h^*(P,Q) \end{array}\right\} \tag{38a}$$

$$R_l^*(P,Q) = \min_{V:\substack{PV\gg PW \\ (PV)_Y=Q}} \mathsf{I}(P,V) \tag{38b}$$

$$R_h^*(P,Q) = \min_R \left\{ R : \min_{V:\substack{\mathsf{I}(P,V)\le R \\ (PV)_Y=Q}} \mathsf{D}(V\| W|P) = \min_{V:(PV)_Y=Q} \mathsf{D}(V\| W|P) \right\} \tag{38c}$$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \min_{\substack{Q,T,R_1,R_2: \\ R_1\ge R_2\ge 0,\ T\ge 0 \\ \alpha\boldsymbol{\zeta}(\frac{R_1}{\alpha},P,Q)+|R_2-R|^+ +T\le E_{\mathbf{x}}}} \alpha\boldsymbol{\zeta}\left(\frac{R_2}{\alpha},P,Q\right) + |R_1-R|^+ + (1-\alpha)\Gamma\left(\frac{T}{1-\alpha},\Pi\right). \tag{41}$$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \left\{\begin{array}{ll} \alpha E_r\left(\frac{R}{\alpha},P\right), & \text{if } E_{\mathbf{x}} > \alpha E_r\left(\frac{R}{\alpha},P\right) \\ \min_{\substack{Q,T,R_1,R_2: \\ R_1\ge R_2\ge R,\ T\ge 0 \\ \alpha\boldsymbol{\zeta}(\frac{R_1}{\alpha},P,Q)+R_2-R+T\le E_{\mathbf{x}}}} \alpha\boldsymbol{\zeta}\left(\frac{R_2}{\alpha},P,Q\right) + R_1 - R + (1-\alpha)\Gamma\left(\frac{T}{1-\alpha},\Pi\right), & \text{if } E_{\mathbf{x}} \le \alpha E_r\left(\frac{R}{\alpha},P\right) \end{array}\right\} \tag{42}$$

enlarge the constrained set of the minimization in (42). Thus, we get an exponent $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$, given in (46), shown at the bottom of the page, which is smaller than or equal to $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in all channels and for all rate erasure-exponent pairs.

After an investigation, very similar to the one we have already done for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in Section III-E, we conclude that $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ given in (47), shown at the bottom of the page, where $\alpha^*(R, E_{\mathbf{x}})$, $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$, and $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ are given in (33), (34), and (46), respectively, is an achievable error exponent for the for reliable sequences emerging from (46).

G. Special Cases

Zero Erasure-Exponent Case $\mathcal{E}_{\mathbf{e}}(R, 0)$: Using a simple repetition-at-erasures scheme, fixed-length errors-and-erasures codes can be converted into variable-length block codes, with the same error exponent. Thus, the error exponents of variable-length block codes given by Burnashev [3] are an upper bound to the error exponent of fixed-length block codes with erasures

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq \left(1 - \frac{R}{\mathcal{C}}\right) \mathcal{D} \quad \forall R \geq 0, E_{\mathbf{x}} \geq 0$$

where $\mathcal{D} = \max_{x, \tilde{x}} \sum_y W(y|x) \log \frac{W(y|x)}{W(y|\tilde{x})}$.

We show below that $\tilde{E}_{\mathbf{e}}(R, 0) \geq (1 - \frac{R}{\mathcal{C}})\mathcal{D}$. This implies that our coding scheme is optimal for $E_{\mathbf{x}} = 0$ for all rates, i.e., $\tilde{E}_{\mathbf{e}}(R, 0) = \mathcal{E}_{\mathbf{e}}(R, 0) = (1 - \frac{R}{\mathcal{C}})\mathcal{D}$.

Recall that, for all R less than the capacity, $\alpha^*(R, 0) = \frac{R}{\mathcal{C}}$. Furthermore, for any $\alpha \geq \frac{R}{\mathcal{C}}$

$$\mathcal{P}(R, 0, \alpha) = \left\{ P : \mathsf{I}(P, W) \geq \frac{R}{\alpha} \right\}.$$

Thus, for any $(R, 0, \alpha, P)$ such that $P \in \mathcal{P}(R, 0, \alpha)$, $R'' \geq R' \geq R$, $T \geq 0$, and $\alpha E_{sp}(\frac{R''}{\alpha}, P) + R' - R + T \leq 0$ imply that $R' = R$, $R'' = \alpha \mathsf{I}(P, W)$, $T = 0$. Consequently

$$\tilde{E}_{\mathbf{e}}(R, 0, \alpha, P, \Pi) = \alpha \left[E_{sp}\left(\frac{R}{\alpha}, P\right) + \mathsf{I}(P, W) - \frac{R}{\alpha} \right] + (1 - \alpha) \mathsf{D}(W_r \| W_a | \Pi).$$

When we maximize over Π and $P \in \mathcal{P}(R, 0, \alpha)$, we get

$$\tilde{E}_{\mathbf{e}}(R, 0, \alpha) = \max_{P \in \mathcal{P}(R, 0, \alpha)} \alpha E_{sp}\left(\frac{R}{\alpha}, P\right) + \alpha \mathsf{I}(P, W) - R + (1 - \alpha)\mathcal{D}$$

for all $\alpha \in [\frac{R}{\mathcal{C}}, 1]$. Simply inserting the minimum possible value of α, i.e., $\alpha^*(R, 0) = \frac{R}{\mathcal{C}}$, we get

$$\begin{aligned} \tilde{E}_{\mathbf{e}}\left(R, 0, \frac{R}{\mathcal{C}}\right) &= \max_{P \in \mathcal{P}\left(R, 0, \frac{R}{\mathcal{C}}\right)} \frac{R}{\mathcal{C}} E_{sp}(\mathcal{C}, P) + \frac{R}{\mathcal{C}} \mathsf{I}(P, W) - R + \left(1 - \frac{R}{\mathcal{C}}\right) \mathcal{D} \\ &= \left(1 - \frac{R}{\mathcal{C}}\right) \mathcal{D}. \end{aligned}$$

Thus, $\tilde{E}_{\mathbf{e}}(R, 0) \geq \left(1 - \frac{R}{\mathcal{C}}\right) \mathcal{D}$.

Indeed one need not to rely on the converse of variable-length block codes in order to establish the fact that $\tilde{E}_{\mathbf{e}}(R, 0) = (1 - \frac{R}{\mathcal{C}})\mathcal{D}$. The lower bound to the probability of the error presented in the next section not only recovers this particular optimality result but also upper bounds the optimal error exponent $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ as a function of rate R and erasure exponents $E_{\mathbf{x}}$.

Channels With Nonzero Zero-Error Capacity: For channels with a nonzero zero-error capacity, as a result of (21), $E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty$ for any $E_{\mathbf{x}} < E_r(R)$. This implies that we can get error-free block codes with this two-phase coding scheme

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \begin{cases} E_r(R), & \text{if } E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\Pi} \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R \; T \geq 0 \\ \alpha^* E_{sp}(\frac{R''}{\alpha^*}) + R' - R + T \leq E_{\mathbf{x}}}} \alpha^* E_{sp}\left(\frac{R'}{\alpha^*}\right) + R'' - R + (1 - \alpha^*)\Gamma\left(\frac{T}{1 - \alpha^*}, \Pi\right), & \text{if } E_{\mathbf{x}} \leq E_r(R) \end{cases} \tag{44}$$

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \begin{cases} \alpha E_r\left(\frac{R}{\alpha}, P\right), & \text{if } E_{\mathbf{x}} > \alpha E_r\left(\frac{R}{\alpha}, P\right) \\ \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R \; T \geq 0 \\ \alpha E_{sp}(\frac{R''}{\alpha}, P) + R' - R + T \leq E_{\mathbf{x}}}} \alpha E_{sp}\left(\frac{R'}{\alpha}, P\right) + R'' - R + (1 - \alpha)\Gamma\left(\frac{T}{1 - \alpha}, \Pi\right), & \text{if } E_{\mathbf{x}} \leq \alpha E_r\left(\frac{R}{\alpha}, P\right) \end{cases}. \tag{46}$$

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \begin{cases} E_r(R) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max\limits_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max\limits_{\Pi} \tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} \leq E_r(R) \end{cases} \tag{47}$$

for any rate $R < \mathcal{C}$ and any erasure exponent $E_{\mathbf{x}} \leq E_r(R)$. As we discuss in Section V in more detail, this is the best erasure exponent for rates over the critical rate, at least for the symmetric channels.

IV. An Outer Bound for Error-Exponent/Erasure-Exponent Tradeoff

In this section, we derive an upper bound on $\mathcal{E}_{\mathrm{e}}(R, E_{\mathbf{x}})$ using previously known results on the erasure-free block codes with feedback and a generalization of the straight-line bound of Shannon *et al.* [30]. We first present a lower bound on the minimum error probability of block codes with feedback and erasures, in terms of that of shorter codes in Section IV-A. Then, in Section IV-B, we make a brief overview of the outer bounds on the error exponents of erasure-free block codes with feedback. Finally, in Section IV-C, we use the relation we have derived in Section IV-A to tie the previously known results we have summarized in Section IV-B to bound $\mathcal{E}_{\mathrm{e}}(R, E_{\mathbf{x}})$.

A. A Trait of Minimum Error Probability of Block Codes With Erasures

Shannon *et al.* [30] considered fixed-length block codes, with a list decoding and established a family of lower bounds on the minimum error probability in terms of the product of minimum error probabilities of certain shorter codes. They have shown [30, Th. 1] that for fixed-length block codes with a list decoding and without feedback

$$\tilde{\mathcal{P}}_{\mathrm{e}}(M, \mathsf{n}, L) \geq \tilde{\mathcal{P}}_{\mathrm{e}}(M, \mathsf{n}_1, L_1)\tilde{\mathcal{P}}_{\mathrm{e}}(L_1+1, \mathsf{n}-\mathsf{n}_1, L) \quad (48)$$

where $\tilde{\mathcal{P}}_{\mathrm{e}}(M, \mathsf{n}, L)$ denotes the minimum error probability of erasure-free block codes of length n with M equally probable messages and with decoding list size L. As they have already pointed out in [30], this theorem continues to hold in the case when a feedback link is available from the receiver to the transmitter; although $\tilde{\mathcal{P}}_{\mathrm{e}}$'s are different when feedback is available, the relation given in (48) still holds. They were interested in erasure-free codes. On the other hand, we are interested in block codes which might have nonzero erasure probability. Accordingly, we need to incorporate erasure probability as one of the parameters of the optimal error probability. This is what this section is dedicated to.

In a size L list decoder with erasures, decoded set $\hat{\mathsf{M}}$ is either a subset[17] of $\mathcal{M}$ whose size is at most L, like the erasure-free case, or a set which only includes the erasure symbol, i.e., either $\hat{\mathsf{M}} \subset \mathcal{M}$ such that $|\hat{\mathsf{M}}| \leq L$ or $\hat{\mathsf{M}} = \{\mathbf{x}\}$. An erasure occurs whenever $\hat{\mathsf{M}} = \{\mathbf{x}\}$ and an error occurs whenever $\hat{\mathsf{M}} \neq \{\mathbf{x}\}$ and $\mathsf{M} \notin \hat{\mathsf{M}}$. We will denote the minimum error probability of length n block codes, with M equally probable messages, decoding list size L, and erasure probability $P_{\mathbf{x}}$ by $\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$.

Theorem 2 bounds the error probability of block codes with erasures and list decoding using the error probabilities of shorter codes with erasures and list decoding, as [30, Th. 1] does in the erasure-free case. As its counterpart in the erasure-free case, Theorem 2 is later used to establish the outer bounds to the error exponents.

Theorem 2: For any $\mathsf{n}, M, L, P_{\mathbf{x}}, \mathsf{n}_1 \leq \mathsf{n}, L_1$, and $0 \leq s \leq 1$, the minimum error probability of fixed-length block codes with feedback satisfies

$$\begin{aligned}(1-s)&\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})\\ &\geq \mathcal{P}_{\mathrm{e}}(M, \mathsf{n}_1, L_1, s)\\ &\quad\times \mathcal{P}_{\mathrm{e}}\left(L_1+1, \mathsf{n}-\mathsf{n}_1, L, \frac{(1-s)P_{\mathbf{x}}}{\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}_1, L_1, s)}\right). \quad (49)\end{aligned}$$

Note that given a (M, n, L) triple if the error-probability/erasure-probability pairs $(P_{\mathrm{e}1}, P_{\mathbf{x}1})$ and $(P_{\mathrm{e}2}, P_{\mathbf{x}2})$ are achievable, then for any $\gamma \in [0,1]$ using the initial symbol A_0 of the feedback link we can construct a code that uses the code achieving $(P_{\mathrm{e}1}, P_{\mathbf{x}1})$ with probability γ and the code achieving $(P_{\mathrm{e}2}, P_{\mathbf{x}2})$ with probability $(1-\gamma)$. This new code achieves error-probability/erasure-probability pair $(\gamma P_{\mathrm{e}a} + (1-\gamma)P_{\mathrm{e}b}, \gamma P_{\mathbf{x}a} + (1-\gamma)P_{\mathbf{x}b})$. As a result, for any (M, n, L) triple, the set of achievable error probability erasure probability pairs is convex. We use this fact twice in order to prove Theorem 2.

Let us first consider the following lemma which bounds the achievable error-probability/erasure-probability, pairs for block codes with nonuniform *a priori* probability distribution, in terms of block codes with a uniform *a priori* probability distribution but fewer messages.

Lemma 2: For any length n block code with message set $\mathcal{M}$, *a priori* probability distribution $\varphi(\cdot)$ on $\mathcal{M}$, erasure probability $P_{\mathbf{x}}$, decoding list size L, and integer K

$$P_{\mathrm{e}} \geq \Omega(\varphi, K)\,\mathcal{P}_{\mathrm{e}}\left(K+1, \mathsf{n}, L, \frac{P_{\mathbf{x}}}{\Omega(\varphi, K)}\right) \quad (50)$$

where $\Omega(\varphi, K) = \min_{\mathcal{S}\subset\mathcal{M}, |\mathcal{S}|\leq K} \varphi(\mathcal{S}^{\mathrm{c}})$ and $\mathcal{S}^{\mathrm{c}} = \mathcal{M}/\mathcal{S}$.

Recall that $\mathcal{P}_{\mathrm{e}}(K+1, \mathsf{n}, L, P_{\mathbf{x}})$ is the minimum error probability of length n codes with $(K+1)$ equally probable messages and decoding list size L, with feedback if the original code does have feedback, and without feedback if the original code does not.

Note that $\Omega(\varphi, K)$ is the error probability of a decoder which decodes to the set of K most likely messages under φ. In other words, $\Omega(\varphi, K)$ is the minimum error probability for a size K list decoder when the posterior probability distribution is φ.

Proof: If $\Omega(\varphi, K) = 0$, the lemma holds trivially. Thus, we assume $\Omega(\varphi, K) > 0$ henceforth. For any size $(K+1)$ subset $\mathcal{M}'$ of $\mathcal{M}$, one can use the encoding scheme and the decoding rule of the original code for $\mathcal{M}$, to construct the following block code for $\mathcal{M}'$.

- **Encoder**: $\forall m \in \mathcal{M}'$ use the encoding scheme for message m in the original code, i.e., for all $m \in \mathcal{M}'$, $t \in [1, \mathsf{n}]$, and $z^{t-1} \in \mathcal{Z}^{t-1}$

$$X_t'(m, z^{t-1}) = X_t(m, z^{t-1}).$$

[17]Note that if $\hat{\mathsf{M}} \subset \mathcal{M}$, then $\mathbf{x} \notin \hat{\mathsf{M}}$, because $\mathbf{x} \notin \mathcal{M}$.

- **Decoder**: For all $z^{\mathsf{n}} \in \mathcal{Z}^{\mathsf{n}}$ if the original decoding rule declares erasure, declare erasure; else decode to the intersection of the original decoded list and $\mathcal{M}'$

$$\hat{\mathsf{M}}' = \begin{cases} \mathbf{x}, & \text{if } \hat{\mathsf{M}} = \mathbf{x} \\ \hat{\mathsf{M}} \cap \mathcal{M}', & \text{else.} \end{cases}$$

This is a length n code with $(K+1)$ messages and decoding list size L. Furthermore, for all m in $\mathcal{M}'$, the conditional error probability $P'_{\mathrm{e}|m}$ and the conditional erasure probability $P'_{\mathbf{x}|m}$ are equal to the conditional error probability $P_{\mathrm{e}|m}$ and the conditional erasure probability $P_{\mathbf{x}|m}$ in the original code, respectively.

Note that for all $\mathcal{M}' \subset \mathcal{M}$ such that $|\mathcal{M}'| = K+1$

$$\frac{1}{K+1} \sum\nolimits_{m \in \mathcal{M}'} \left(P_{\mathrm{e}|m}, P_{\mathbf{x}|m}\right) \in \Psi(K+1, \mathsf{n}, L) \quad (51)$$

where $\Psi(K+1, \mathsf{n}, L)$ is the set of the achievable error probability, erasure probability pairs for length n block codes with $(K+1)$ equally probable messages, and with decoding list size L.

Let the smallest nonzero element of $\{\varphi(1), \varphi(2), \ldots \varphi(|\mathcal{M}|)\}$ be $\varphi(\xi_1)$. For any size $(K+1)$ subset of $\mathcal{M}$, which includes ξ_1, and all whose elements have nonzero probabilities, say $\mathcal{M}_1$, we have

$$\begin{aligned}(P_{\mathrm{e}}, P_{\mathbf{x}}) &= \sum\nolimits_{m \in \mathcal{M}} \varphi(m)(P_{\mathrm{e}|m}, P_{\mathbf{x}|m}) \\ &= \sum\nolimits_{m \in \mathcal{M}} [\varphi(m) - \varphi(\xi_1)\mathbb{1}_{\{m \in \mathcal{M}_1\}}](P_{\mathrm{e}|m}, P_{\mathbf{x}|m}) \\ &\quad + \varphi(\xi_1) \sum\nolimits_{m \in \mathcal{M}_1} (P_{\mathrm{e}|m}, P_{\mathbf{x}|m}).\end{aligned}$$

Equation (51) and the definition of $\Psi(K+1, \mathsf{n}, L)$ imply that $\exists \psi_1 \in \Psi(K+1, \mathsf{n}, L)$ such that

$$(P_{\mathrm{e}}, P_{\mathbf{x}}) = \sum\nolimits_{m \in \mathcal{M}} \varphi^{(1)}(m)(P_{\mathrm{e}|m}, P_{\mathbf{x}|m}) + \varphi(\psi_1)\psi_1 \quad (52)$$

$$1 = \varphi(\psi_1) + \sum\nolimits_{m \in \mathcal{M}} \varphi^{(1)}(m) \quad (53)$$

where

$$\varphi(\psi_1) = (K+1)\varphi(\xi_1) \quad \varphi^{(1)}(m) = \varphi(m) - \varphi(\xi_1)\mathbb{1}_{\{m \in \mathcal{M}_1\}}.$$

Furthermore, the number of nonzero $\varphi^{(1)}(m)$'s is at least one fewer than that of nonzero $\varphi(m)$'s. The remaining probabilities $\varphi^{(1)}(m)$ have a minimum $\varphi^{(1)}(\xi_2)$ among their nonzero elements. One can repeat the same argument once more using that element and reduce the number of nonzero elements to at least one more. After at most $|\mathcal{M}| - K$ such iterations, one reaches to $\varphi^{(\ell)}$, which is nonzero for K or fewer messages

$$(P_{\mathrm{e}}, P_{\mathbf{x}}) = \sum_{j=1}^{\ell} \varphi(\psi_j)\psi_j + \sum_{m \in \mathcal{M}} \varphi^{(\ell)}(m)(P_{\mathrm{e}|m}, P_{\mathbf{x}|m}) \quad (54)$$

where

$$\varphi^{(\ell)}(m) \le \varphi(m) \quad \forall m \in \mathcal{M} \quad \text{and} \quad \sum_{m \in \mathcal{M}} \mathbb{1}_{\{\varphi^{(\ell)}(m) > 0\}} \le K.$$

In (54), the first sum is equal to a convex combination of ψ_j's multiplied by $\sum_{j=1}^{\ell} \varphi(\psi_j)$; the second sum is equal to a pair with nonnegative entries. As a result of the definition of $\Omega(\varphi, K)$ given in (50)

$$\Omega(\varphi, K) \le \sum_{j=1}^{\ell} \varphi(\psi_j). \quad (55)$$

Then, as a result of convexity of $\Psi(K+1, \mathsf{n}, L)$, we can conclude that there exists a $\psi \in \Psi(K+1, \mathsf{n}, L)$ such that $(P_{\mathrm{e}}, P_{\mathbf{x}}) = a\Omega(\varphi, K)\tilde{\psi} + (b_1, b_2)$ for some $a \ge 1$, $b_1 \ge 0$, and $b_2 \ge 0$. Thus

$$\exists \psi \in \Psi(K+1, \mathsf{n}, L) : \left(\frac{P_{\mathrm{e}}}{\Omega(\varphi, K)}, \frac{P_{\mathbf{x}}}{\Omega(\varphi, K)}\right) = \psi + (b_3, b_4) \quad (56)$$

for some $b_3 \ge 0, b_4 \ge 0$.

Then, the lemma follows from (56), the fact that $\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$ is decreasing in $P_{\mathbf{x}}$, and the fact that $\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$ is uniquely determined by $\Psi(M, \mathsf{n}, L)$ for $s_{\mathbf{x}} \in [0, 1]$ for all $(M, \mathsf{n}, L, \psi_{\mathbf{x}})$ as follows:

$$\mathcal{P}_{\mathrm{e}}(M, \mathsf{n}, L, P_{\mathbf{x}}) = \min_{\psi_{\mathbf{x}} : (\psi_{\mathrm{e}}, \psi_{\mathbf{x}}) \in \Psi(M, \mathsf{n}, L)} \psi_{\mathrm{e}}. \quad (57)$$

□

For proving Theorem 2, we express the error-and-erasure probabilities as a convex combination of error-and-erasure probabilities of $(\mathsf{n} - \mathsf{n}_1)$ long block codes with *a priori* probability distribution $\varphi_{z^{\mathsf{n}_1}}(m) = \mathbf{P}[m|\, z^{\mathsf{n}_1}]$ over the messages, and apply Lemma 2 together with convexity arguments similar to the ones above.

Proof of Theorem 2: For all m in $\mathcal{M}$, let $\Upsilon(m)$ be the decoding region of m, let $\Upsilon(\mathbf{x})$ be the decoding region of the erasure symbol $\mathbf{x}$, and let $\tilde{\Upsilon}(m)$ be the error region of m

$$\begin{aligned}\Upsilon(m) &\triangleq \{z^{\mathsf{n}} : m \in \hat{\mathsf{M}}\} \\ \Upsilon(\mathbf{x}) &\triangleq \{z^{\mathsf{n}} : \mathbf{x} \in \hat{\mathsf{M}}\} \\ \tilde{\Upsilon}(m) &\triangleq \Upsilon(m)^c \cap \Upsilon(\mathbf{x})^c\end{aligned}$$

where $\Upsilon^c = \mathcal{Z}^{\mathsf{n}}/\Upsilon$. Then, for all $m \in \mathcal{M}$. Then

$$(P_{\mathrm{e}|m}, P_{\mathbf{x}|m}) = \left(\mathbf{P}\left[\tilde{\Upsilon}(m)\middle|\, m\right], \mathbf{P}\left[\Upsilon(\mathbf{x})|\, m\right]\right). \quad (58)$$

Note that[18]

$$\begin{aligned}P_{\mathbf{x}|m} &= \sum_{z^{\mathsf{n}} : z^{\mathsf{n}} \in \Upsilon(\mathbf{x})} \mathbf{P}\left[z^{\mathsf{n}}|\, m\right] \\ &= \sum_{z^{\mathsf{n}_1}} \mathbf{P}\left[z^{\mathsf{n}_1}|\, m\right] \sum_{z^{\mathsf{n}}_{\mathsf{n}_1+1} : (z^{\mathsf{n}_1}, z^{\mathsf{n}}_{\mathsf{n}_1+1}) \in \Upsilon(\mathbf{x})} \mathbf{P}\left[z^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, z^{\mathsf{n}_1}\right].\end{aligned}$$

Then, the erasure probability is

$$P_{\mathbf{x}} = \sum_{m \in \mathcal{M}} \frac{1}{|\mathcal{M}|} \sum_{z^{\mathsf{n}_1}} \mathbf{P}\left[z^{\mathsf{n}_1}|\, m\right] \sum_{z^{\mathsf{n}}_{\mathsf{n}_1+1} : (z^{\mathsf{n}_1}, z^{\mathsf{n}}_{\mathsf{n}_1+1}) \in \Upsilon(\mathbf{x})} \mathbf{P}\left[z^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, z^{\mathsf{n}_1}\right]$$

[18]There is a slight abuse of notation here: if A's include real-valued random variables with densities, we should integrate, rather than sum, over them. Since it is clear from the context what needs to be done we omit that subtlety in the following calculations.

$$= \sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right] \sum_{m \in \mathcal{M}} \mathbf{P}\left[m \middle| z^{n_1}\right] \sum_{z^{n}_{n_1+1}:(z^{n_1},z^{n}_{n_1+1})\in\Upsilon(\mathbf{x})} \mathbf{P}\left[z^{n}_{n_1+1}\middle| m, z^{n_1}\right]$$
$$= \sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right] P_{\mathbf{x}}(z^{n_1}).$$

Note that for every z^{n_1}, $P_{\mathbf{x}}(z^{n_1})$ is the erasure probability of a code of length $(n - n_1)$ with *a priori* probability distribution $\varphi_{z^{n_1}}(m) = \mathbf{P}[m|z^{n_1}]$. Furthermore, one can write the error probability $P_{\mathbf{e}}$ as

$$P_{\mathbf{e}} = \sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right] \sum_{m \in \mathcal{M}} \mathbf{P}\left[m \middle| z^{n_1}\right] \sum_{z^{n}_{n_1+1}:(z^{n_1},z^{n}_{n_1+1})\in\tilde{\Upsilon}(m)} \mathbf{P}\left[z^{n}_{n_1+1}\middle| m, z^{n_1}\right]$$
$$= \sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right] P_{\mathbf{e}}(z^{n_1})$$

where $P_{\mathbf{e}}(z^{n_1})$ is the error probability of the very same length $(n - n_1)$ code. As a result of Lemma 2, the pair $(P_{\mathbf{e}}(z^{n_1}), P_{\mathbf{x}}(z^{n_1}))$ satisfies

$$P_{\mathbf{e}}(z^{n_1}) \geq \Omega\left(\varphi_{z^{n_1}}, L_1\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (n-n_1), L, \frac{P_{\mathbf{x}}(z^{n_1})}{\Omega\left(\varphi_{z^{n_1}}, L_1\right)}\right). \tag{59}$$

Then, using the convexity[19] of $\mathcal{P}_{\mathbf{e}}(M, n, L, P_{\mathbf{x}})$ in $P_{\mathbf{x}}$, we conclude that for any $s \in [0,1]$, $P_{\mathbf{e}}$ satisfies (60), shown at the bottom of the page.

Now consider a code which uses the first n_1 time units of the original encoding scheme as its encoding scheme. Decoder of this new code draws a real number from $[0,1]$ uniformly at random, independently of Z^{n_1} of the original code (and the message evidently). If this number is less than s, it declares erasure; else it makes a maximum-likelihood decoding with list of size L_1. Then, the sum on the left-hand side of (61) is its error probability. But that probability is lower bounded by $\mathcal{P}_{\mathbf{e}}(M, n_1, L_1, s)$, which is the minimum error probability over all length n_1 block codes with M messages and decoding list size L_1, i.e.,

$$\sum_{z^{n_1}} \mathbf{P}[z^{n_1}](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right) \geq \mathcal{P}_{\mathbf{e}}\left(M, n_1, L_1, s\right). \tag{61}$$

Then, the theorem follows from (60), (61) and the fact that $\mathcal{P}_{\mathbf{e}}(M, n, L_1, P_{\mathbf{x}})$ is a decreasing function of $P_{\mathbf{x}}$. □

As the result of Shannon *et al.* [30, Th. 1], Theorem 2 is correct both with and without feedback. Although $\mathcal{P}_{\mathbf{e}}$'s are different in each case, the relationship between them given in (49) holds in both cases.

B. Classical Results on Error Exponent of Erasure-Free Block Codes With Feedback

In this section, we give a very brief overview of the previously known results on the error probability of erasure-free block codes with feedback. These results are used in Section IV-C together with Theorem 2 to bound $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ from above. Note that Theorem 2 only relates the error probability of longer codes to that of the shorter ones. It does not in and of itself bound the error probability. It is in a sense a tool to glue together various bounds on the error probability.

First bound we consider is on the error exponent of erasure-free block codes with feedback. Haroutunian [17] proved that, for any (M_n, n, L_n) sequence of triples, such that $\lim_{n\to\infty} \frac{\ln M_n - \ln L_n}{n} = R$

$$\lim_{n\to\infty} \frac{-\ln \mathcal{P}_{\mathbf{e}}(M_n, n, L_n, 0)}{n} \leq E_H(R) \tag{62}$$

where

$$E_H(R) = \min_{V:\mathcal{C}(V)\leq R} \max_P \mathsf{D}\left(V \| W | P\right) \tag{63a}$$
$$\mathcal{C}(V) = \max_P \mathsf{I}\left(P, V\right). \tag{63b}$$

Second bound we consider is on the tradeoff between the error exponents of two messages in a two-message erasure-free block code with feedback. Berlekamp mentions this result in passing [1] and attributes it to Gallager and Shannon.

Lemma 3: For any feedback encoding scheme with two messages and erasure-free decision rule and for all $T \geq T_0$

$$\text{either} \qquad P_{\mathbf{e}1} \geq \frac{1}{4} e^{-nT + \sqrt{n}4 \ln P_{\min}} \tag{64a}$$
$$\text{or} \qquad P_{\mathbf{e}2} > \frac{1}{4} e^{-n\Gamma(T) + \sqrt{n}4 \ln P_{\min}} \tag{64b}$$

where $P_{\min} = \min_{x,y:W(y|x)>0} W(y|x)$

$$T_0 \triangleq \max_{x,\tilde{x}} -\ln \sum_{y:W(y|\tilde{x})>0} W(y|x) \tag{65}$$
$$\Gamma\left(T\right) \triangleq \max_{\Pi} \Gamma\left(T, \Pi\right). \tag{66}$$

Result is old and somewhat intuitive to those who are familiar with the calculations in the nonfeedback case. Thus, probably it

[19] The convexity of $\mathcal{P}_{\mathbf{e}}(M, n, L, P_{\mathbf{x}})$ in $P_{\mathbf{x}}$ follows from (57) and the convexity of the region $\Psi(M, n, L)$.

$$(1-s)P_{\mathbf{e}} \geq \sum\nolimits_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right)\mathcal{P}_{\mathbf{e}}\left(L_1+1, (n-n_1), L, \frac{P_{\mathbf{x}}(z^{n_1})}{\Omega\left(\varphi_{z^{n_1}}, L_1\right)}\right)$$
$$\geq \left(\sum\nolimits_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right)\right)\mathcal{P}_{\mathbf{e}}\left(L_1+1, (n-n_1), L, \frac{\sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)P_{\mathbf{x}}(z^{n_1})}{\sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right)}\right)$$
$$= \left(\sum\nolimits_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right)\right)\mathcal{P}_{\mathbf{e}}\left(L_1+1, (n-n_1), L, \frac{(1-s)P_{\mathbf{x}}}{\sum_{z^{n_1}} \mathbf{P}\left[z^{n_1}\right](1-s)\Omega\left(\varphi_{z^{n_1}}, L_1\right)}\right). \tag{60}$$

has been proven a number of times. But we are not aware of a published proof, hence we have included one in Appendix A.

Although Lemma 3 establishes only the converse part, $(T, \Gamma(T))$ is indeed the optimal tradeoff for the error exponents of two messages in an erasure-free block code, both with and without feedback. Achievablity of this tradeoff has already been established in [30, Th. 5] for the case without feedback; evidently this implies the achievablity with feedback. Furthermore, T_0 does have an operational meaning; it is the maximum error exponent first message can have, while the second message has zero-error probability. This fact is also proved in Appendix A.

For some channels Lemma 3 gives us a bound on the error exponent of erasure-free codes at zero rate, which is tighter than Haroutunian's bound at zero rate. In order to see this let us first define T^* to be

$$T^* = \max_T \min\{T, \Gamma(T)\}. \tag{67}$$

Note that T^* is finite iff $\sum_y W(y|x)W(y|\tilde{x}) > 0$ for all x, $\tilde{x}$ pairs. Recall that this is also the necessary and sufficient condition of zero-error capacity $\mathcal{C}_0$ to be zero. $E_H(R)$ on the other hand is infinite for all $R \leq R_\infty$ like $E_{sp}(R)$ where R_∞ is given by

$$R_\infty = -\min_{P(\cdot)} \max_y \ln \sum_{x:W(y|x)>0} P(x). \tag{68}$$

Even in the cases where $E_H(0)$ is finite, $E_H(0) \geq T^*$. We can use this fact, Lemma 3, and Theorem 2, or [30, Th. 1] for that matter, to strengthen Haroutunian bound at low rates, as follows.

Lemma 4: For all channels with zero zero-error capacity, $\mathcal{C}_0 = 0$ and any sequence of M_n, such that $\lim_{\mathsf{n}\to\infty} \frac{\ln M_\mathsf{n}}{\mathsf{n}} = R$

$$\lim_{\mathsf{n}\to\infty} \frac{-\ln \mathcal{P}_\mathbf{e}(M_\mathsf{n}, \mathsf{n}, 1, 0)}{\mathsf{n}} \leq \tilde{E}_H(R) \tag{69}$$

where

$$\tilde{E}_H(R) = \left\{ \begin{array}{lll} E_H(R), & \text{if} & R \geq R_{ht} \\ T^* + \frac{E_H(R_{ht})-T^*}{R_{ht}}R, & \text{if} & R \in [0, R_{ht}) \end{array} \right\}$$

and R_{ht} is the unique solution of the equation $T^* = E_H(R) - RE'_H(R)$ if it exists; $R_{ht} = \mathcal{C}$ otherwise.

Before going into the proof let us note that $\tilde{E}_H(R)$ is obtained simply by drawing the tangent line to the curve $(R, E_H(R))$ from the point $(0, T^*)$. The curve $(R, \tilde{E}_H(R))$ is the same as the tangent line, for the rates between 0 and R_{ht}, and it is the same as the curve $(R, E_H(R))$ from then on where R_{ht} is the rate of the point at which the tangent from $(0, T^*)$ meets the curve $(R, E_H(R))$.

Proof: For $R \geq R_{ht}$, this Lemma immediately follows from Haroutunian's result [17] for $L_1 = 1$. If $R < R_{ht}$, then we apply Theorem 2

$$\begin{aligned}(1-s)&\mathcal{P}_\mathbf{e}(M, \mathsf{n}, L_1, P_\mathbf{x}) \\ &\geq \mathcal{P}_\mathbf{e}(M, \tilde{\mathsf{n}}, L_1, s)\mathcal{P}_\mathbf{e}\left(L_1+1, \mathsf{n}-\tilde{\mathsf{n}}, \tilde{L}, \frac{(1-s)P_\mathbf{x}}{\mathcal{P}_\mathbf{e}(M, \mathsf{n}, L_1, s)}\right)\end{aligned} \tag{70}$$

with[20] $s = 0$, $P_\mathbf{x} = 0$, $L_1 = 1$, and $\tilde{\mathsf{n}} = \lfloor \frac{R}{R_{ht}} \rfloor$. Furthermore, by Lemma 3 and the definition of T^* given in (67), we have

$$\mathcal{P}_\mathbf{e}(2, \mathsf{n}-\tilde{\mathsf{n}}, L, 0) \geq \frac{e^{-(\mathsf{n}-\tilde{\mathsf{n}})T^* + \sqrt{\mathsf{n}-\tilde{\mathsf{n}}}\ln P_{\min}}}{8}. \tag{71}$$

Using (70) and (71), we get

$$\begin{aligned}\frac{-\ln \mathcal{P}_\mathbf{e}(M, \mathsf{n}, 1, 0)}{\mathsf{n}} &\leq \frac{-\ln \mathcal{P}_\mathbf{e}(M, \tilde{\mathsf{n}}, 1, 0)}{\tilde{\mathsf{n}}} \frac{R}{R_{ht}} + \left[1 - \frac{R}{R_{ht}} + \frac{1}{\tilde{\mathsf{n}}}\right] T^* \\ &\quad + \left(\sqrt{\frac{1}{\tilde{\mathsf{n}}}}\right)\left(\sqrt{\frac{R_{ht}-R}{R_{ht}}}\right) \ln \frac{P_{\min}}{8}\end{aligned}$$

where $\frac{\ln M_\mathsf{n}}{\tilde{\mathsf{n}}} = R_{ht}$. The lemma follows by simply applying Haroutunian's result to the first terms on the right-hand side. □

C. Generalized Straight-Line Bound for Block Codes With Erasures

Theorem 2 bounds the minimum error probability length n block codes from below in terms of the minimum error probability of length n_1 and length $(\mathsf{n}-\mathsf{n}_1)$ block codes. The rate and erasure probability of the longer code constraints the rates and erasure probabilities of the shorter ones, but does not specify them completely. We use this fact together with the improved Haroutunian's bound on the error exponents of erasure-free block codes with feedback, i.e., Lemma 4, and the error-exponent tradeoff of the erasure-free feedback block codes with two messages, i.e., Lemma 3, to obtain a family of upper bounds on the error exponents of feedback block codes with erasure.

Theorem 3: For any DMC with $\mathcal{C}_0 = 0$ rate $R \in [0, \mathcal{C}]$ and $E_\mathbf{x} \in [0, E_H(R)]$ and for any $r \in [r_h(R, E_\mathbf{x}), \mathcal{C}]$

$$\mathcal{E}_\mathbf{e}(R, E_\mathbf{x}) \leq \frac{R}{r}\tilde{E}_H(r) + \left(1 - \frac{R}{r}\right)\Gamma\left(\frac{E_\mathbf{x} - \frac{R}{r}\tilde{E}_H(r)}{1 - \frac{R}{r}}\right)$$

where $r_h(R, E_\mathbf{x})$, is the unique solution of

$$R\tilde{E}_H(r) - rE_\mathbf{x} = 0.$$

Theorem 3 simply states that any line connecting any two points of the curves $(R, E_\mathbf{x}, E_\mathbf{e}) = (R, \tilde{E}_H(R), \tilde{E}_H(R))$ and $(R, E_\mathbf{x}, E_\mathbf{e}) = (0, E_\mathbf{x}, \Gamma(E_\mathbf{x}))$ lies above the surface $(R, E_\mathbf{x}, E_\mathbf{e}) = (R, E_\mathbf{x}, \mathcal{E}_\mathbf{e}(R, E_\mathbf{x}))$. The condition $\mathcal{C}_0 = 0$ is not merely a technical condition due to the proof technique; as we will see in Section V for channels with $\mathcal{C}_0 > 0$, there are zero-error codes with erasure exponent as high as $E_{sp}(R)$ for any rate $R \leq \mathcal{C}$.

Proof: We will consider the cases $r \in (r_h(R, E_\mathbf{x}), \mathcal{C}]$ and $r = r_h(R, E_\mathbf{x})$ separately.

- $r \in (r_h(R, E_\mathbf{x}), \mathcal{C}]$: For any $(M, \mathsf{n}, P_\mathbf{x})$ sequence such that $\liminf_{\mathsf{n}\to\infty} \frac{\ln M}{\mathsf{n}} = R$, $\liminf_{\mathsf{n}\to\infty} \frac{-\ln P_\mathbf{x}}{\mathsf{n}} = E_\mathbf{x}$, if we choose $\mathsf{n}_1 = \lfloor \frac{R}{r}\mathsf{n} \rfloor$, as a result of Lemma 4, we know that there exists a ϱ_n sequence such that

$$\varrho_\mathsf{n} \leq \mathcal{P}_\mathbf{e}(M, \mathsf{n}_1, 1, 0) \quad \text{and} \quad \lim_{\mathsf{n}\to\infty} -\frac{\ln \varrho_\mathsf{n}}{\mathsf{n}_1} = \tilde{E}_H(r)$$

[20]Or [30, Th. 1] with $L_1 = 1$ and $\mathsf{n}_1 = \lfloor \frac{R}{R_{ht}} \rfloor$.

Theorem 2 for $s=0$, $L=1$, $L_1=1$ implies

$$\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}\right) \times \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0) \geq \mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\varrho_{\mathsf{n}}}\right)\varrho_{\mathsf{n}}$$

where the second inequality follows from the fact that $\mathcal{P}_{\mathbf{e}}(2,\mathsf{n}-\mathsf{n}_1,1,s)$ is a decreasing function of s. If we take the logarithm of both sides and divide by n, we get

$$\frac{\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,P_{\mathbf{x}})}{\mathsf{n}} \geq \frac{\mathsf{n}-\mathsf{n}_1}{\mathsf{n}} \frac{\ln \mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\varrho_{\mathsf{n}}}\right)}{\mathsf{n}-\mathsf{n}_1} + \frac{\mathsf{n}_1}{\mathsf{n}}\frac{\ln \varrho_{\mathsf{n}}}{\mathsf{n}_1}. \tag{72}$$

Since $r > r_h(R,E_{\mathbf{x}})$ and $\tilde{E}_H(R)$ is convex, we have

$$0 < \liminf_{\mathsf{n}\longrightarrow\infty} \frac{-1}{\mathsf{n}-\mathsf{n}_1} \ln \frac{P_{\mathbf{x}}}{\varrho_{\mathsf{n}}} \leq T^*.$$

Assume for the moment for any $T\in(0,T^*]$ and sequence of $P_{\mathbf{x}}^{(\mathsf{n})}$ such that $\liminf_{\mathsf{n}\longrightarrow\infty}\frac{-\ln P_{\mathbf{x}}{}^{(\mathsf{n})}}{\mathsf{n}} = T$, we have

$$\liminf_{\mathsf{n}\longrightarrow\infty} \frac{-\ln \mathcal{P}_{\mathbf{e}}(2,\mathsf{n},1,P_{\mathbf{x}}{}^{(\mathsf{n})})}{\mathsf{n}} \leq \Gamma(T). \tag{73}$$

Then, Theorem 3 follows from (72) if we take the limit as n goes to infinity and use the fact that $\Gamma(T)$ is a nondecreasing function of T.
In order to establish (73), note that if $T_0>0$ and $T\leq T_0$, then $\Gamma(T)=\infty$. Thus, (73) holds trivially. For $T>T_0$ case, we prove (73) by contradiction. Assume that (73) is wrong. Then, there exist an $\epsilon>0$ and a block code with erasures that satisfies

$$\begin{aligned}\mathbf{P}\left[\tilde{\Upsilon}(\tilde{m})\middle|\, \tilde{m}\right] &\leq e^{-\mathsf{n}(\Gamma(T)+\epsilon+o(1))}\\ \mathbf{P}\left[\Upsilon(\mathbf{x})\middle|\, \tilde{m}\right] &\leq e^{-\mathsf{n}(T+o(1))}\\ \mathbf{P}\left[\tilde{\Upsilon}(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] &\leq e^{-\mathsf{n}(\Gamma(T)+\epsilon+o(1))}\\ \mathbf{P}\left[\Upsilon(\mathbf{x})\middle|\, \tilde{\tilde{m}}\right] &\leq e^{-\mathsf{n}(T+o(1))}.\end{aligned}$$

Enlarge the decoding region of $\tilde{m}$ by taking its union with the erasure region

$$\Upsilon'(\tilde{m}) = \Upsilon(\tilde{m})\cup\Upsilon(\mathbf{x}) \quad \Upsilon'(\tilde{\tilde{m}}) = \Upsilon(\tilde{\tilde{m}}) \quad \Upsilon'(\mathbf{x}) = \emptyset.$$

The resulting code is an erasure-free code with

$$\begin{aligned}\mathbf{P}\left[\Upsilon'(\tilde{\tilde{m}})\middle|\, \tilde{m}\right] &\leq e^{-\mathsf{n}(\Gamma(T)+\epsilon+o(1))}\\ \mathbf{P}\left[\Upsilon'(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] &\leq e^{-\mathsf{n}(\min\{\Gamma(T),T\}+o(1))}.\end{aligned}$$

Since $T_0 < T \leq T^*$, $\Gamma(T)\geq T$, this contradicts with Lemma 3, thus (73) holds.

- $r = r_h(R,E_{\mathbf{x}})$: Apply Theorem 2 with $s=0$, $L=1$, $L_1=1$, and $\mathsf{n}_1 = \max\{\ell : \mathcal{P}_{\mathbf{e}}(M,\ell,1,0) > P_{\mathbf{x}}\ln\frac{1}{P_{\mathbf{x}}}\}$

$$\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)\mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}\right) \geq P_{\mathbf{x}}\ln\frac{1}{P_{\mathbf{x}}}\mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{1}{-\ln P_{\mathbf{x}}}\right). \tag{74}$$

Note that for $\mathsf{n}_1 = \max\{\ell : \mathcal{P}_{\mathbf{e}}(M,\ell,1,0) > P_{\mathbf{x}}\ln\frac{1}{P_{\mathbf{x}}}\}$

$$\liminf_{\mathsf{n}\longrightarrow\infty}\frac{\mathsf{n}_1}{\mathsf{n}}\mathcal{E}\left(\frac{R\mathsf{n}}{\mathsf{n}_1}\right) = E_{\mathbf{x}}.$$

Then, as a result of Lemma 4, we have

$$\liminf_{\mathsf{n}\longrightarrow\infty}\frac{\mathsf{n}_1}{\mathsf{n}}\tilde{E}_H\left(\frac{R\mathsf{n}}{\mathsf{n}_1}\right) \geq E_{\mathbf{x}}.$$

Then

$$\liminf_{\mathsf{n}\longrightarrow\infty}\frac{\mathsf{n}_1}{\mathsf{n}} \geq \frac{R}{r_h(R,E_{\mathbf{x}})}. \tag{75}$$

Assume for the moment that for any ϵ_{n} such that $\liminf_{\mathsf{n}\longrightarrow\infty}\epsilon_{\mathsf{n}} = 0$

$$\liminf_{\mathsf{n}\longrightarrow\infty}\frac{-\ln\mathcal{P}_{\mathbf{e}}(2,\mathsf{n},1,\epsilon_{\mathsf{n}})}{\mathsf{n}} \leq \Gamma(0). \tag{76}$$

Then, taking the logarithm of both sides of (74), dividing both sides by n, taking the limit as n tends to infinity, and substituting (75) and (76), we get

$$\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}}) \leq E_{\mathbf{x}} + \left(1-\frac{R}{r_h(R,E_{\mathbf{x}})}\right)\Gamma(0). \tag{77}$$

Note that Theorem 3 for $r=r_h(R,E_{\mathbf{x}})$ case is equivalent to (77). Identity given in (76) follows from an analysis similar to the one used for establishing (73), in which instead of Lemma 3 we use a simple typicality argument such as [10, Corollary 1.2]. □

We have set $L_1=1$ in the proof. If instead of $L_1=1$ we had chosen L_1 to be a subexponential function of n which grew to infinity with n, the logic and the mechanics of the proof would still work but we would have replaced $\Gamma(T)$ with $\mathcal{E}_{\mathbf{e}}(0,E_{\mathbf{x}})$, while keeping the term including $\tilde{E}_H(R)$ the same. Since the best known upper bound for $\mathcal{E}_{\mathbf{e}}(0,E_{\mathbf{x}})$ is $\Gamma(E_{\mathbf{x}})$ for $E_{\mathbf{x}}\leq T^*$, the final result is the same for the case with feedback.[21] On the other hand, for the case without feedback, which is not the main focus of this paper, this does make a difference. By choosing L_1 to be a function of block length that goes to infinity subexponentially with block length, one can use Telatar's converse result [31, Th. 4.4], on the error exponent at zero rate and zero erasure exponent without feedback.

In Fig. 1, the upper and lower bounds we have derived for error exponent are plotted as a function of the erasure exponent for a binary symmetric channel with crossover probability $\epsilon=0.25$ at rate $R=8.62\times10^{-2}$ nats per channel use. Solid lines are the lower bounds to the error exponent for block codes with feedback, which have been established in Section III, and without feedback, which was established previously [10], [14],

[21]In binary symmetric channels, these results can be strengthened using the value of $\mathcal{E}(0)$, [35]. However, those changes will improve the upper bound on error exponent only at low rates and high erasure exponents.

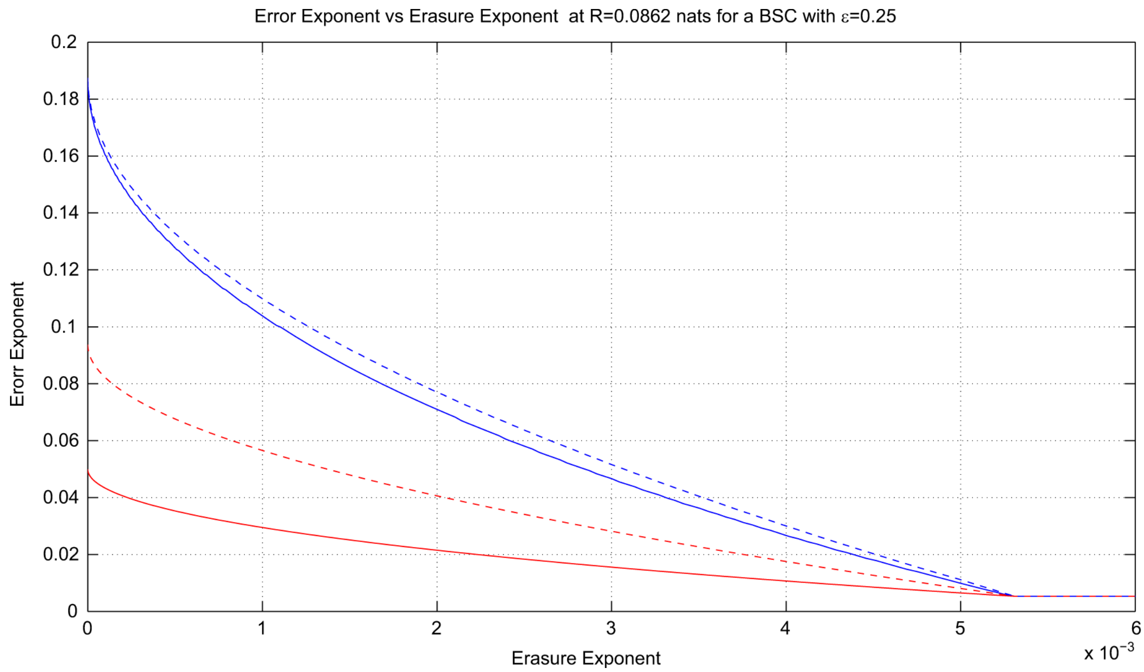

Fig. 1. Error exponent versus erasure exponent.

[31]. Dashed lines are the upper bounds obtained using Theorem 3.

Note that all four curves meet at a point on the bottom right; this is the point that corresponds to the error exponent of block codes at the rate $R = 8.62 \times 10^{-2}$ nats per channel use and its values are the same with and without feedback since we are on a symmetric channel and our rate is over the critical rate. Any point to the lower right of this point is achievable both with and without feedback.

The proximity of the inner and outer bounds demonstrated in Fig. 1 is not particular to the channel we have chosen. A discussion of the closeness of the inner and outer bounds is given in Section VI.

V. Erasure Exponent of Error-Free Codes: $\mathcal{E}_{\mathbf{x}}(R)$

For all DMCs which have one or more zero probability transitions, for all rates below capacity, $R \le \mathcal{C}$ and for small enough $E_{\mathbf{x}}$'s, $E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty$. For such $(R, E_{\mathbf{x}})$ pairs, coding scheme we have described in Section III gives us an error-free code. The connection between the erasure exponent of error-free block codes and the error exponent of block codes with erasures is not confined to this particular encoding scheme. In order to explain those connections in more detail let us first define the error-free codes more formally.

Definition 3: A sequence $\mathcal{Q}_0$ of block codes with feedback is an error-free reliable sequence iff

$$P_{\mathbf{e}}{}^{(\mathsf{n})} = 0 \quad \forall \mathsf{n}, \quad \text{and} \quad \limsup_{\mathsf{n} \longrightarrow \infty} \left(P_{\mathbf{x}}{}^{(\mathsf{n})} + \frac{1}{|\mathcal{M}^{(\mathsf{n})}|} \right) = 0.$$

The highest rate achievable for error-free reliable codes is the zero-error capacity with feedback and erasures $\mathcal{C}_{x,0}$.

If all the transition probabilities are positive, i.e., $\min_{x,y} W(y|x) = \delta > 0$, then $\frac{\mathbf{P}[z^{\mathsf{n}}|m]}{\mathbf{P}[z^{\mathsf{n}}|\tilde{m}]} \ge (\frac{\delta}{1-\delta})^n$ for all $m, \tilde{m} \in \mathcal{M}$ and $z^{\mathsf{n}} \in \mathcal{Z}^{\mathsf{n}}$. Thus, we have

$$\mathbf{P}\left[m|\, z^{\mathsf{n}}\right] \ge \left(\frac{\delta}{1-\delta}\right)^n \mathbf{P}\left[\tilde{m}|\, z^{\mathsf{n}}\right] \quad \forall m, \tilde{m} \in \mathcal{M}, \forall z^{\mathsf{n}} \in \mathcal{Z}^{\mathsf{n}}. \tag{78}$$

Consequently, we have $P_{\mathbf{e}} \ge \frac{(e^{\mathsf{n}R}-1)\delta^{\mathsf{n}}}{(e^{\mathsf{n}R}-1)\delta^{\mathsf{n}}+(1-\delta)^{\mathsf{n}}}$ and $\mathcal{C}_{x,0}$ is zero. On the other hand, as an immediate consequence of the encoding scheme suggested by Yamamoto and Itoh [33], if there is one or more zero probability transitions, $\mathcal{C}_{x,0}$ is equal to channel capacity $\mathcal{C}$.

Definition 4: For all DMCs with at least one (x, y) pair such that $W(y|x) = 0$, $\forall R \le \mathcal{C}$ erasure exponent of error-free block codes with feedback is defined as

$$\mathcal{E}_{\mathbf{x}}(R) \triangleq \sup_{\mathcal{Q}_0: R(\mathcal{Q}_0) \ge R} E_{\mathbf{x}}(\mathcal{Q}_0). \tag{79}$$

For any erasure exponent $E_{\mathbf{x}}$ less than $\mathcal{E}_{\mathbf{x}}(R)$, there is an error-free reliable sequence, i.e., there is a reliable sequence with an infinite error exponent

$$E_{\mathbf{x}} \le \mathcal{E}_{\mathbf{x}}(R) \Rightarrow \mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty. \tag{80}$$

More interestingly, if $E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R)$, then $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) < \infty$. In order to see this, let δ be the minimum nonzero transition probability. Then, for any $m, \tilde{m} \in \mathcal{M}$ and $z \in \mathcal{Z}$ such that $\mathbf{P}[m|\, z^{\mathsf{n}}]\, \mathbf{P}[\tilde{m}|\, z^{\mathsf{n}}] > 0$, we have $\mathbf{P}[z^{\mathsf{n}}|\, m] \ge \delta^n \mathbf{P}[z^{\mathsf{n}}|\, \tilde{m}]$. Thus, if $\mathbf{P}\left[\hat{\mathsf{M}} \notin \{\mathsf{M}, \mathbf{x}\}\middle|\, \mathsf{Z}^{\mathsf{n}}\right] \neq 0$, then $\mathbf{P}\left[\hat{\mathsf{M}} \notin \{\mathsf{M}, \mathbf{x}\}\middle|\, \mathsf{Z}^{\mathsf{n}}\right] \ge \frac{\delta^{\mathsf{n}}}{1+\delta^{\mathsf{n}}}$. Using this, we get

$$\mathbf{E}\left[\mathbb{1}_{\{\mathbf{P}[\hat{\mathsf{M}} \notin \{\mathsf{M},\mathbf{x}\} | \mathsf{Z}^{\mathsf{n}}] \neq 0\}}\right]$$

$$\leq \frac{1+\delta^{\mathsf{n}}}{\delta^{\mathsf{n}}}\mathbf{E}\left[\mathbb{1}_{\{\mathbf{P}[\hat{\mathsf{M}}\notin\{\mathsf{M},\mathbf{x}\}|\mathsf{Z}^{\mathsf{n}}]\neq 0\}}\mathbf{P}\left[\hat{\mathsf{M}}\notin\{\mathsf{M},\mathbf{x}\}\middle|\mathsf{Z}^{\mathsf{n}}\right]\right]$$
$$= (1+\delta^{-\mathsf{n}})\mathbf{P}\left[\hat{\mathsf{M}}\notin\{\mathsf{M},\mathbf{x}\}\right]. \tag{81}$$

Equation (81) reveals that the total probability of z^{n}'s, at which the receiver chooses to decode to a message rather than declaring an erasure despite the fact that it is not certain about the message, is upper bounded by $(1+\delta^{-\mathsf{n}})$ times the undetected error probability. Thus, if we replace the decoder with a new decoder, which declares an erasure unless it is sure about the transmitted message, i.e., unless there is a message with posterior probability one, resulting erasure probability $P_{\mathbf{x}}{}'$ will be bounded in terms of original error-and-erasure probabilities as follows:

$$P_{\mathbf{x}}{}' \leq P_{\mathbf{x}} + (1+\delta^{-\mathsf{n}})P_{\mathbf{e}}. \tag{82}$$

Thus, by changing the decoding rule, any length n code with error probability $P_{\mathbf{e}}$ and erasure probability $P_{\mathbf{x}}$ can be transformed into error-free code with erasure probability $P_{\mathbf{x}}{}'$, where $P_{\mathbf{x}}{}'$ satisfies (82). Using this transformation, we can change any code with errors-and-erasure decoding into an error-free block code with erasures. Evidently, we can use the very same transformation to convert reliable sequences into error-free reliable sequences. Considering error-and-erasure exponents of the original reliable sequences and erasure exponents of the resulting error-free reliable sequences, we get

$$\mathcal{E}_{\mathbf{x}}(R) \geq \min\{E_{\mathbf{x}}, \mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}}) + \ln\delta\} \qquad \forall R, E_{\mathbf{x}}. \tag{83}$$

Consequently

$$E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R) \Rightarrow \mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}}) \leq \mathcal{E}_{\mathbf{x}}(R) - \ln\delta < \infty. \tag{84}$$

As a result of (80) and (84), we can conclude that $\mathcal{E}_{\mathbf{e}}(R,E_{\mathbf{x}}) = \infty$ if and only if $E_{\mathbf{x}} \leq \mathcal{E}_{\mathbf{x}}(R)$. In a sense, similarly to the error exponent of erasure-free block codes $\mathcal{E}(R)$, the erasure exponent of the error-free bock codes $\mathcal{E}_{\mathbf{x}}(R)$ gives a partial description of $\mathcal{E}(R,E_{\mathbf{x}})$. $\mathcal{E}(R)$ gives the value of error exponents below which the erasure exponent can be pushed to infinity and $\mathcal{E}_{\mathbf{x}}(R)$ gives the value of the erasure exponent below which the error exponent can be pushed to infinity.

In the following, the erasure exponent of zero-error codes $\mathcal{E}_{\mathbf{x}}(R)$ is investigated separately for two families of channels: channels which have a positive zero-error capacity, i.e., $\mathcal{C}_0 > 0$ and channels which have zero zero-error capacity, i.e., $\mathcal{C}_0 = 0$.

A. Case 1: $\mathcal{C}_0 > 0$

Theorem 4: For a DMC, if $\mathcal{C}_0 > 0$, then

$$E_H(R) \geq \mathcal{E}_{\mathbf{x}}(R) \geq E_{sp}(R).$$

Proof: If zero-error capacity is strictly greater than zero, i.e., $\mathcal{C}_0 > 0$, then one can achieve the sphere packing exponent, with zero-error probability using a two-phase scheme. In the first phase, the transmitter uses a length $\mathsf{n}_1 = \lceil e^{\mathsf{n}_1 R}\rceil$ block code without feedback with a list decoder of size $L = \lceil \frac{\partial}{\partial R}E_{sp}(R,P_R^*)\rceil$ where P_R^* is the input distribution satisfying $E_{sp}(R) = E_{sp}(R,P_R^*)$. Note that with this list size the sphere packing exponent[22] is achievable at rate R. Thus, the correct message is in the list with at least probability $(1-e^{-\mathsf{n}_1 E_{sp}(R)})$; see [10, p. 196]. In the second phase, the transmitter uses a zero-error code of length[23] $\mathsf{n}_2 = \lceil\frac{\ln(L+1)}{\mathcal{C}_0}\rceil$ with $L+1$ messages, to tell the receiver whether the correct message is in that list, and the correct message itself if it is in the list. Clearly such a feedback code with two phases is error-free, and it has erasures only when there exists an error in the first phase. Thus, the erasure probability of the overall code is upper bounded by $e^{-\mathsf{n}_1 E_{sp}(R)}$. Note that n_2 is fixed for a given R. Consequently, as the length of the first phase n_1 grows to infinity, the rate and erasure exponent of $(\mathsf{n}_1+\mathsf{n}_2)$ long block code converges to the rate and error exponent of n_1 long code of the first phase, i.e., to R and $E_{sp}(R)$. Thus

$$\mathcal{E}_{\mathbf{x}}(R) \geq E_{sp}(R).$$

Any error-free block code with erasures can be forced to decode, at erasures. The resulting fixed-length code has an error probability no larger than the erasure probability of the original code. However, we know that [17] the error probability of the erasure-free block codes with feedback decreases with an exponent no larger than $E_H(R)$. Thus

$$\mathcal{E}_{\mathbf{x}}(R) \leq E_H(R).$$

This upper bound on the erasure exponent also follows from the converse result we present in Theorem 6. □

For symmetric channels, $E_H(R) = E_{sp}(R)$ and Theorem 4 determines the erasure exponent of error-free codes on symmetric channels with nonzero zero-error capacity completely.

B. Case 2: $\mathcal{C}_0 = 0$

This case is more involved than the previous one. First, we establish an upper bound on $\mathcal{E}_{\mathbf{x}}(R)$ in terms of the improved version of Haroutunian's bound, i.e., Lemma 4, and the erasure exponent of error-free codes at zero rate $\mathcal{E}_{\mathbf{x}}(0)$. Then, we show that $\mathcal{E}_{\mathbf{x}}(0)$ is equal to the erasure-exponent error-free block codes with two messages $\mathcal{E}_{\mathbf{x},2}$ and bound $\mathcal{E}_{\mathbf{x},2}$ from below.

For any M, n, and L, $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},L,P_{\mathbf{x}}) = 0$ for large enough $P_{\mathbf{x}}$. We denote the minimum of such $P_{\mathbf{x}}$'s by $\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)$. Thus, we can write $\mathcal{E}_{\mathbf{x},2}$ as

$$\mathcal{E}_{\mathbf{x},2} = \liminf_{\mathsf{n}\longrightarrow\infty}\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1).$$

Theorem 5: For any n, M, L, $\mathsf{n}_1 \leq \mathsf{n}$ and L_1, minimum erasure probability of fixed-length error-free block codes with feedback $\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)$ satisfies

$$\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L) \geq \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)\mathcal{P}_{0,\mathbf{x}}(L_1+1,\mathsf{n}-\mathsf{n}_1,L). \tag{85}$$

[22]Indeed this upper bound on error probability is tight exponentially for block codes without feedback.

[23]For some DMCs with $\mathcal{C}_0 > 0$ and for some L, one may need more than $\lceil\frac{\ln(L+1)}{\mathcal{C}_0}\rceil$ time units to convey one of the $(L+1)$ messages without any errors, because $\mathcal{C}_0$ itself is defined as a limit. But even in those cases we are guaranteed to have a fixed amount of time for these transmissions, which do not change with n_1. Thus, the above argument holds as is, even in those cases.

As Theorem 2, Theorem 5 is correct both with and without feedback. Although $\mathcal{P}_{0,\mathbf{x}}$'s and $\mathcal{P}_{\mathbf{e}}$ will be different in each case, the relationship between them given in (85) holds in both cases.

Proof: If $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)=0$, the theorem holds trivially. Thus, we assume henceforth that $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)>0$. Using Theorem 2 with $P_{\mathbf{x}}=\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)$, we get

$$\begin{aligned}&\mathcal{P}_{\mathbf{e}}\left(M,\mathsf{n},L,\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)\right)\\ &\geq \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)\\ &\quad\times \mathcal{P}_{\mathbf{e}}\left(L_1+1,(\mathsf{n}-\mathsf{n}_1),L,\frac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)}\right).\end{aligned}$$

Since $\mathcal{P}_{\mathbf{e}}\left(M,\mathsf{n},L,\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)\right)=0$ and $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)>0$, we have

$$\mathcal{P}_{\mathbf{e}}\left(L_1+1,(\mathsf{n}-\mathsf{n}_1),L,\frac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)}\right)=0.$$

Thus

$$\frac{\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)}\geq \mathcal{P}_{0,\mathbf{x}}\left(L_1+1,(\mathsf{n}-\mathsf{n}_1),L\right). \qquad \square$$

As we have done in the errors-and-erasures case, we can convert this into a bound on exponents. If we use the improved version of Haroutunian's bound, i.e., Lemma 4, as an upper bound on the error exponent of erasure-free block codes, we get the following.

Theorem 6: For any rate $R\geq 0$ for any $\alpha\in\left[\frac{R}{C},1\right]$

$$\mathcal{E}_{\mathbf{x}}(R)\leq \alpha\tilde{E}_H\left(\frac{R}{\alpha}\right)+(1-\tilde{\alpha})\mathcal{E}_{\mathbf{x}}(0).$$

Now let us focus on the value of the erasure exponent at zero rate.

Lemma 5: For the channels which have zero zero-error capacity, i.e., $\mathcal{C}_0=0$, the erasure exponent of error-free block codes at zero rate $\mathcal{E}_{\mathbf{x}}(0)$ is equal to the erasure exponent of error-free block codes with two messages $\mathcal{E}_{\mathbf{x},2}$.

Note that unlike the two-message case $\mathcal{E}_{\mathbf{x},2}$, in the zero rate case $\mathcal{E}_{\mathbf{x}}(0)$, the number of messages increases with block length to infinity, thus we cannot claim $\mathcal{E}_{\mathbf{x},2}=\mathcal{E}_{\mathbf{x}}(0)$ just as a result of their definitions.

Proof: If we write Theorem 5 for $L=1$, $\mathsf{n}_1=0$ and $L_1=1$

$$\begin{aligned}\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},1)&\geq \mathcal{P}_{\mathbf{e}}(M,0,1)\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\\ &=\frac{M-1}{M}\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\qquad \forall M,\mathsf{n}.\end{aligned}$$

Thus, as an immediate result of the definitions of $\mathcal{E}_{\mathbf{x}}(0)$ and $\mathcal{E}_{\mathbf{x},2}$, we have $\mathcal{E}_{\mathbf{x}}(0)\leq\mathcal{E}_{\mathbf{x},2}$.

In order to prove the equality, one needs to prove $\mathcal{E}_{\mathbf{x}}(0)\geq\mathcal{E}_{\mathbf{x},2}$. For doing that let us assume that it is possible to send one bit with erasure probability ϵ with a block code of length $\ell(\epsilon)$

$$\epsilon\geq\mathcal{P}_{0,\mathbf{x}}(2,\ell(\epsilon),1). \tag{86}$$

One can use this code to send r bits, by repeating each bit whenever there exists an erasure. If the block length is $\mathsf{n}=k\ell(\epsilon)$, then a message erasure occurs only when the number of bit erasures in k trials is more than $k-r$. Let $\#e$ denote the number of erasures out of k trials, then

$$\mathbf{P}[\#e=l]=\frac{k!}{(k-l)!l!}(1-\epsilon)^{k-l}\epsilon^l$$

and

$$P_{\mathbf{x}}=\sum_{l=k-r+1}^{k}\mathbf{P}[\#e=l].$$

Thus

$$\begin{aligned}P_{\mathbf{x}}&=\sum_{l=k-r+1}^{k}\frac{k!}{l!(k-l)!}(1-\epsilon)^{k-l}\epsilon^l\\ &=\sum_{l=k-r+1}^{k}\frac{k!}{l!(k-l)!}\left(\frac{l}{k}\right)^l\left(1-\frac{l}{k}\right)^{k-l}e^{-\left[l\ln\frac{l/k}{\epsilon}+(k-l)\ln\frac{1-l/k}{1-\epsilon}\right]}\\ &=\sum_{l=k-r+1}^{k}\frac{k!}{l!(k-l)!}\left(\frac{l}{k}\right)^l\left(1-\frac{l}{k}\right)^{k-l}e^{-k\mathsf{D}\left(\frac{l}{k}\middle\|\epsilon\right)}.\end{aligned}$$

Then, for any $\epsilon\leq 1-\frac{r}{k}$, we have

$$P_{\mathbf{x}}\leq e^{-k\mathsf{D}\left(1-\frac{r}{k}\middle\|\epsilon\right)}.$$

Evidently $P_{\mathbf{x}}\geq\mathcal{P}_{0,\mathbf{x}}(2^r,\mathsf{n},1)$ for $\mathsf{n}=k\ell(\epsilon)$. Thus

$$\frac{-\ln\mathcal{P}_{0,\mathbf{x}}(2^r,\mathsf{n},1)}{\mathsf{n}}\geq\frac{\mathsf{D}\left(1-\frac{r}{k}\middle\|\epsilon\right)}{\ell(\epsilon)}.$$

Then, $\frac{-\ln\epsilon}{\ell(\epsilon)}$ is an achievable erasure exponent for any sequence of (r,k)'s such that $\lim_{k\to\infty}\frac{r}{k}=0$, i.e., $\mathcal{E}_{\mathbf{x}}(0)\geq\frac{-\ln\epsilon}{\ell(\epsilon)}$. Thus, any exponent achievable for the two-message case is achievable for zero rate case: $\mathcal{E}_{\mathbf{x}}(0)\geq\mathcal{E}_{\mathbf{x},2}$. $\square$

As a result of Lemma 6, which is presented in the next section, we know that

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq\left(\sup_{s\in(0,.5)}\beta(s)\right)^{\mathsf{n}}$$

where

$$\beta(s)=\min_{x,\tilde{x}}\sum_y W(y|x)^{(1-s)}W(y|\tilde{x})^s.$$

Thus, as a result of Lemma 5, we have

$$\mathcal{E}_{\mathbf{x}}(0)=\mathcal{E}_{\mathbf{x},2}\leq-\ln\sup_{s\in(0,0.5)}\beta(s).$$

C. Lower Bounds on $\mathcal{P}_{0,\mathbf{X}}(2,\mathsf{N},1)$

Suppose at time t the correct message M is assigned to the input letter x and the other message is assigned to the input letter $\tilde{x}$, then the receiver cannot rule out the incorrect message at time t with probability $\sum_{y:W(y|\tilde{x})>0}W(y|x)$. Using this fact, one can prove that

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq\left(\min_{x,\tilde{x}}\sum_{y:W(y|\tilde{x})>0}W(y|x)\right)^{\mathsf{n}}. \tag{87}$$

Now let us consider channels whose transition probability matrix W is of the form

$$W=\begin{bmatrix}1-q & q\\ 0 & 1\end{bmatrix}. \tag{88}$$

We denote the output letter that can be reached from both of the input letters by $\tilde{y}$. For the moment, we consider only the deterministic encoding schemes, i.e., $\mathsf{Z}_t=\mathsf{Y}_t$. Note that in the optimal encoding scheme

$$X_t(1,y^{t-1})\neq X_t(2,y^{t-1})\quad \forall t,\ \forall y^{t-1}\in\mathcal{Y}^{t-1}.$$

Then, for all t and $y^{t-1}\in\mathcal{Y}^{t-1}$

$$\mathbf{P}\left[\mathsf{Y}_t=\tilde{y}|\,\mathsf{M}=1,y^{t-1}\right]\mathbf{P}\left[\mathsf{Y}_t=\tilde{y}|\,\mathsf{M}=2,y^{t-1}\right]=q. \tag{89}$$

Furthermore, if $\mathsf{Y}^{\mathsf{n}}=\tilde{y}^{\mathsf{n}}$, then the receiver cannot decode without errors, i.e., it has to declare an erasure. Then

$$\begin{aligned}&\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\\ &\quad\geq\frac{1}{2}(\mathbf{P}[\mathsf{Y}^{\mathsf{n}}=\tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=1]+\mathbf{P}[\mathsf{Y}^{\mathsf{n}}=\tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=2])\\ &\quad\overset{(a)}{\geq}\sqrt{\mathbf{P}[\mathsf{Y}^{\mathsf{n}}=\tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=1]\,\mathbf{P}[\mathsf{Y}^{\mathsf{n}}=\tilde{y}\tilde{y}\ldots\tilde{y}|\,\mathsf{M}=2]}\\ &\quad\overset{(b)}{=}q^{\frac{\mathsf{n}}{2}}\end{aligned} \tag{90}$$

where (a) hods because the arithmetic mean is larger than the geometric mean and (b) follows from (89).

For W given in (88) the bound given in (90) is very tight. If the encoder assigns the first message to the input letter that always leads to $\tilde{y}$ and the second message to the other input letter in first $\lfloor\frac{\mathsf{n}}{2}\rfloor$ instances, and does the flipped assignment in the last $\lceil\frac{\mathsf{n}}{2}\rceil$ instances, then an erasure happens with a probability less than $q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$, i.e., $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\leq q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$.

On the other hand, for W given in (88), the bound given in (87) ensures only $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq q^{\mathsf{n}}$, rather than $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$. Thus, for the channel given in (88), the bound given in (90) is tighter than the one in (87).

The idea used in deriving the bound given in (90) for this particular W can be applied to a general DMC to prove the following lower bound:

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq\left(\min_{x,\tilde{x}}\sum\nolimits_y\sqrt{W(y|x)W(y|\tilde{x})}\right)^{\mathsf{n}}. \tag{91}$$

The bound given in (91) decays exponentially in n, even when all entries of W are positive, however for those channels the bound given in (87) implies $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq 1$. Thus, the bound given in (91) cannot be superior to the bound given in (87) in general. The following bound implies bounds given in both (87) and (91). Furthermore, for certain channels, it is strictly better than both.

Lemma 6: Erasure probability of all error-free block codes with two messages is lower bounded as

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)\geq(\sup_{s\in(0,.5)}\beta(s))^{\mathsf{n}} \tag{92}$$

where

$$\beta(s)=\min_{x,\tilde{x}}\sum_y W(y|x)^{(1-s)}W(y|\tilde{x})^s.$$

Note that bounds given in (87) and (91) are implied by $\lim_{s\to 0^+}\beta(s)$ and $\lim_{s\to 0.5^-}\beta(s)$, respectively.

Although $\sum_y W(y|x)^sW(y|\tilde{x})^{1-s}$ is convex in s on $(0,0.5)$ for all $(x,\tilde{x})$ pairs, $\beta(s)$ is not convex in s because of the minimization in its definition. Thus, the supremum over s does not necessarily occur on the boundaries. Indeed there are channels for which the bound given in Lemma 6 is strictly better than the bounds given in (87) and (91). Following is the transition probability matrix of one such channel

$$W=\begin{bmatrix}0.1600 & 0.0200 & 0.2200 & 0.3000 & 0.3000\\ 0.0900 & 0.4000 & 0.2700 & 0.0002 & 0.2398\\ 0.1800 & 0.2000 & 0.3000 & 0.3200 & 0\end{bmatrix}$$

$\lim_{s\to 0}\beta(s)=0.7000$, $\lim_{s\to 0.5}\beta(s)=0.7027$, and $\beta(0.18)=0.7299$.

Proof: Let μ_t and $\tilde{\mu}_t(z^{t-1})$ be

$$\begin{aligned}\mu_t&=\left\{z^t:\mathbf{P}\left[\mathsf{M}=1|\,z^t\right]\mathbf{P}\left[\mathsf{M}=2|\,z^t\right]>0\right\}\\ \tilde{\mu}_t(z^{t-1})&=\left\{y_t:\mathbf{P}\left[y_t|\,\mathsf{M}=1,z^{t-1}\right]\mathbf{P}\left[y_t|\,\mathsf{M}=2,z^{t-1}\right]>0\right\}.\end{aligned}$$

Then, for any error-free code and for any $s\in(0,0.5)$, we have

$$\begin{aligned}P_{\mathbf{x}}&=\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}\right]\\ &=\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}(\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]+\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}])\right]\\ &=\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}((1-s)\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]+s\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}])\right]\\ &\quad+\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}(s\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]+(1-s)\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}])\right]\\ &\geq\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]^{1-s}\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}]^s\right]\\ &\quad+\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]^s\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}]^{1-s}\right]\end{aligned} \tag{93}$$

where the last inequality follows from the fact that the arithmetic mean is lower bounded by the geometric mean.

Furthermore, using the law of total expectation, we can rewrite the first terms in (93) as described in (94), shown at the bottom of the page. Note that

$$\begin{aligned}\frac{\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]}{\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}-1}]}&=\frac{\mathbf{P}\left[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}-1},\mathsf{Y}_{\mathsf{n}}\right]}{\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}-1}]}\\ &=\frac{\mathbf{P}\left[\mathsf{Y}_{\mathsf{n}}|\,\mathsf{M}=1,\mathsf{Z}^{\mathsf{n}-1}\right]}{\mathbf{P}[\mathsf{Y}_{\mathsf{n}}|\,\mathsf{Z}^{\mathsf{n}-1}]}.\end{aligned} \tag{95}$$

$$\begin{aligned}&\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}}\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]^{1-s}\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}]^s\right]\\ &\quad=\mathbf{E}\left[\mathbf{E}\left[\mathbb{1}_{\{\tilde{\mu}_{\mathsf{n}}(\mathsf{Z}^{\mathsf{n}-1})\}}\left(\frac{\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}}]}{\mathbf{P}[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}-1}]}\right)^{1-s}\left(\frac{\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}}]}{\mathbf{P}[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}-1}]}\right)^s\middle|\,\mathsf{Z}^{\mathsf{n}-1}\right]\mathbb{1}_{\{\mu_{(\mathsf{n}-1)}\}}\mathbf{P}\left[\mathsf{M}=1|\,\mathsf{Z}^{\mathsf{n}-1}\right]^{1-s}\mathbf{P}\left[\mathsf{M}=2|\,\mathsf{Z}^{\mathsf{n}-1}\right]^s\right].\end{aligned} \tag{94}$$

Similarly

$$\frac{\mathbf{P}\left[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}}\right]}{\mathbf{P}\left[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}-1}\right]}=\frac{\mathbf{P}\left[\mathsf{Y}_{\mathsf{n}} \mid \mathsf{M}=2, \mathsf{Z}^{\mathsf{n}-1}\right]}{\mathbf{P}\left[\mathsf{Y}_{\mathsf{n}} \mid \mathsf{Z}^{\mathsf{n}-1}\right]}. \tag{96}$$

Thus, using (95) and (96), we have

$$\begin{aligned}
&\mathbf{E}\left[\mathbb{1}_{\{\tilde{\mu}_{\mathsf{n}}(\mathsf{Z}^{\mathsf{n}-1})\}}\left(\frac{\mathbf{P}[\mathsf{M}=1 \mid \mathsf{Z}^{\mathsf{n}}]}{\mathbf{P}[\mathsf{M}=1 \mid \mathsf{Z}^{\mathsf{n}-1}]}\right)^{1-s}\left(\frac{\mathbf{P}[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}}]}{\mathbf{P}[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}-1}]}\right)^{s} \middle| \mathsf{Z}^{\mathsf{n}-1}\right] \\
&=\mathbf{E}\left[\frac{\mathbf{P}\left[\mathsf{Y}_{\mathsf{n}} \mid \mathsf{M}=1, \mathsf{Z}^{\mathsf{n}-1}\right]^{1-s} \mathbf{P}\left[\mathsf{Y}_{\mathsf{n}} \mid \mathsf{M}=2, \mathsf{Z}^{\mathsf{n}-1}\right]^{s}}{\mathbf{P}\left[\mathsf{Y}_{\mathsf{n}} \mid \mathsf{Z}^{\mathsf{n}-1}\right]} \middle| \mathsf{Z}^{\mathsf{n}-1}\right] \\
&=\sum_{y_{\mathsf{n}}} \mathbf{P}\left[y_{\mathsf{n}} \mid \mathsf{M}=1, \mathsf{Z}^{\mathsf{n}-1}\right]^{1-s} \mathbf{P}\left[y_{\mathsf{n}} \mid \mathsf{M}=2, \mathsf{Z}^{\mathsf{n}-1}\right]^{s} \\
&\geq \beta(s)
\end{aligned} \tag{97}$$

where the last inequality follows from the definition of $\beta(s)$ given in (92).

Using (94) and (97), we get

$$\begin{aligned}
&\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}} \mathbf{P}\left[\mathsf{M}=1 \mid \mathsf{Z}^{\mathsf{n}}\right]^{1-s} \mathbf{P}\left[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}}\right]^{s}\right] \\
&\quad \geq \mathbf{E}\left[\mathbf{P}\left[\mathsf{M}=1 \mid \mathsf{Z}^{0}\right]^{1-s} \mathbf{P}\left[\mathsf{M}=2 \mid \mathsf{Z}^{0}\right]^{s}\right] \beta(s)^{\mathsf{n}} \\
&\quad \geq \frac{1}{2} \beta(s)^{\mathsf{n}}.
\end{aligned} \tag{98}$$

If we follow a similar line of reasoning for the second term in (93), we get

$$\begin{aligned}
&\mathbf{E}\left[\mathbb{1}_{\{\mu_{\mathsf{n}}\}} \mathbf{P}\left[\mathsf{M}=1 \mid \mathsf{Z}^{\mathsf{n}}\right]^{s} \mathbf{P}\left[\mathsf{M}=2 \mid \mathsf{Z}^{\mathsf{n}}\right]^{1-s}\right] \\
&\quad \geq \frac{1}{2} \beta(1-s)^{\mathsf{n}} \\
&\quad =\frac{1}{2} \beta(s)^{\mathsf{n}}.
\end{aligned} \tag{99}$$

The lemma follows from (93), (98), and (99) by taking the supremum over $s \in (0, 0.5)$. □

VI. Discussion

The value of the error exponent is not known for erasure-free fixed-length block codes with feedback on a general DMC. We do not even know if it is still upper bounded by the sphere packing exponent for nonsymmetric DMCs. Yet the value of the error exponent for fixed-length block codes with feedback and errors-and-erasures decoding can be deduced, for the zero-erasure-exponent case, from the results on the variable-length block codes [3], [33]. Our main aim in this paper was establishing upper and lower bounds that extend the bounds at the zero erasure-exponent case gracefully and nontrivially to the positive erasure-exponents values. Our results are best understood in this framework and should be interpreted accordingly.

By finding the optimal error-exponent/erasure-exponent tradeoff, one solves the open problem of finding the optimal error exponent of erasure-free fixed-length block codes with feedback. This is an important and difficult problem on its own. We did not attempt to solve that problem, yet the inner and outer bounds we have derived for the case with erasure quantify how much we loose from the optimal performance by using the encoding schemes inspired by the optimal encoding schemes for variable-length block codes.

We derived inner bounds using two-phase encoding schemes, which are known to be optimal at zero-erasure-exponent case. We have improved the performance of these two-phase schemes at positive erasure-exponent values by choosing relative durations of the phases considering the desired values of the rate and erasure exponent, and by using a decoder that takes into account the outputs of both phases while deciding between decoding to a message and declaring an erasure. However within each phase the assignment of messages to input letters is fixed. In a general feedback encoder, on the other hand, the assignment of the messages to input symbols at each time can depend on the previous channel outputs, and such encoding schemes have proven to improve the error exponent at low rates [6], [13], [23], [34] for some DMCs. Using such an encoding in the communication phase will improve the performance at low rates. In addition, instead of committing to a fixed duration for the communication phase, one might consider using a stopping time to switch from the communication phase to the control phase. However, in order to apply these ideas effectively for a general DMC, it seems that one first needs to solve the problem for the erasure-free block codes for a general DMC.

We derived the outer bounds without making any assumption about the feedback encoding scheme. Thus, they are valid for any fixed-length block code with feedback and erasures. The principal idea of the straight-line bound is making use of the bounds derived for different rate, erasure-exponent pairs by taking their convex combinations. This approach can be interpreted as a generalization of the outer bounds used for variable-length block codes [2], [3]. As was the case for the inner bounds, it seems that in order to improve the outer bounds one needs to establish the outer bounds on two related problems, i.e., on the error exponents of erasure-free block codes with feedback and on the error-exponent/erasure-exponent tradeoff at zero rate.

The inner and outer bounds we have derived do not coincide for arbitrary values of the erasure exponent. But they do coincide for all channels at all rates at zero erasure exponent.

- If the channel does not have a zero probability transition, both the inner bound and the outer bound are equal to $(1-\frac{R}{C})\mathcal{D}$.
- If the channel does have a zero probability transition, the inner bound is equal to infinity and there are fixed-length block codes with zero-error probability for all large enough block lengths.

Furthermore, on the plane where the erasure exponent is equal to the error exponent, the outer bound we have derived is loose only as much as the best outer bound we know for the error exponent of the erasure-free block codes with feedback is loose. Thus, the proximity we have observed between the inner and outer bounds in Fig. 1 is not peculiar to the particular channel we have chosen for Fig. 1. For all channels, the inner and outer bounds we have derived coincide in the upper left corner as they do in Fig. 1. If the channel is symmetric and if we are considering a rate over critical rate they will also coincide in the lower right corner. Furthermore, if the sphere packing exponent is shown to be an upper bound for the error exponent of erasure-free fixed-length block codes, this behavior will extend to nonsymmetric channels.

APPENDIX A
THE ERROR-EXPONENT TRADEOFF FOR FEEDBACK ENCODING SCHEMES WITH TWO-MESSAGE AND ERASURE-FREE DECODERS

In this appendix, we will first establish an alternative expression for the $\Gamma(T,\Pi)$ function defined in (39) in Lemma 7. After that, we will prove that in a two-message code with feedback on a DMC, if the error exponent of one of the messages is greater than some $T \geq T_0$, then the error exponent of the other message cannot be greater than $\Gamma(T)$, where T_0 and $\Gamma(T)$ are defined in (65) and (66), respectively. Furthermore, we will prove that if the error probability of the one of the messages is zero, then the error probability of the other message cannot be lower than $e^{-\mathsf{n}T_0}$; we will also prove that it can be as low as $e^{-\mathsf{n}T_0}$; see Lemma 8. These results will imply that the error performance of a two-message code does not improve with feedback. Berlekamp attributes this result to Shannon and Gallager in [1].

Lemma 7: $\Gamma(T,\Pi)$ defined in (39) is equal to the expression given in (40)

Proof: $\Gamma(T,\Pi)$ satisfies

$$\begin{aligned}\Gamma(T,\Pi) &= \min_{U:\mathsf{D}(U\|W_a|\Pi)\leq T} \mathsf{D}(U\|W_r|\Pi)\\ &= \min_U \sup_{\lambda>0} \mathsf{D}(U\|W_r|\Pi) + \lambda(\mathsf{D}(U\|W_a|\Pi) - T)\\ &\overset{(a)}{=} \sup_{\lambda>0}\min_U \mathsf{D}(U\|W_r|\Pi) + \lambda(\mathsf{D}(U\|W_a|\Pi) - T)\\ &= \sup_{\lambda>0}\min_U -\lambda T + (1+\lambda)\sum_{x,\tilde{x},y}\Pi(x,\tilde{x})U(y|x,\tilde{x})\\ &\quad\times \ln\frac{U(y|x,\tilde{x})}{W(y|x)^{\frac{\lambda}{1+\lambda}}W(y|\tilde{x})^{\frac{1}{1+\lambda}}}\\ &\overset{(b)}{=} \sup_{\lambda>0} -\lambda T - (1+\lambda)\sum_{x,\tilde{x}}\Pi(x,\tilde{x})\\ &\quad\times \ln\sum_y W(y|x)^{\frac{\lambda}{1+\lambda}}W(y|\tilde{x})^{\frac{1}{1+\lambda}}.\end{aligned} \tag{100}$$

Note that (a) follows from convexity of $\mathsf{D}(U\|W_r|\Pi) + \lambda(\mathsf{D}(U\|W_a|\Pi) - T)$ in U and linearity (concavity) of it in λ; (b) holds because minimizing U is U_s for $s = \frac{1}{1+\lambda}$.

The function on the right-hand side of (100) is maximized at a positive and finite λ iff there is a λ such that $\mathsf{D}(U_{\frac{1}{1+\lambda}}\|W_a|\Pi) = T$. Thus, by substituting $\lambda = \frac{1-s}{s}$, we receive (101), shown at the bottom of the page. Lemma follows from the definition U_s at $s = 0, 1$ and (101). ☐

Now we are ready to present the proof of Lemma 3

Proof of Lemma 3: Our proof is very much like the one for the converse part of [30, Th. 5], except few modifications that allow us to handle the fact that encoding schemes we are considering are feedback encoding schemes. As [30, Th. 5], we construct a probability measure $P_T[\cdot]$ on $\mathcal{Z}^{\mathsf{n}}$ as a function of T and the encoding scheme. Then, we bound the error probability of each message from below using the probability of the decoding region of the other message under $P_T[\cdot]$. We consider probability measures on $\mathcal{Z}^{\mathsf{n}}$ rather than $\mathcal{Y}^{\mathsf{n}}$ to include the possible randomization in the encoding and decoding schemes.

For any $T \geq T_0$ and Π, let $\mathbf{S}_{T,\Pi}$ be

$$\mathbf{S}_{T,\Pi} = \begin{cases} 0, & \text{if } T < \mathsf{D}(U_0\|W_a|\Pi),\\ s, & \text{if } \exists s\in[0,1] \text{ s.t. } \mathsf{D}(U_s\|W_a|\Pi) = T\\ 1, & \text{if } T > \mathsf{D}(U_1\|W_a|\Pi).\end{cases} \tag{102}$$

Recall that

$$T_0 = \max_{x,\tilde{x}} -\ln\sum_{y:W(y|\tilde{x})>0} W(y|x)$$

and

$$\mathsf{D}(U_0\|W_a|\Pi) = -\sum_{x,\tilde{x}}\Pi(x,\tilde{x})\ln\sum_{y:W(y|\tilde{x})>0}W(y|x).$$

Then, for all Π, we have

$$T_0 \geq \mathsf{D}(U_0\|W_a|\Pi). \tag{103}$$

Thus, as a result of the definition of $\mathbf{S}_{T,\Pi}$ and (103), we have

$$\mathsf{D}\left(U_{\mathbf{S}_{T,\Pi}}\middle\|W_a|\Pi\right) \leq T \qquad \forall T\geq T_0. \tag{104}$$

Using Lemma 7, the definition of $\mathbf{S}_{T,\Pi}$, and (103), we can also conclude that

$$\mathsf{D}\left(U_{\mathbf{S}_{T,\Pi}}\middle\|W_r|\Pi\right) = \Gamma(T,\Pi) \leq \Gamma(T) \qquad \forall T\geq T_0. \tag{105}$$

Note that given $\mathsf{Z}^{t-1} = z^{t-1}$ channel input letters assigned to each message at time t, $X_t(m_1, z^{t-1})$ and $X_t(m_2, z^{t-1})$ are fixed for any feedback encoding scheme $X_t(\cdot) : \{m_1, m_2\}\times\mathcal{Z}^{t-1}$. Thus, the corresponding Π is given by

$$\Pi(x,\tilde{x}) = \begin{cases} 0, & \text{if } (x,\tilde{x}) \neq (X_t(m_1,z^{t-1}), X_t(m_2,z^{t-1}))\\ 1, & \text{if } (x,\tilde{x}) = (X_t(m_1,z^{t-1}), X_t(m_2,z^{t-1})).\end{cases} \tag{106}$$

Then, for any $T\geq T_0$, let $P_T\left[y_t\middle| z^{t-1}\right]$ be

$$P_T\left[y_t\middle| z^{t-1}\right] = U_{\mathbf{S}_{T,\Pi}}(y_t|X_t(m_1,z^{t-1}), X_t(m_2,z^{t-1})). \tag{107}$$

$$\Gamma(T,\Pi) = \left\{\begin{array}{lll} \infty, & \text{if} & T < \lim_{s\to 0^+}\mathsf{D}(U_s\|W_a|\Pi)\\ \lim_{s\to 0^+}\mathsf{D}(U_s\|W_r|\Pi), & \text{if} & T = \lim_{s\to 0^+}\mathsf{D}(U_s\|W_a|\Pi)\\ \mathsf{D}(U_s\|W_r|\Pi), & \text{if} & T = \mathsf{D}(U_s\|W_a|\Pi) \quad \text{for some } s\in(0,1)\\ \lim_{s\to 1^-}\mathsf{D}(U_s\|W_r|\Pi), & \text{if} & T = \lim_{s\to 1^-}\mathsf{D}(U_s\|W_a|\Pi)\\ \lim_{s\to 1^-}\mathsf{D}(U_s\|W_r|\Pi), & \text{if} & T > \lim_{s\to 1^-}\mathsf{D}(U_s\|W_a|\Pi)\end{array}\right\} \tag{101}$$

Furthermore, let us assume that the conditional distribution of A_t given $(\mathsf{M},\mathsf{Z}^{t-1},\mathsf{Y}_t)$ under $P_T[\cdot]$ is identical to the conditional distribution of A_t given $(\mathsf{M},\mathsf{Z}^{t-1},\mathsf{Y}_t)$ under $\mathbf{P}[\cdot]$, i.e., the original conditional distribution.

Note that as a result of (104) and (105), we have

$$E_T\left[\ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]}\middle|\,\mathsf{Z}^{t-1}\right] \le T \quad \text{w.p.1}$$

and

$$E_T\left[\ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_2,\mathsf{Z}^{t-1}\right]}\middle|\,\mathsf{Z}^{t-1}\right] \le \Gamma(T) \quad \text{w.p.1.}$$

Now we make a standard measure change argument

$$\begin{aligned}\mathbf{P}&\left[\mathsf{Y}_t|\,\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]\\ &=e^{-\ln\frac{P_T\left[\mathsf{Y}_t|\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]}}P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]\\ &=e^{-E_T\left[\ln\frac{P_T\left[\mathsf{Y}_t|\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]}\middle|\mathsf{Z}^{t-1}\right]}e^{\chi_{t,m_1}(\mathsf{Y}_t|\mathsf{Z}^{t-1})}P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]\\ &\ge e^{-T}e^{\chi_{t,m_1}(\mathsf{Y}_t|\mathsf{Z}^{t-1})}P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right] \qquad (108)\end{aligned}$$

where

$$\chi_{t,m_1}(\mathsf{Y}_t|\mathsf{Z}^{t-1})=E_T\left[\ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]}\middle|\,\mathsf{Z}^{t-1}\right] - \ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]}. \qquad (109)$$

For $m=m_1,m_2$, let $\chi(m)$ be

$$\chi(m)=\left\{z^{\mathsf{n}}:\left|\sum_{t=1}^{\mathsf{n}}\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1})\right|\le 4\sqrt{\mathsf{n}}\ln\frac{1}{P_{\min}}\right\}. \qquad (110)$$

For any event B measurable in the sigma field generated by Z^{n} as a result of equation (108), we have

$$\begin{aligned}\mathbf{P}[\mathsf{B}]&\ge \mathbf{E}\left[\mathbb{1}_{\{\mathsf{B}\}}\mathbb{1}_{\{\chi(m_1)\}}\right]\\ &\ge e^{-\mathsf{n}T}e^{-4\sqrt{\mathsf{n}}\ln\frac{1}{P_{\min}}}E_T\left[\mathbb{1}_{\{\mathsf{B}\}}\mathbb{1}_{\{\chi(m_1)\}}\right]\\ &\ge e^{-\mathsf{n}T}e^{-4\sqrt{\mathsf{n}}\ln\frac{1}{P_{\min}}}P_T\left[\{\mathsf{B}\text{ and }\chi(m_1)\}\right]. \qquad (111)\end{aligned}$$

Following a similar line of reasoning, we get

$$\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_2,\mathsf{Z}^{t-1}\right]\ge e^{-\mathsf{n}\Gamma(T)}e^{\chi_{t,m_2}(\mathsf{Y}_t|\mathsf{Z}^{t-1})}P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right] \qquad (112)$$

where

$$\chi_{t,m_2}(\mathsf{Y}_t|\mathsf{Z}^{t-1})=E_T\left[\ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}m_2,\mathsf{Z}^{t-1}\right]}\middle|\,\mathsf{Z}^{t-1}\right] - \ln \frac{P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]}{\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_2,\mathsf{Z}^{t-1}\right]} \qquad (113)$$

and for any event B measurable in the sigma field generated by Z^{n}, we have

$$\mathbf{P}[\mathsf{B}]\ge e^{-\mathsf{n}\Gamma(T)}e^{-4\sqrt{\mathsf{n}}\ln\frac{1}{P_{\min}}}P_T\left[\{\mathsf{B}\text{ and }\chi(m_2)\}\right]. \qquad (114)$$

Note that for all $m=\{m_1,m_2\}$, $t\in\{1,2,\ldots,\mathsf{n}\}$, and $k\in\{1,2,\ldots,t-1\}$

$$E_T\left[\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1})\middle|\,\mathsf{Z}^{t-1}\right]=0 \qquad (115a)$$

$$E_T\left[(\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1}))^2\middle|\,\mathsf{Z}^{t-1}\right]\le 4(\ln P_{\min})^2 \qquad (115b)$$

$$E_T\left[\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1})\chi_{t-k,m}(\mathsf{Y}_{t-k}|\mathsf{Z}^{t-k-1})\middle|\mathsf{Z}^{t-1}\right]=0. \qquad (115c)$$

Thus, as a result of (115), for $m\in\{m_1,m_2\}$

$$E_T\left[\sum_{t=1}^{\mathsf{n}}\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1})\right]=0 \qquad (116a)$$

$$E_T\left[\left(\sum_{t=1}^{\mathsf{n}}\chi_{t,m}(\mathsf{Y}_t|\mathsf{Z}^{t-1})\right)^2\right]\le 4\mathsf{n}(\ln P_{\min})^2. \qquad (116b)$$

Using (116) and Chebychev's inequality, we conclude that

$$P_T[\chi(m)]\ge 3/4 \qquad m=m_1,m_2.$$

Hence

$$P_T[\chi(m_1)\cap\chi(m_2)]\ge 1/2.$$

Thus, either the total probability of intersection of $\chi(m_1)\cap\chi(m_2)$ with the decoding region of the second message is equal to or larger than $1/4$ or the total probability of intersection of $\chi(m_1)\cap\chi(m_2)$ with the decoding region of the first message is strictly larger than $1/4$. Then, the lemma follows from (111) and (114). □

As we have noted previously T_0 does have an operational meaning: it is the maximum error exponent first message can have, when the error probability of the second message is zero.

Lemma 8: For any feedback encoding scheme with two messages, if $P_{\mathbf{e}|m_2}=0$, then $P_{\mathbf{e}|m_1}\ge e^{-\mathsf{n}T_0}$. Furthermore, there does exist an encoding scheme such that $P_{\mathbf{e}|m_2}=0$, then $P_{\mathbf{e}|m_1}=e^{-\mathsf{n}T_0}$.

Proof: Let us use a construction similar to the one used in the proof of Lemma 3

$$P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]=U_0(\mathsf{Y}_t|X_t(m_1,\mathsf{Z}^{t-1}),X_t(m_2,\mathsf{Z}^{t-1})).$$

Recall that

$$U_0(y_t|x,\tilde{x})=\frac{\mathbb{1}_{\{W(y|\tilde{x})>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|\tilde{x})>0}W(\tilde{y}|x)}W(y|x).$$

Thus

$$P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]\le e^{T_0}\mathbf{P}\left[\mathsf{Y}_t|\,\mathsf{M}=m_1,\mathsf{Z}^{t-1}\right]$$
$$P_T\left[\mathsf{Y}_t|\,\mathsf{Z}^{t-1}\right]\le \mathbb{1}_{\{\mathbf{P}[\mathsf{Y}_t|\mathsf{M}=m_2,\mathsf{Z}^{t-1}]>0\}}.$$

As we have done in the proof of Lemma 3, we will assume that the conditional distribution of A_t given $(\mathsf{M},\mathsf{Z}^{t-1},\mathsf{Y}_t)$ under $P_T[\cdot]$ is identical to the conditional distribution of A_t given $(\mathsf{M},\mathsf{Z}^{t-1},\mathsf{Y}_t)$ under $\mathbf{P}[\cdot]$, i.e., the original conditional distribution.

Then, for any event B measurable in the sigma field generated by Z^{n}, we have

$$\mathbf{P}[\mathsf{B}|\,\mathsf{M}=m_1]\ge e^{-\mathsf{n}T_0}P_T[\mathsf{B}] \qquad (117)$$

$$\mathbf{P}[\mathsf{B}|\,\mathsf{M}=m_2]\ge e^{\mathsf{n}\ln P_{\min}}P_T[\mathsf{B}] \qquad (118)$$

where $P_{\min}$ is the minimum nonzero element of W.

Since $P_{\mathbf{e}|m_2} = 0$, (118) implies that $P_T\left[\hat{\mathsf{M}} \neq m_2\right] = 0$ and $P_T\left[\hat{\mathsf{M}} \neq m_1\right] = 1$. Using this fact together with (117), we conclude that

$$P_{\mathbf{e}|m_1} \geq e^{-\mathsf{n}T_0}. \tag{119}$$

Let us assume that maximizing x-pair in (65) is (x_1^*, x_2^*), i.e., $T_0 = -\ln \sum_{y:W(y|x_2^*)>0} W(y|x_1^*)$. If the encoding scheme sends x_1^* for the first message and x_2^* for the second message, the decoder decodes to the second message unless $\mathsf{Y}_t = y^*$ for some $t \in \{1, 2, \ldots, \mathsf{n}\}$ and for some y^* such that $W(y^*|x_2^*) = 0$. Then, $P_{\mathbf{e}m_2} = 0$ and $P_{\mathbf{e}m_1} = e^{-\mathsf{n}T_0}$. □

Appendix B
Convexity of $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in α

Lemma 9: For any probability distribution P on input alphabet $\mathcal{X}$, $\boldsymbol{\zeta}(P, Q, R)$ is convex in (Q, R) pair.

Proof: Note that

$$\gamma\boldsymbol{\zeta}(R_a, P, Q_a) + (1-\gamma)\boldsymbol{\zeta}(R_b, P, Q_b) = \min_{\substack{V_a, V_b:\, \mathsf{I}(P,V_a)\leq R_a \\ \mathsf{I}(P,V_b)\leq R_b \\ (PV_a)_Y = Q_a \\ (PV_b)_Y = Q_b}} \gamma \mathsf{D}\left(V_a \| W|P\right) + (1-\gamma)\mathsf{D}\left(V_b \| W|P\right).$$

Using the convexity of $\mathsf{D}\left(V \| W|P\right)$ in V and Jensen's inequality, we get

$$\gamma\boldsymbol{\zeta}(R_a, P, Q_a) + (1-\gamma)\boldsymbol{\zeta}(R_b, P, Q_b) \geq \min_{\substack{V_a, V_b:\, \mathsf{I}(P,V_a)\leq R_a \\ \mathsf{I}(P,V_b)\leq R_b \\ (PV_a)_Y = Q_a \\ (PV_b)_Y = Q_b}} \mathsf{D}\left(V_\gamma \| W|P\right)$$

where $V_\gamma = \gamma V_a + (1-\gamma)V_b$.

If the constraint set of a minimization is enlarged, then the resulting minimum does not increase. Using this fact together with the convexity of $\mathsf{I}(P, V)$ in V and Jensen's inequality, we get

$$\begin{aligned}\gamma\boldsymbol{\zeta}(R_a, P, Q_a) + (1-\gamma)\boldsymbol{\zeta}(R_b, P, Q_b) &\geq \min_{\substack{V_\gamma:\, \mathsf{I}(P,V_\gamma)\leq R_\gamma \\ (PV_\gamma)_Y = Q_\gamma}} \mathsf{D}\left(V_\gamma \| W|P\right) \\ &= \boldsymbol{\zeta}(R_\gamma, P, Q_\gamma)\end{aligned}$$

where $R_\gamma = \gamma R_a + (1-\gamma)R_b$, $Q_\gamma = \gamma Q_a + (1-\gamma)Q_b$. □

Lemma 10: For all $(R, E_{\mathbf{x}}, P, \Pi)$ quadruples such that $E_r(R, P) \geq E_{\mathbf{x}}$, $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is a convex function of α on the interval $[\alpha^*(R, E_{\mathbf{x}}, P), 1]$ where $\alpha^*(R, E_{\mathbf{x}}, P)$ is the unique solution[24] of $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$.

Proof: For any P such that $E_r, (R, P)$ is a nonnegative, convex, and decreasing function of R in the interval $[0, \mathsf{I}(P, W)]$. Thus, $\alpha E_r(\frac{R}{\alpha}, P)$ is strictly an increasing continuous function of $\alpha \in [\frac{R}{\mathsf{I}(P,W)}, 1]$. Furthermore, for $\alpha = \frac{R}{\mathsf{I}(P,W)}$, $\alpha E_r(\frac{R}{\alpha}, P) = 0$, and for $\alpha = 1$, $\alpha E_r(\frac{R}{\alpha}, P) \geq E_{\mathbf{x}}$. Thus, $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$ has a unique solution.

Using the convexity arguments analogous to the ones used in the proof of Lemma 9 one can prove that the inequality leading to (120) holds for any $\gamma \in [0, 1]$. Then, the convexity of $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in α follows from (120), shown at the bottom of the page, where α_γ, T_γ, Q_γ, $R_{1\gamma}$, and $R_{2\gamma}$ are given by

$$\begin{aligned}\alpha_\gamma &= \gamma\alpha_a + (1-\gamma)\alpha_b \\ T_\gamma &= \gamma T_a + (1-\gamma)T_b \\ Q_\gamma &= \frac{\gamma\alpha_a}{\alpha_\gamma}Q_a + \frac{(1-\gamma)\alpha_b}{\alpha_\gamma}Q_b \\ R_{1\gamma} &= \gamma R_{1a} + (1-\gamma)R_{1b} \\ R_{2\gamma} &= \gamma R_{2a} + (1-\gamma)R_{2b}\end{aligned}$$
□

[24]The equation $\alpha E_r(\frac{R}{\alpha}, P) = 0$ has multiple solutions; we choose the minimum of those to be α^*, i.e., $\alpha^*(R, 0, P) = \frac{R}{\mathsf{I}(P,W)}$.

$$\begin{aligned}&\gamma E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_a, P, \Pi) + (1-\gamma)E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_b, P, \Pi) \\ &= \min_{\substack{Q_a, R_{1a}, R_{2a}, T_a, Q_b, R_{1b}, R_{2b}, T_b: \\ R_{1a}\geq R_{2a}\geq R \;\; T_a\geq 0 \\ R_{1b}\geq R_{2b}\geq R \;\; T_b\geq 0 \\ \alpha_a\boldsymbol{\zeta}(\frac{R_{1a}}{\alpha_a}, P, Q_a) + R_{2a} - R + T_a \leq E_{\mathbf{x}} \\ \alpha_b\boldsymbol{\zeta}(\frac{R_{1b}}{\alpha_b}, P, Q_b) + R_{2b} - R + T_b \leq E_{\mathbf{x}}}} \begin{array}{l}\gamma\left[\alpha_a\boldsymbol{\zeta}\left(\frac{R_{2a}}{\alpha_a}, P, Q_a\right) + R_{1a} - R + (1-\alpha_a)\Gamma\left(\frac{T_a}{1-\alpha_a}, \Pi\right)\right] \\ +(1-\gamma)\left[\alpha_b\boldsymbol{\zeta}\left(\frac{R_{2b}}{\alpha_b}, P, Q_b\right) + R_{1b} - R + (1-\alpha_b)\Gamma\left(\frac{T_b}{1-\alpha_b}, \Pi\right)\right]\end{array} \\ &\geq \min_{\substack{Q_\gamma, R_{1\gamma}, R_{2\gamma}, T_\gamma: \\ R_{1\gamma}\geq R_{2\gamma}\geq R \;\; T_\gamma\geq 0 \\ \alpha_\gamma\boldsymbol{\zeta}(\frac{R_{1\gamma}}{\alpha_\gamma}, P, Q_\gamma) + R_{2\gamma} - R + T_\gamma \leq E_{\mathbf{x}}}} \alpha_\gamma\boldsymbol{\zeta}\left(\frac{R_{2\gamma}}{\alpha_\gamma}, P, Q_\gamma\right) + R_{1\gamma} - R + (1-\alpha_\gamma)\Gamma\left(\frac{T}{1-\alpha_\gamma}, \Pi\right) \\ &= E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha_\gamma, P, \Pi).\end{aligned} \tag{120}$$

APPENDIX C
$\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi)$, $\forall P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$

Let us first consider a control phase type $\Pi_P(x_1, x_2) = \frac{P(x_1)P(x_2)\mathbb{1}_{\{x_1 \neq x_2\}}}{1-\sum_x (P(x))^2}$ and establish for all $P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi_P) > E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P). \tag{121}$$

First consider

$$\begin{aligned}
&\left(1 - \sum_x (P(x))^2\right) \mathsf{D}\left(U \| W_a | \Pi_P\right) \\
&= \sum_{\substack{x_1, x_2: \\ x_1 \neq x_2}} P(x_1)P(x_2) \sum_y U(y|x_1, x_2) \log \frac{U(y|x_1, x_2)}{W(y|x_1)} \\
&= \sum_{\substack{x_1, x_2: \\ x_1 \neq x_2}} P(x_1)P(x_2) \sum_y U(y|x_1, x_2) \log \frac{U(y|x_1, x_2)}{V_U(y|x_1)} \\
&\quad - \sum_{\substack{x_1, x_2: \\ x_1 \neq x_2}} P(x_1)P(x_2) \sum_y U(y|x_1, x_2) \log \frac{V_U(y|x_1)}{W(y|x_1)} \\
&\geq \mathsf{I}\left(P, \hat{V}_U\right) + \mathsf{D}\left(V_U \| W | P\right)
\end{aligned} \tag{122}$$

where the last step follows from the log sum inequality, and transition probability matrices V_U and $\hat{V}_U$ are given by

$$\begin{aligned}
V_U(y|x_1) &= W(y|x_1)P(x_1) + \sum_{x_2: x_2 \neq x_1} U(y|x_1, x_2)P(x_2) \\
\hat{V}_U(y|x_2) &= W(y|x_2)P(x_2) + \sum_{x_1: x_1 \neq x_2} U(y|x_1, x_2)P(x_1).
\end{aligned}$$

Using a similar line of reasoning, we get

$$\mathsf{D}\left(U \| W_r | \Pi_P\right) \geq \frac{\mathsf{D}\left(\hat{V}_U \middle\| W|P\right) + \mathsf{I}\left(P, V_U\right)}{1 - \sum_x (P(x))^2}. \tag{123}$$

Note that for all $P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ if we use the inequalities (122) and (123) together with the definition of $E_{\mathbf{e}}$ given in (16) and (21), we get

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha(R, E_{\mathbf{x}}), P, \Pi_P) \geq E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) + \delta_P$$

for some $\delta_P > 0$. Consequently, for all $P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$, (121) holds.

Note that for all Π and for all $P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) = E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi).$$

Thus, for all $P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$, we have

$$\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi). \tag{124}$$

ACKNOWLEDGMENT

The authors would like to thank E. Telatar for his encouragement on the problem and for numerous discussions on error-free codes. In particular, the observations presented about z-channels are his and Lemma 5 was proved during summer 2006 at Ecole Polytechnique Federale de Lausanne (EPFL). They would also like to thank T. Weissman and A. Lapidoth for bringing the Shannon–Gallager result mentioned in Elwyn Berlekamp's thesis to their attention; A. Sahai for various discussions on communication problems with feedback; and R. G. Gallager for various discussions on the encoding scheme presented in Section III and the two-message error-exponent tradeoff. The authors would also like to acknowledge the thorough review provided by the anonymous reviewer B, which has helped them improve the presentation of this paper in general. In addition, the $(1 - s)$ factor on the left-hand side of Theorem 2 was pointed out to the authors by reviewer B.

REFERENCES

[1] E. R. Berlekamp, "Block coding with noiseless feedback," Ph.D. dissertation, Dept. Electr. Eng., Massachusetts Inst. Technol., Cambridge, MA, 1964.
[2] 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.
[3] M. V. Burnashev, "Data transmission over a discrete channel with feedback, random transmission time," *Probl. Inf. Transm.*, vol. 12, no. 4, pp. 10–30, 1976.
[4] M. V. Burnashev, "Note: On the article data transmission over a discrete channel with feedback, random transmission time," *Probl. Inf. Transm.*, vol. 13, no. 1, pp. 108–108, 1977.
[5] M. V. Burnashev, "Sequential discrimination of hypotheses with control of observations," *Math. USSR Izvestija*, vol. 15, no. 3, pp. 419–440, 1980.
[6] M. V. Burnashev, "On the reliability function of a binary symmetrical channel with feedback," *Probl. Inf. Transm.*, vol. 24, no. 1, pp. 3–10, 1988.
[7] M. V. Burnashev and H. Yamamoto, "On the zero-rate error exponent for a BSC with noisy feedback," *Probl. Inf. Transm.*, vol. 44, no. 3, pp. 198–213, 2008.
[8] M. V. Burnashev and H. Yamamoto, "On the reliability function for a BSC with noisy feedback," *Probl. Inf. Transm.*, vol. 46, no. 2, pp. 103–121, 2010.
[9] I. Csiszàr, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2502–2523, Oct. 1998.
[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, FL: Academic, 1982.
[11] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Kibernetiki*, vol. 8, pp. 161–168, 1962.
[12] S. Draper and A. Sahai, "Variable-length coding with noisy feedback," *Eur. Trans. Telecommun.*, vol. 19, no. 4, pp. 355–370, May 2008.
[13] A. G. D'yachkov, "Upper bounds on the error probability for discrete memoryless channels with feedback," *Probl. Inf. Transm.*, vol. 11, no. 4, pp. 13–28, 1975.
[14] G. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 2, pp. 206–220, Mar. 1968.
[15] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by Schalkwijk and Kailath," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 6–17, Jan. 2010.
[16] P. K. Gopala, Y.-H. Nam, and H. E. Gamal, "On the error exponents of ARQ channels with deadlines," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4265–4273, Nov. 2007.
[17] E. A. Haroutunian, "A lower bound on the probability of error for channels with feedback," *Probl. Inf. Transm.*, vol. 13, no. 2, pp. 36–44, 1977.
[18] E. Hof, I. Sason, and S. Shamai, "Performance bounds for erasure, list, and decision feedback schemes with linear block codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3754–3778, Aug. 2010.
[19] Y.-H. Kim, A. Lapidoth, and T. Weissman, "Error exponents for the Gaussian channel with active noisy feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1223–1236, Mar. 2011.
[20] N. Merhav, "Error exponents of erasure list decoding revisited via moments of distance enumerators," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4439–4447, Oct. 2008.

[21] N. Merhav and M. Feder, "Minimax universal decoding with an erasure option," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1664–1675, May 2007.

[22] P. Moulin, "A Neyman-Pearson approach to universal erasure and list decoding," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4462–4478, Oct. 2009.

[23] B. Nakiboğlu, "Exponential bounds on error probability with feedback," Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 2011.

[24] B. Nakiboğlu and R. G. Gallager, "Error exponents for variable-length block codes with feedback and cost constraints," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 945–963, Mar. 2008.

[25] M. S. Pinsker, "The probability of error in block transmission in a memoryless Gaussian channel with feedback," *Probl. Inf. Transm.*, vol. 4, no. 4, pp. 1–4, 1968.

[26] E. Sabbag and N. Merhav, "Achievable error exponents for channels with side information—Erasure and list decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5424–5431, Nov. 2010.

[27] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback—Part II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 183–189, Apr. 1966.

[28] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—Part I: No bandwidth constraint," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 172–182, Apr. 1966.

[29] C. E. Shannon, "The zero error capacity of a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.

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

[31] I. E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, 1992.

[32] I. E. Telatar and R. G. Gallager, "New exponential upper bounds to error and erasure probabilities," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 1994, DOI: 10.1109/ISIT.1994.394640.

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

[34] K. S. Zigangirov, "Upper bounds for the error probability for channels with feedback," *Probl. Inf. Transm.*, vol. 6, no. 2, pp. 87–92, 1970.

[35] K. S. Zigangirov, "Optimum zero rate date transmission through binary symmetric channel with feedback," *Probl. Control Inf. Theory*, vol. 7, no. 3, pp. 21–35, 1978.

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, in 2005 and 2011, respectively.

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

Since 2002, he has been with the Department of Electrical Engineering and Computer Sciences, Massachusetts Institute of Technology (MIT), Cambridge, where he is currently an Associate Professor. His research interests include information theory, wireless communications, and wireless networks.

Dr. Zheng received the Eli Jury award from the University of California at Berkeley in 2002, the IEEE Information Theory Society Paper Award in 2003, the National Science Foundation (NSF) CAREER award in 2004, and the AFOSR Young Investigator Award in 2007. He is currently an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY.