Exponential Bounds On Error Probability With Feedback

by

Barış Nakiboğlu

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2011

Author ....
Department of Electrical Engineering and Computer Science
January 27, 2011

Certified by ....
Lizhong Zheng
Associate Professor
Thesis Supervisor

Accepted by ....
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Exponential Bounds On Error Probability With Feedback
by
Barış Nakiboğlu

Submitted to the Department of Electrical Engineering and Computer Science
on January 27, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Feedback is useful in memoryless channels for decreasing complexity and increasing reliability; the capacity of the memoryless channels, however, can not be increased by feedback. For fixed length block codes even the decay rate of error probability with block length does not increase with feedback for most channel models. Consequently for making the physical layer more reliable for higher layers one needs go beyond the framework of fixed length block codes and consider relaxations like variable-length coding, error- erasure decoding. We strengthen and quantify this observation by investigating three problems.

1. *Error-Erasure Decoding for Fixed-Length Block Codes with Feedback:* Error-erasure codes with communication and control phases, introduced by Yamamoto and Itoh, are building blocks for optimal variable-length block codes. We improve their performance by changing the decoding scheme and tuning the durations of the phases, and establish inner bounds to the tradeoff between error exponent, erasure exponent and rate. We bound the loss of performance due to the encoding scheme of Yamamoto-Itoh from above by deriving outer bounds to the tradeoff between error exponent, erasure exponent and rate both with and without feedback. We also consider the zero error codes with erasures and establish inner and outer bounds to the optimal erasure exponent of zero error codes. In addition we present a proof of the long known fact that, the error exponent tradeoff between two messages is not improved with feedback.
2. *Unequal Error Protection for Variable-Length Block Codes with Feedback:* We use Kudrayashov's idea of implicit confirmations and explicit rejections in the framework of unequal error protection to establish inner bounds to the achievable pairs of rate vectors and error exponent vectors. Then we derive an outer bound that matches the inner bound using a new bounding technique. As a result we characterize the region of achievable rate vector and error exponent vector pairs for bit-wise unequal error protection problem for variable-length block codes with feedback. Furthermore we consider the single message message-wise unequal error protection problem and determine an analytical expression for the missed detection exponent in terms of rate and error exponent, for variable-length block codes with feedback.
3. *Feedback Encoding Schemes for Fixed-Length Block Codes:* We modify the analysis technique of Gallager to bound the error probability of feedback encoding schemes. Using the encoding schemes suggested by Zigangirov, D'yachkov and Burnashev we recover or improve all previously known lower bounds on the error exponents of fixed-length block codes.

Thesis Supervisor: Lizhong Zheng
Title: Associate Professor

to my brother, my parents and my grandmother...

Contents

Chapter 1

Introduction

The effects of feedback in point to point communication problem have been studied from the early days of the information theory. Arguably the most important qualitative conclusion of those studies has been that, for memoryless channels at high rates, in order to use feedback to increase the decay rate of error probability with block length, one needs to go beyond the framework of fixed-length block codes and change the way constraints are imposed on cost, decoding time, decoding algorithm or encoding rules. In other words, one needs to consider variable-length block codes, error-erasure decoding or non-block encoding schemes in order to use feedback to decrease error probability substantially, at least at high rates. Throughout the thesis we will strengthen and quantify this conclusion in a number of ways on different problems. First in Chapter 2 we analyze the error performance of fixed-length block codes with error-erasure decoding and establish inner and outer bounds to the optimal error exponent erasure exponent tradeoff. In Chapter 3 we consider variable-length block codes with feedback for the transmission of a message composed of multiple groups of bits each with different priority and determine the achievable pairs of rate vector and error exponent vector. Then in Chapter 4 we return to the framework of fixed-length block codes and suggest an analysis technique and encoding scheme which increases the error exponent at low rates. We explain below some of the results which suggest going beyond the framework of fixed-length block codes and the origins of some of the ideas enhanced in later chapters. More detailed literature surveys, specific to the problems investigated, are given at the beginning of each chapter.

The channel capacity of a memoryless channel is not increased by the addition of a noiseless and delay-free feedback link from the receiver to the transmitter. This was first shown by Shannon [37] for discrete memoryless channels (DMCs). Furthermore Dobrushin [13] for symmetric DMCs and Pinsker [32] for additive white Gaussian noise channel (AWGNC) showed that even the exponential decay rate of error probability with block length, i.e. error exponent, does not increase with feedback for rates over the critical rate. Both [32] and [13] established the sphere packing exponent as an upper bound to the error exponent at all rates[1] smaller than channel capacity.

The result of Pinsker [32] seems to dispute widely known results of Schalkwijk and Kailath [35] and Schalkwijk [34], according to which error probability can decay as a doubly exponential function of the block length. Similarly, Dobrushin's result [13], seems to disagree

[1]Later Haroutunian, [20], established an upper bound on the error exponent of block codes with feedback. His upper bound is equal to the sphere packing exponent for symmetric channels but it is strictly larger than the sphere packing exponent for non-symmetric channels. It is still not known, whether or not the same is true for general DMCs.

with Burnashev's results in [4], which claims the existence of a scheme for DMCs which allows for an exponent much larger than sphere packing exponent. These contradictions however are illusive, because the models used in these papers are different. In [32] the power constraint $\mathcal{P}$ on the total energy spent for transmission $\mathcal{S}_{\mathsf{n}}$ holds with probability one, i.e. $\mathbf{P}[\mathcal{S}_{\mathsf{n}} \leq \mathcal{P}\mathsf{n}] = 1$; whereas in [35] and [34] the power constraint is on the expected value of energy spent on a block $\mathbf{E}[\mathcal{S}_{\mathsf{n}}]$, i.e. the power constraint is of the form $\mathbf{E}[\mathcal{S}_{\mathsf{n}}] \leq \mathcal{P}\mathsf{n}$. In a similar fashion, both in [13] and [20], fixed-length block codes are considered, i.e. the duration of the transmission n is fixed and does not change with channel output; whereas in [4] variable-length block codes are considered, i.e. the duration of the transmission τ is a random variable which depends on the channel output and the error exponent is defined as the decay rate of error probability with $\mathbf{E}[\tau]$.

Though the contradictions are illusive, the difference in the behavior of error exponent for different models, or more precisely different families of codes, is real. By changing the architecture of the communication system and allowing for transmission time to be variable, one can obtain substantial gains in terms of error performance.

1.1 Block Codes: Variable-Length vs Fixed-Length with Erasures

The main result of [4] was promising, but the encoding scheme suggested to achieve that performance was fairly complicated and hard to implement. Later Yamamoto and Itoh [43] suggested a much simpler scheme to achieve the optimal performance. They considered a fixed-length block code with error-erasure decoding and two phases. In the communication phase, the message is transmitted with a non-feedback code and a temporary decision is made by the receiver at the end of the phase. In the control phase, the transmitter tries to confirm the temporary decision if it is correct and deny it if it is not. The receiver decodes to the temporary decision or declares an erasure depending on the control phase channel output. Yamamoto and Itoh [43] showed that, by using this fixed-length block code repeatedly until a non-erasure decoding happens, one can achieve the optimal performance for variable-length block codes.

The result of Yamamoto and Itoh [43] is interesting in a number of ways. First it demonstrates that the error exponent of fixed-length block codes can be as high as that of variable-length block codes if their decoders are error-erasure decoders. Secondly it demonstrates that a very limited use of feedback, merely to inform the transmitter about the receiver's temporary decision is good enough. Furthermore the Yamamoto-Itoh scheme achieves the optimal performance by using a code, in which communication and the bulk of error correction are decoupled for both encoding and decoding purposes:

(a) The message is transmitted in the communication phase and confirmed in the control phase.

(b) Temporary decoding is made using just the output of the communication phase. The decision between declaring an erasure and decoding to the temporary decision is made using solely the control phase output.

In [26] and [27] we have generalized all these results to channels with cost constraints and potentially infinite alphabets.

Unfortunately these observations which simplify the architecture of the communication scheme hold only when the erasure probability is desired to decay to zero slowly, i.e. subexponentially. When the erasure probability is desired to decay exponentially with block length, the Yamamoto-Itoh scheme is not necessarily optimal, because of the decoupling in the encoder (a) and the decoupling the decoder (b), discussed above. In Chapter 2 we address this issue by considering the situation when erasure probability is decreasing with a positive exponent.

Finding an encoder that uses feedback properly, even for the case when there are no-erasures, is a challenging problem. Thus we do not attempt to find the optimal encoder. Instead we use the Yamamoto and Itoh encoding scheme with the inherent decoupling mentioned in (a), but tune the relative durations of the phases in the encoding scheme and get rid of the decoupling in the decoder mentioned in (b). We use a fixed-length block code with communication and control phases, like Yamamoto and Itoh [43], together with a decoder, that uses the outputs of both communication and control phases while deciding between declaring an erasure and decoding to the temporary decision. The inner bound obtained for such an encoding scheme is better than the best inner bound for the error exponent, erasure exponent, rate tradeoff for non-feedback schemes found previously, [15], [41], [42].

In order to bound the loss in performance because of the particular family of encoding schemes we have used in the inner bounds, we derive outer bounds to the error exponent, erasure exponent, rate tradeoff that are valid for all fixed-length block codes. For doing that we first generalize the straight line bound of Shannon Gallager and Berlekamp [38] from erasure-free decoders to error-erasure decoders, using the fact that the region of achievable error probability, erasure probability pairs for a given block length, message set size and list decoding size triple is convex. Then we recall the outer bounds to error exponents in two related problems:

- The error exponents of erasure free block codes with feedback
- The error exponent tradeoff between two messages with feedback

We use the generalized straight line bound to combine the outer bounds on these two related problems into a family of outer bounds on the error exponent, erasure exponent, rate tradeoff for fixed length block codes with feedback encoders and error-erasure decoders.

The inner and outer bounds derived in Chapter 2 on error exponent, erasure exponent, rate tradeoff will allow us to bound the loss in error exponent at a given rate and erasure exponent because of the two phase scheme we have assumed.

In the last part of the Chapter 2 we investigate the problem of finding the optimal erasure exponent for zero-error codes as a complement to the analysis of the tradeoff between error exponent and erasure exponent, and derive inner and outer bounds to the optimal erasure exponent for a given rate.

1.2 Non-Block Encoding Schemes and Bit-Wise *UEP*

The Yamamoto-Itoh scheme has the peculiar property that it almost always sends the same letter in the control phase. Yet the Yamamoto-Itoh scheme spends a considerable part of its time in the control phase. Such an explicit control phase exists in order to ensure that each message is decoded before the transmission of the next message starts. The Yamamoto-Itoh scheme is a block coding scheme as a result of this property. The schemes in which transmission of successive messages are overlapping are non-block encoding schemes. If non-block encoding schemes are allowed one can drop the explicit control phase all together and use the implicit-confirmation-explicit-rejection protocol suggested by Kudrayshov [22] in order to increase the exponential decay rate of error probability with expected delay.

In a communication system the transmitter transmits successive messages $\mathsf{M}_1, \mathsf{M}_2, \ldots$ and the receiver decodes them as $\hat{\mathsf{M}}_1, \hat{\mathsf{M}}_2, \ldots$. In a block coding scheme disjoint time intervals are allocated for different messages, i.e. transmission of M_k starts only after $\hat{\mathsf{M}}_{k-1}$ is decoded. This is why block coding schemes with feedback use explicit control phases in order to decrease the error probability. However if the transmission of successive messages are allowed to overlap, one can use the following scheme to decrease error probability without an explicit control phase. The transmitter sends M_1 using a fixed-length block code of rate R, and the receiver makes a temporary decision $\tilde{\mathsf{M}}_1$. If $\mathsf{M}_1 = \tilde{\mathsf{M}}_1$ then the transmitter sends M_2 and the receiver makes a temporary decision $\tilde{\mathsf{M}}_2$ and so on. After the first incorrect temporary decision, the transmitter starts sending the special codeword Ξ, until it is detected by the receiver. Once it is detected by the receiver, the transmitter starts re-sending the last message and the previous ℓ messages from scratch. Thus if the j^{th} temporary decision is incorrect and the special codeword Ξ is sent without detection for k times, $|k-\ell|^+$ messages[2] would be in error. The decoder decodes to a temporary decision if the following ℓ temporary decisions are all ordinary codewords, i.e. $\hat{\mathsf{M}}_j = \tilde{\mathsf{M}}_j$ if $\tilde{\mathsf{M}}_t \neq \Xi$ for $t \in \{j, j+1, \ldots j+\ell\}$. Using an implicit confirmation and explicit rejection protocol like the described one above, Kudryashov [22] showed that non-block encoding schemes can have much faster decay of error probability with expected delay than the block encoding schemes.

The implicit confirmation explicit rejection protocols are also useful for block coding schemes. But in order to appreciate that, one needs to consider unequal error protection problems. Consider for example the situation where each message, M_j, is composed of two groups of bits:

$$\mathsf{M}_j = (\mathsf{M}_j^{(a)}, \mathsf{M}_j^{(b)}) \qquad\qquad j \in \{1, 2, 3, \ldots\}$$

where $\mathsf{M}^{(a)}$'s require a better protection than $\mathsf{M}^{(b)}$'s. One way of giving that protection is using an implicit confirmation explicit rejection protocol in a three phase scheme as follows. The transmitter first sends $\mathsf{M}^{(a)}$ and a temporary decision $\tilde{\mathsf{M}}^{(a)}$ is made at the receiver. If $\tilde{\mathsf{M}}^{(a)} = \mathsf{M}^{(a)}$ then in the second phase $\mathsf{M}^{(b)}$ is sent, if not the special codeword Ξ is sent. At the end of the second phase a temporary decision $\tilde{\mathsf{M}}^{(b)}$ is made for $\mathsf{M}^{(b)}$. The third phase is an explicit control phase in order to confirm or reject $(\tilde{\mathsf{M}}^{(a)}, \tilde{\mathsf{M}}^{(b)})$. At the end of the third phase either an erasure is declared or $\mathsf{M} = (\mathsf{M}^{(a)}, \mathsf{M}^{(b)})$ is decoded as $\hat{\mathsf{M}} = (\tilde{\mathsf{M}}^{(a)}, \tilde{\mathsf{M}}^{(b)})$. In Chapter 3 we present a detailed analysis of this scheme for the case with k layers instead of just two layers and obtain a sufficient condition for the achievablity of a $(\vec{R}, \vec{E})$ pair, in terms of the relative durations of the phases.

Implicit confirmation and explicit rejection protocols provide us a sufficient condition for the achievablity of a $(\vec{R}, \vec{E})$ pair. In order to prove that the above architecture is optimal, we prove that the sufficient condition for the achievablity of a $(\vec{R}, \vec{E})$ pair is also a necessary

[2]
$$|s|^+ = \begin{cases} 0 & s < 0 \\ s & s \geq 0 \end{cases}.$$

condition. This is done using a new technique in Chapter 3. Previously in order to derive such outer bounds the average error probability associated with a query posed at stopping time is used [2], [3]; instead we use the missed detection probability of a hypothesis chosen at a stopping time. This gives us a necessary condition for the achievablity of a $(\vec{R}, \vec{E})$ pair in terms of the relative durations of the phases, which is identical to the sufficient condition. Thus we conclude that in a bit-wise unequal error protection problem for variable-length block codes, communication and error correction of each layer of bits can be decoupled both for the purposes of encoding and for the purposes of decoding. Furthermore error correction phase of each layer can be combined with the communication and error correction phases of less important layers using implicit confirmation explicit rejection protocols.

In order to introduce some of the ideas in a simpler form we also investigate the message-wise unequal error protection problem for the single message case In Chapter 3. In that problem we are interested in the minimum conditional error probability that a message can have when the overall rate is R and the overall error exponent is E. We determine the exponent of the minimum error probability message, which is called the missed detection exponent, for any rate R and any error exponent E for variable-length block codes with feedback.

The results of Chapter 3 generalize the corresponding results in [3], which were derived for the case when the overall rate is (very close to) capacity. In Chapter 3 there is no such assumption and we calculate the tradeoffs for the whole rate region.

1.3 Error Probability and Posterior Matching for Fixed-Length Block Codes

The interest in error-erasure decoding and variable-length block codes were partly because of the negative results about the error performance of fixed-length block codes with feedback. But those negative results about error exponents of fixed-length block codes with feedback imply only that there can not be any improvement, at high rates i.e. rates above the critical rate[3] R_{crit}. For the rates below the critical rate there are encoding schemes that improve the error exponents in binary symmetric channels, [44], [7], k-ary symmetric channels [14] and binary input channels [14].

The encoding schemes of [44] and [14] are matching schemes. The messages are equally likely at the beginning of the transmission both in [44] and [14]. At each time step the encoding scheme tries to match an input distribution on the input alphabet as closely as it can with the given pseudo posterior probabilities of the messages. After the observation of each channel output the pseudo posterior probabilities of the messages are updated according to a channel which is "noisier" than the actual one. In binary symmetric channels such a scheme is optimal in all rates below capacity except some interval of the form $(0, R_{Zcrit})$ where $R_{Zcrit} < R_{crit}$. In general binary input channels and k-ary symmetric channels performance of this scheme is better than that of random coding. The principle insight of [44] and [14] is that by using a "noisier" channel in updating the posterior probabilities, one can improve the performance of the posterior matching schemes.

In Chapter 4 we suggest an alternative analysis technique, based on the error analysis technique of Gallager [16] to achieve similar conclusions. We demonstrate that instead of working with pseudo posteriors, the encoder can work with the actual posterior probabilities

[3]The critical rate, R_{crit}, is the rate at which the random coding exponent and sphere packing exponent diverge from one another.

of the messages. However, in order to do so the encoder needs to apply a titling to the posterior probabilities before the matching. Tilting the posterior probabilities before the matching and using the pseudo posteriors calculated assuming a noisier channel have the same effect; both operations damp down the effect of past observations in the matching. Using an analysis technique based on the analysis technique of Gallager [16] we recover the results of [44] and [14] for binary input channels and improve the results of [14] for k-ary symmetric channels.

For binary symmetric channels Burnashev [7] has suggested a modification to the encoding scheme of Zigangirov given in [44]. Burnashev's modification improves the performance in the range $(0, R_{Zcrit})$. In the second part of Chapter 4 we extend the results of [7] to a broader class of DMCs. This is done by modifying the analysis technique we have suggested in the first part of the chapter in order to accommodate the modified encoding scheme of Burnashev [7].

Chapter 2

Error-Erasure Decoding for Block Codes with Feedback[1]

In this chapter we investigate the performance of fixed length block codes on discrete memoryless channels (DMCs) with feedback and error-erasure decoding. We derive inner and outer bounds to the optimal tradeoff between the rate, the erasure exponent and the error exponent. This analysis complements on one hand the results on error-erasure decoding without feedback and on the other hand the results on variable-length block codes with feedback.

We start with a brief overview of previous results on variable-length block codes and error-erasure decoding without feedback in Section 2.1. This will motivate the problem at hand and relate it to the previous studies. Then in Section 2.2, we introduce the channel model and block codes with error-erasure decoding formally. After that, in Section 2.3, we derive an inner bound bound using a two-phase coding algorithm (similar to the one described by Yamamoto and Itoh in [43]) combined with a decoding rule and analysis techniques inspired by Telatar in [41] for the non-feedback case. The analysis and the decoding rule in [41] is tailored for a single phase scheme without feedback, and the two-phase scheme of [43] is tuned specifically to the zero-erasure exponent; coming up with a framework in which both of the ideas can be used efficiently is the main technical challenge here. In Section 2.4, we first generalize the straight line bound introduced by Shannon, Gallager and Berlekamp in [38] to block code which have decoding rules with erasures. This is then combined with the error exponent tradeoff between two codewords with feedback to establish an outer bound. In Section 2.5, we introduce error free block codes with erasures, discuss their relation to block codes with errors and erasures and present inner and outer bounds to their erasure exponents.

2.1 Error-Erasure Decoding and Variable-Length Block Codes

Burnashev [4], [5], [6] considered variable-length block codes with feedback and determined the decay rate of their error probability with expected transmission time exactly for all rates. Later Yamamoto and Itoh [43] showed that the optimal performance for variable-length block codes with feedback can be achieved by using an appropriate fixed-length block code with error-erasure decoding repetitively until a non-erasure decoding occurs.[2]

[1] Results presented in this chapter have been reported previously in [28].

[2] Note that introducing erasures can not decrease the optimal error exponent and any variable-length block code with erasure can be used as a variable-length block code without erasures with feedback, simply

In fact any fixed-length block code with erasures can be used repetitively, as done in [43], to get a variable-length block code with essentially the same error exponent as the original fixed-length block code. Thus [4] 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 fixed-length block codes with erasures described in [43], when erasure probability decays to zero sub-exponentially with block length. However, the techniques used in this stream 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 paragraph below, the case with strictly positive erasure exponents is important both for engineering applications and for a more comprehensive understanding of error-erasure 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, while recovering 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. In this respect the probability of erasure is a criterion for the quality of service of the physical layer. The average transmission time is only a first order measure of the burden of the retransmissions to the higher layers: 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 [43], the cost of retransmissions are 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 the error-erasure exponent tradeoff is useful in coming up with designs to meet those constraints. An example of this phenomena is variable-length block coding schemes with hard deadlines for decoding time, which have already been investigated by Gopala *et. al.* [18] for block codes without feedback. They have used a block coding scheme with erasures and they have resent the message whenever an erasure occurs. But because of the hard deadline, they employed this scheme only for some fixed number of trials. If all the trials prior to the last one fail, i.e. lead to an erasure, they use a non-erasure 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 stream of research we complement with our investigation: error-erasure decoding for block codes without feedback. Forney [15] was the first one to consider error-erasure decoding without feedback. He obtained an achievable tradeoff between the exponents of error and erasure probabilities. Then Csiszár and Körner, [12] achieved the same performance using universal coding and decoding algorithms. Later Telatar and Gallager [42] introduced a strict improvement on certain channels over the results presented in [15] and [12]. Recently there has been revived interest in errors and erasures decoding for universally achievable performance [25], [24], for alternative methods of analysis [23], for extensions to channels with side information [33] and implementation

by retransmitting the message at hand whenever there is an erasure. Using these two observations one can check in a few lines that the best error exponent for the variable-length block codes with feedback is the same with and without erasures.

with linear block codes [21]. The encoding schemes in these codes do not have access to any feedback. However if the transmitter learns whether the decoded message was an erasure or not, it can resend the message whenever it is erased. Because of this block retransmission variant these problems are sometimes called decision feedback problems.

2.2 Channel Model and Reliable Sequences For Error-Erasure Decoding

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 are 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 discrete memoryless channel characterized by an $|\mathcal{X}|$-by-$|\mathcal{Y}|$ transition probability matrix W.

$$\mathbf{P}\left[\mathsf{Y}_t | \mathsf{X}^t, \mathsf{Y}^{t-1}\right] = \mathbf{P}[\mathsf{Y}_t | \mathsf{X}_t] = W(\mathsf{Y}_t|\mathsf{X}_t) \qquad \forall t. \tag{2.1}$$

The feedback channel is noiseless and delay free, i.e. the transmitter observes Y_{t-1} before transmitting X_t.

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 $\Phi_t(\mathsf{M}, \mathsf{Y}^{t-1})$ is sent. The sequence of functions $\Phi_t(\cdot) : \mathcal{M} \times \mathcal{Y}^{t-1}$ which assigns an input symbol for each $m \in \mathcal{M}$ and $y^{t-1} \in \mathcal{Y}^{t-1}$ is called the encoding function.

After receiving Y^n the receiver decodes $\hat{\mathsf{M}}(\mathsf{Y}^\mathsf{n}) \in \{\mathbf{x}\} \cup \mathcal{M}$ where $\mathbf{x}$ is the erasure symbol. The conditional erasure and error probabilities $P_{\mathbf{x}|\mathsf{M}}$ and $P_{\mathbf{e}|\mathsf{M}}$ and unconditional erasure and error probabilities, $P_\mathbf{x}$ and $P_\mathbf{e}$ are defined as,

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

Since all the messages are equally likely,

$$P_\mathbf{x} = \tfrac{1}{|\mathcal{M}|} \sum\nolimits_{m\in\mathcal{M}} P_{\mathbf{x}|m} \qquad\qquad P_\mathbf{e} = \tfrac{1}{|\mathcal{M}|} \sum\nolimits_{m\in\mathcal{M}} P_{\mathbf{e}|m}$$

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

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

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

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

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

Haroutunian, [20, Theorem 2], established a strong converse for erasure free block codes with feedback which implies that $\lim_{\mathsf{n}\to\infty}(P_{\mathbf{e}}{}^{(\mathsf{n})} + P_{\mathbf{x}}{}^{(\mathsf{n})}) = 1$ 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 non-negative erasure exponents $E_{\mathbf{x}}$, the (true) error exponent $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ of fixed-length block codes with feedback is defined to be the best error exponent of the reliable sequences[3] 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, $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ is,*

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \triangleq \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R, E_{\mathbf{x}\mathcal{Q}} \geq E_{\mathbf{x}}} E_{\mathbf{e}\mathcal{Q}}. \tag{2.2}$$

Note that

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \mathcal{E}(R) \qquad \forall E_{\mathbf{x}} > \mathcal{E}(R) \tag{2.3}$$

where $\mathcal{E}(R)$ is the (true) error exponent of erasure-free block codes on DMCs with feedback.[4] Thus the benefit of error-erasure 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 and for all channels is still an open problem; only upper and lower bounds to $\mathcal{E}(R)$ are known. In this chapter we are only interested in quantifying the gains of error-erasure decoding with feedback instead of finding $\mathcal{E}(R)$. We will analyze the performance of generalizations of the simple two-phase schemes that are known to be optimal when the erasure exponent is zero. In order to quantify how much is lost by using such a restricted architecture we will derive general outer bounds to $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ and compare them with the inner bounds.

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

$$E_r(R, P) = \min_{V} \mathsf{D}\left(V \| W|P\right) + |\mathsf{I}\left(P, V\right) - R|^{+} \qquad E_r(R) = \max_{P} E_r(R, P) \tag{2.4}$$

$$E_{sp}(R, P) = \min_{V: \mathsf{I}(P,V) \leq R} \mathsf{D}\left(V \| W|P\right) \qquad E_{sp}(R) = \max_{P} E_{sp}(R, P) \tag{2.5}$$

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:

$$\mathsf{D}\left(V \| W|P\right) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x) V(y|x) \ln \tfrac{V(y|x)}{W(y|x)}$$

$$\mathsf{I}\left(P, V\right) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(x) V(y|x) \ln \tfrac{V(y|x)}{\sum_{\tilde{x} \in \mathcal{X}} P(\tilde{x}) V(y|\tilde{x})}$$

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 $\mathsf{supp}P$.

---

[3] We only consider 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.

[4] In order to see this consider a reliable sequence with erasures $\mathcal{Q}$ and replace its decoding algorithm by any erasure free one such that $\hat{\mathsf{M}}'(y^{\mathsf{n}}) = \hat{\mathsf{M}}(y^{\mathsf{n}})$ if $\hat{\mathsf{M}}(y^{\mathsf{n}}) \neq \mathbf{x}$ to obtain an erasure free reliable sequence, $\mathcal{Q}'$. Then $P_{\mathbf{e}}{}^{(\mathsf{n})}_{\mathcal{Q}'} \leq P_{\mathbf{x}}{}^{(\mathsf{n})}_{\mathcal{Q}} + P_{\mathbf{e}}{}^{(\mathsf{n})}_{\mathcal{Q}}$; 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 equation (2.3).

2.3 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 Itoh in [43] together with a decoding rule similar to the one described by Telatar in [41]. First, in the communication phase, the transmitter uses a fixed-composition code of length $\alpha \mathsf{n}$ and rate $\frac{R}{\alpha}$. At the end of the communication 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. Then in the $(\mathsf{n}-\mathsf{n}_1)$ long control 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 control phase if the tentative decision dominates all other messages, the receiver decodes to the tentative decision, if not the receiver declares an erasure. The word "dominate" will be made precise later in Section 2.3.2.

At the encoder our scheme is similar to that of Yamamoto and Itoh [43], in that the communication and the bulk of the error correction are done in two disjoint phases. At the decoder, however, unlike [43] we consider the channel outputs in both of the phases while deciding between declaring an erasure and decoding to the tentative decision.

In the rest of this section, we analyze the performance of this coding architecture and derive the 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 $E_{\mathbf{x}}$.

2.3.1 Fixed-Composition Codes and the Packing Lemma

Let us start with a very brief overview of certain properties of types, a thorough handling of types can be found in [12]. The empirical distribution of $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ is called the type of x^{n} and the empirical distribution of transitions from $x^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ to $y^{\mathsf{n}} \in \mathcal{X}^{\mathsf{n}}$ is called the conditional type:[5]

$$\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) \triangleq \tfrac{1}{\mathsf{n}} \sum\nolimits_{t=1}^{\mathsf{n}} \mathbb{1}_{\{x_t=\tilde{x}\}} \qquad \tilde{x} \in \mathcal{X}. \tag{2.6}$$

$$\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}(\tilde{y}|\tilde{x}) \triangleq \tfrac{1}{\mathsf{n}\mathsf{P}_{x^{\mathsf{n}}}(\tilde{x})} \sum\nolimits_{t=1}^{\mathsf{n}} \mathbb{1}_{\{x_t=\tilde{x}\}} \mathbb{1}_{\{y_t=\tilde{y}\}} \qquad \forall \tilde{y} \in \mathcal{Y}, \ \ \forall \tilde{x} \text{ s.t. } \mathsf{P}_{x^{\mathsf{n}}}(\tilde{x}) > 0. \tag{2.7}$$

For any probability transition matrix $\hat{V} : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \to \mathcal{Y}$ we have[6]

$$\prod\nolimits_{t=1}^{\mathsf{n}} \hat{V}(y_t|x_t) = e^{-\mathsf{n}\left(\mathsf{D}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} \middle\| \hat{V} \middle| \mathsf{P}_{x^{\mathsf{n}}}\right) + \mathsf{H}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} \middle| \mathsf{P}_{x^{\mathsf{n}}}\right)\right)} \tag{2.8}$$

V-shell of x^{n}, $T_V(x^{\mathsf{n}})$, is the set of all y^{n}'s whose conditional type with respect to x^{n} is V:

$$T_V(x^{\mathsf{n}}) \triangleq \{y^{\mathsf{n}} : \mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} = V\}. \tag{2.9}$$

The total probability of $T_V(x^{\mathsf{n}})$ has to be less than one for any transition probability matrix from $\mathcal{X}$ to $\mathcal{Y}$ and resulting channel. Thus using equation (2.8) for $\hat{V} = V$ we get,

$$|T_V(x^{\mathsf{n}})| \leq e^{\mathsf{n}\mathsf{H}\left(\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}} \middle| \mathsf{P}_{x^{\mathsf{n}}}\right)} \tag{2.10}$$

[5] Note that $\mathsf{P}_{x^{\mathsf{n}}}$ is a distribution on $\mathcal{X}$, whereas $\mathsf{V}_{y^{\mathsf{n}}|x^{\mathsf{n}}}$ is a channel from the support of $\mathsf{P}_{x^{\mathsf{n}}}$ to $\mathcal{Y}$.

[6] Note that for any $\hat{V} : \mathcal{X} \to \mathcal{Y}$ there is unique consistent $\hat{V}' : \mathsf{supp}\mathsf{P}_{x^{\mathsf{n}}} \to \mathcal{Y}$.

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 2.3.4 the error and erasure events are described 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 the V-shell of $x^{\mathsf{n}}(m)$ with the $\hat{V}$-shells of other codewords:

$$F^{(\mathsf{n})}\left(V, \hat{V}, m\right) \triangleq T_V\left(x^{\mathsf{n}}(m)\right) \bigcap \left(\cup_{\tilde{m} \neq m} T_{\hat{V}}\left(x^{\mathsf{n}}(\tilde{m})\right)\right). \tag{2.11}$$

The following packing lemma, proved by Csiszár and Körner [12, Lemma 2.5.1], claims the existence of a code with a guaranteed upper bound on the size of $F^{(\mathsf{n})}\left(V, \hat{V}, m\right)$.

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 $\forall m \in \mathcal{M}$*

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

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 in [12, Lemma 2.5.1], for a fixed δ and large enough n. However, this form follows immediately from their proof.

Using Lemma 1 together with equations (2.8) and (2.10) one 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 getting 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,*

$$\mathbf{P}\left[F^{(\mathsf{n})}\left(V, \hat{V}, \mathsf{M}\right) \middle| \mathsf{M}\right] \leq e^{-\mathsf{n}\eta\left(R,P,V,\hat{V}\right)} \tag{2.12}$$

where

$$\eta\left(R, P, V, \hat{V}\right) \triangleq \mathsf{D}\left(V \| W|P\right) + \left|\mathsf{I}\left(P, \hat{V}\right) - R\right|^{+}. \tag{2.13}$$

2.3.2 Coding Algorithm

For the length $\mathsf{n}_1 = \lceil \alpha \mathsf{n} \rceil$ communication phase, we use a 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 communication 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 more than one codewords have the maximum value of empirical mutual information, the codeword with the lowest index is chosen.[7]

$$\tilde{\mathsf{M}} = \left\{ m : \begin{array}{ll} \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) > \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) & \forall \tilde{m} < m \\ \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}\right) \geq \mathsf{I}\left(P, \mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{m})}\right) & \forall \tilde{m} > m \end{array} \right\} \tag{2.14}$$

[7]The choice of lowest index message for ties is arbitrary; the receiver could decode to any one of the messages with the highest mutual information. Results will continue to hold in that case. Indeed the tradeoff between the exponents will not be effected even if receiver declares erasures whenever such a tie happens.

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.

The 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 time 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.

As was done in Corollary 1, one 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 $W_r(y|\tilde{x}, \tilde{\tilde{x}}) = W(y|\tilde{\tilde{x}})$. Note that $|T_U| \le e^{(\mathsf{n}-\mathsf{n}_1)\mathsf{H}(U|\Pi)}$, thus

$$\mathbf{P}\left[T_U \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(a)\right] \le e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_a|\Pi)} \tag{2.15a}$$
$$\mathbf{P}\left[T_U \middle| \mathsf{X}^{\mathsf{n}}_{\mathsf{n}_1+1} = x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right] \le e^{-(\mathsf{n}-\mathsf{n}_1)\mathsf{D}(U\|W_r|\Pi)}. \tag{2.15b}$$

2.3.3 Decoding Rule

For an encoder like the one in Section 2.3.2, a decoder that depends only on the conditional type of $\mathsf{Y}^{\mathsf{n}_1}$ for different codewords in the communication phase, i.e. $\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}_1}(m)}$ 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 an analytically intractable problem. Instead we will only consider 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.[8]

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.

$$\hat{\mathsf{M}} = \begin{cases} \tilde{\mathsf{M}} & \text{if } \forall m \neq \tilde{\mathsf{M}} \quad (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{\mathsf{M}})}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m) \\ \mathbf{x} & \text{if } \exists m \neq \tilde{\mathsf{M}} \text{ s.t. } (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(\tilde{\mathsf{M}})}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, \tilde{\mathsf{M}}) \not\succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1}|x^{\mathsf{n}}(m)}, \mathsf{U}_{\mathsf{Y}^{\mathsf{n}}_{\mathsf{n}_1+1}}, m) \end{cases} \tag{2.16}$$

The domination rule used will depend on whether $E_{\mathbf{x}}$ is less than $\alpha E_r(\frac{R}{\alpha}, P)$ or not. If

[8] Note that conditional probability, $\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\, \mathsf{M} = m]$, 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, $\frac{\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\mathsf{M}=\tilde{m}]}{\mathbf{P}[\mathsf{Y}^{\mathsf{n}}|\mathsf{M}=m]}$, for $m \neq \tilde{m}$ are in this family of decoding rules.

$E_{\mathbf{x}} \geq \alpha E_r(\frac{R}{\alpha}, P)$ the trivial domination rule leading to the trivial decoder $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$ is used. If $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ then the domination rule given in equation (2.17) is used.

$$(V, U, m) \succ (\hat{V}, U, \tilde{m}) \Leftrightarrow \begin{cases} \mathsf{I}(P, V) > \mathsf{I}\left(P, \hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}\left(U \| W_a | \Pi\right) \leq E_{\mathbf{x}} & \text{if } m \geq \tilde{m} \\ \mathsf{I}(P, V) \geq \mathsf{I}\left(P, \hat{V}\right) \text{ and } \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}\left(U \| W_a | \Pi\right) \leq E_{\mathbf{x}} & \text{if } m < \tilde{m} \end{cases} \tag{2.17}$$

where $\eta\left(R, P, V, \hat{V}\right)$ is given by equation (2.13).

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 (2.16) and (2.17) 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 (2.17), it is guaranteed to dominate the rest of the triplets too.

We allowed the decoding rule to depend on the index of the message in order to avoid declaring erasure whenever there is a tie in terms of the maximum empirical mutual information in the first phase. Alternatively, we could have employed a decoding rule which declares erasures whenever there are two or more messages with the largest empirical mutual information. The resulting tradeoff would be identical to the one we obtain below.

2.3.4 Error Analysis

Using an encoder like the one described in Section 2.3.2 and a decoder like the one in (2.16) we achieve the performance given below. If $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ then the domination rule given in equation (2.17) is used in the decoder; else a trivial domination rule that leads to a non-erasure decoder, $\hat{\mathsf{M}} = \tilde{\mathsf{M}}$, 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*

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

where $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ is given by,

$$E_{\mathbf{e}} = \left\{ \begin{array}{ll} \alpha E_r(\frac{R}{\alpha}, P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \min\limits_{(V,\hat{V},U)\in\mathcal{V}_{\mathbf{e}}} \alpha\eta\left(\frac{R}{\alpha}, P, \hat{V}, V\right) + (1-\alpha)\mathsf{D}\left(U \| W_r | \Pi\right) & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{array} \right\} \tag{2.18a}$$

$$\mathcal{V} = \{(V_1, V_2, U) : \mathsf{I}(P, V_1) \geq \mathsf{I}(P, V_2) \text{ and } (PV_1)_Y = (PV_2)_Y\} \tag{2.18b}$$

$$\mathcal{V}_{\mathbf{e}} = \left\{ \begin{array}{ll} \mathcal{V} & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \left\{(V, \hat{V}, U) \in \mathcal{V} : \ \alpha\eta\left(\frac{R}{\alpha}, P, V, \hat{V}\right) + (1-\alpha)\mathsf{D}\left(U \| W_a | \Pi\right) \leq E_{\mathbf{x}}\right\} & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{array} \right\} \tag{2.18c}$$

$$\delta_{\mathsf{n}}' = \frac{(|\mathcal{X}|+1)^2|\mathcal{Y}|\log(\mathsf{n}+1)}{\mathsf{n}} \tag{2.18d}$$

Equation (2.18) gives the whole achievable region for this family of codes. But for quantifying the gains of error-erasure decoding over the decoding schemes without erasures one need to consider only the region where $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P)$ holds, because for all $\alpha \in (0,1]$, $\alpha E_r(\frac{R}{\alpha}, P) \leq E_r(R, P)$.

The optimization problem given in (2.18) is a convex optimization problem: it is the 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 be calculated numerically 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 equation (2.36)). However the 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 2.3.5.

Proof:
A decoder of the form given in (2.16) decodes correctly when[9] $(\mathsf{Y}^{\mathsf{n}}, \tilde{\mathsf{M}}) \succ (\mathsf{Y}^{\mathsf{n}}, m),\ \forall m \neq \tilde{\mathsf{M}}$ and $\tilde{\mathsf{M}} = \mathsf{M}$. Thus an error or an erasure occur only when the correct message does not dominate all other messages, i.e. when $\exists m \neq \mathsf{M}$ such that $(\mathsf{Y}^{\mathsf{n}}, \mathsf{M}) \not\succ (\mathsf{Y}^{\mathsf{n}}, m)$. Consequently, we can write the sum of the conditional error probability and the conditional erasure probabilities for a message $m \in \mathcal{M}$ as,

$$P_{\mathbf{e}|m} + P_{\mathbf{x}|m} = \mathbf{P}[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.} (y^{\mathsf{n}}, m) \not\succ (y^{\mathsf{n}}, \tilde{m})\} | \, \mathsf{M} = m] \tag{2.19}$$

This can happen in two ways, either there is an error in the first phase, i.e. $\tilde{\mathsf{M}} \neq m$ or the 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 decoder using a domination rule described in Section 2.3.3,

$$\begin{aligned} P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq & \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \geq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \, m] \\ & + \sum_{V} \sum_{\hat{V}: \mathsf{I}(P,\hat{V}) \leq \mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1} \in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1} | \, m] \sum_{U:(V,U,m) \not\succ (\hat{V},U,m+1)} \sum_{y_{\mathsf{n}_1+1}^{\mathsf{n}} \in T_U} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}} \middle| \, \mathsf{X}_{\mathsf{n}_1+1}^{\mathsf{n}} = x_{\mathsf{n}_1+1}^{\mathsf{n}}(a)\right] \end{aligned}$$

where[10] $F^{(\mathsf{n}_1)}\left(V, \hat{V}, m\right)$ is the intersection of V-shell of message $m \in \mathcal{M}$ with the $\hat{V}$-shells of other messages, defined in equation (2.11). As result of Corollary 1,

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

Furthermore, as result of equation (2.15a),

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

---

[9] We use the short hand $(\mathsf{Y}^{\mathsf{n}}, \tilde{\mathsf{M}}) \succ (\mathsf{Y}^{\mathsf{n}}, m)$ for $(\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1} | \tilde{\mathsf{M}}}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, \mathsf{M}) \succ (\mathsf{V}_{\mathsf{Y}^{\mathsf{n}_1} | m}, \mathsf{U}_{\mathsf{Y}_{\mathsf{n}_1+1}^{\mathsf{n}}}, m)$ in this proof.

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

In addition the number of different non-empty V-shells in the communication phase is less than $(\mathsf{n}_1+1)^{|\mathcal{X}||\mathcal{Y}|}$ and the number of non-empty U-shells in the control phase is less than $(\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{X}|^2|\mathcal{Y}|}$. We denote the set of $(V,\hat{V},U)$ triples that corresponds to erasures with a correct tentative decision by $\mathcal{V}_{\mathbf{x}}$:

$$\mathcal{V}_{\mathbf{x}} \triangleq \left\{(V,\hat{V},U) : \mathsf{I}(P,V) \geq \mathsf{I}\left(P,\hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and } (V,U,m) \not\succ (\hat{V},U,m+1)\right\}. \tag{2.20}$$

In the above definition m is a dummy variable and $\mathcal{V}_{\mathbf{x}}$ is the same set for all $m \in \mathcal{M}$. Thus using (2.20) we get

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

Using the definition of $E_r(\frac{R}{\alpha},P)$ given in (2.4) we get

$$P_{\mathbf{e}|m} + P_{\mathbf{x}|m} \leq e^{\mathsf{n}\delta'_{\mathsf{n}}} \max\left\{e^{-\mathsf{n}\alpha E_r(R/\alpha,P)}, e^{-\mathsf{n}\min_{(V,\hat{V},U)\in\mathcal{V}_{\mathbf{x}}} \alpha\eta\left(R/\alpha,P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_a|\Pi)}\right\}. \tag{2.21}$$

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 $\tilde{\tilde{m}} \neq \tilde{m}$:

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

Note that when a $\tilde{m} \in \mathcal{M}$ dominates all other $\tilde{\tilde{m}} \neq \tilde{m}$, it also dominates m, i.e.

$$\left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}},\tilde{m}) \succ (y^{\mathsf{n}},\tilde{\tilde{m}}) \quad \forall \tilde{\tilde{m}} \neq \tilde{m}\right\} \subset \left\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}},\tilde{m}) \succ (y^{\mathsf{n}},m)\right\}.$$

Thus,

$$\begin{aligned}P_{\mathbf{e}|m} &\leq \mathbf{P}[\{y^{\mathsf{n}} : \exists \tilde{m} \neq m \text{ s.t.}(y^{\mathsf{n}},\tilde{m}) \succ (y^{\mathsf{n}},m)\}|\, \mathsf{M}=m] \\ &= \sum_{V} \sum_{\hat{V}:\mathsf{I}(P,\hat{V})\geq\mathsf{I}(P,V)} \sum_{y^{\mathsf{n}_1}\in F^{(\mathsf{n}_1)}(V,\hat{V},m)} \mathbf{P}[y^{\mathsf{n}_1}|\, \mathsf{M}=m] \sum_{U:(\hat{V},U,m-1)\succ(V,U,m)} \sum_{y^{\mathsf{n}}_{\mathsf{n}_1+1}\in T_U} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, x^{\mathsf{n}}_{\mathsf{n}_1+1}(r)\right].\end{aligned} \tag{2.22}$$

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 equal mutual information but smaller index. This is the reason why we sum over $(\hat{V},U,m-1) \succ (V,U,m)$. Using the inequality (2.15b) in the inner most two sums and then applying inequality (2.12) one gets,

$$\begin{aligned}P_{\mathbf{e}|m} &\leq (\mathsf{n}+1)^{(|\mathcal{X}|^2+2|\mathcal{X}|)|\mathcal{Y}|} \max_{\substack{(V,\hat{V},U):\\ \mathsf{I}(P,\hat{V})\geq\mathsf{I}(P,V)\\ (\hat{V},U,m-1)\succ(V,U,m)}} e^{-\mathsf{n}(\alpha\eta\left(R/\alpha,P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi))} \\ &\leq e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n}\min_{(\hat{V},V,U)\in\mathcal{V}_{\mathbf{e}}}(\alpha\eta\left(R/\alpha,P,V,\hat{V}\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi))} \\ &= e^{\mathsf{n}\delta'_{\mathsf{n}}} e^{-\mathsf{n}\min_{(V,\hat{V},U)\in\mathcal{V}_{\mathbf{e}}}(\alpha\eta\left(R/\alpha,P,\hat{V},V\right)+(1-\alpha)\mathsf{D}(U\|W_r|\Pi))}\end{aligned} \tag{2.23}$$

where $\mathcal{V}_{\mathbf{e}}$ is the complement of $\mathcal{V}_{\mathbf{x}}$ in $\mathcal{V}$ given by

$$\mathcal{V}_{\mathbf{e}} \triangleq \left\{(V,\hat{V},U) : \mathsf{I}\,(P,V) \geq \mathsf{I}\left(P,\hat{V}\right) \text{ and } (PV)_Y = (P\hat{V})_Y \text{ and } (V,U,m) \succ (\hat{V},U,m+1)\right\}. \tag{2.24}$$

Note that m in the definition of $\mathcal{V}_{\mathbf{e}}$ is also a dummy variable. 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 domination rule is equivalent to choosing the $\mathcal{V}_{\mathbf{e}}$. Depending on the value of $\alpha E_r(\frac{R}{\alpha},P)$ and $E_{\mathbf{x}}$, $\mathcal{V}_{\mathbf{e}}$ is chosen according to the rule given in (2.18c) then,

(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 equation (2.21).

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

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 (2.21) erasure probability is bounded as $P_{\mathbf{x}} \leq e^{-\mathsf{n}(E_{\mathbf{x}}-\delta_{\mathsf{n}}{}')}$ and Theorem 1 follows from equation (2.23).

QED

2.3.5 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 (2.18), $\forall R \geq 0$ and $\forall E_{\mathbf{x}} > E_r(R)$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \alpha E_r(\tfrac{R}{\alpha},P) \leq E_r(R) \quad \forall\alpha, \quad \forall P, \quad \forall\Pi \tag{2.25a}$$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\tilde{\alpha},\tilde{P},\Pi) = E_r(R) \quad \tilde{\alpha}=1, \quad \tilde{P} = \arg\max_P E_r(R,P), \quad \forall\Pi. \tag{2.25b}$$

Thus for all $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} > E_r(R)$ the optimal time sharing constant is 1, the optimal input distribution is the optimal input distribution for random coding exponent at rate R, and maximum mutual information decoding is used without ever declaring erasure.

$$E_{\mathbf{e}}(R,E_{\mathbf{x}}) = E_{\mathbf{e}}(R,E_{\mathbf{x}},1,P_{r(R)},\Pi) = E_r(R) \qquad \forall R \geq 0 \qquad \forall E_{\mathbf{x}} > E_r(R) \tag{2.26}$$

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

Low Erasure Exponent Region (i.e. $E_{\mathbf{x}} \leq E_r(R)$)

The benefits of error-erasure decoding are observed for $(R,E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$. Since $E_r(R)$ is a non-negative non-increasing and convex function of R, we have

$$\alpha \in [\alpha^*(R,E_{\mathbf{x}}),1] \Leftrightarrow E_{\mathbf{x}} \leq \alpha E_r(\tfrac{R}{\alpha}) \qquad \forall R \geq 0 \qquad \forall 0 < E_{\mathbf{x}} \leq E_r(R)$$

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 the resulting error exponent, $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$, does not correspond to the error exponent of a reliable sequence. The convention introduced below in equation (2.27) 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 $E_{\mathbf{x}} = 0$: $\lim_{E_{\mathbf{x}} \to 0} \alpha^*(R, E_{\mathbf{x}}) = \alpha^*(R, 0)$.

$$\alpha^*(R, E_{\mathbf{x}}) \triangleq \begin{cases} \frac{R}{g^{-1}(\frac{E_{\mathbf{x}}}{R})} & E_{\mathbf{x}} \in (0, E_r(R)] \\ \frac{R}{C} & E_{\mathbf{x}} = 0 \end{cases} \tag{2.27}$$

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

As a result equations (2.18) and (2.27), $\forall R \geq 0$ and $\forall 0 < E_{\mathbf{x}} \leq E_r(R)$ we have

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \alpha E_r(\tfrac{R}{\alpha},P) \quad \forall \alpha \in [0,\alpha^*(R, E_{\mathbf{x}})), \quad \forall P, \qquad \forall \Pi \tag{2.28a}$$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\tilde{\alpha},\tilde{P},\Pi) = E_r(R) \qquad \tilde{\alpha} = 1, \qquad \tilde{P} = \arg\max_P E_r(R, P), \quad \forall \Pi. \tag{2.28b}$$

Thus for all $(R, E_{\mathbf{x}})$ pairs such that $E_{\mathbf{x}} \leq E_r(R)$ the 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

$$\mathcal{P}(R, E_{\mathbf{x}}, \alpha) \triangleq \{P : \alpha E_r(\tfrac{R}{\alpha}, P) \geq E_{\mathbf{x}} \ , \ \mathsf{I}(P, W) \geq \tfrac{R}{\alpha}\}. \tag{2.29}$$

The constraint on mutual information is there to ensure that $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$ corresponds to an error exponent for 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 that $\forall R \geq 0$ and $\forall E_{\mathbf{x}} \in (0, E_r(R)]$,

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \alpha E_r(\tfrac{R}{\alpha}, P) \quad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1], \quad \forall P \notin \mathcal{P}(R, E_{\mathbf{x}}, \alpha), \qquad \forall \Pi \tag{2.30a}$$

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,\tilde{P},\Pi) \geq \alpha E_r(\tfrac{R}{\alpha}) \qquad \forall \alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1], \quad \tilde{P} = \arg\max_P E_r(\tfrac{R}{\alpha}, P), \quad \forall \Pi. \tag{2.30b}$$

Thus as a result of (2.30) one can restrict the optimization over P to $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ when $\forall R \geq 0$ and $\forall E_{\mathbf{x}} \in (0, E_r(R)]$. For $E_{\mathbf{x}} = 0$ case if we require the expression $E_{\mathbf{e}}(R, 0, \alpha, P, \Pi)$ to correspond to the error exponent of a reliable sequence, get the restriction given in equation (2.30). Thus using the definition of $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ given in (2.41) we get:

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \max_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) \qquad \forall R \geq 0 \qquad \forall E_{\mathbf{x}} \leq E_r(R) \tag{2.31}$$

where $\alpha^*(R, E_{\mathbf{x}})$, $\mathcal{P}(R, E_{\mathbf{x}}, \alpha)$ and $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ are given in equations (2.27), (2.29) and (2.18).

Unlike $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ itself, $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ as defined in (2.31) corresponds to error exponent of reliable code sequences even at $E_{\mathbf{x}} = 0$.

If maximizing P for the inner maximization in equation (2.31) is the same for all $\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]$, then the optimal value of α is $\alpha^*(R, E_{\mathbf{x}})$. In order to see that, we first observe that any fixed $(R, E_{\mathbf{x}}, P, \Pi)$ such that $E_r(R, P) \geq E_{\mathbf{x}}$, the function $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ 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[11] $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$, as is shown Lemma 17 in Appendix B.1. Since the maximization preserves the convexity, $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ 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, $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$, 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 B.2 $\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ takes its is 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 $E_{\mathbf{e}}(R, E_{\mathbf{x}})$ by simply maximizing over Π's. In symmetric channels, for example, the uniform distribution is the optimal distribution for all $(R, E_{\mathbf{x}})$ pairs. Thus

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{cc} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P^*, \Pi) & \text{if } E_{\mathbf{x}} > E_r(R, P^*) \\ \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha^*(R, E_{\mathbf{x}}), P^*, \Pi) & \text{if } E_{\mathbf{x}} \leq E_r(R, P^*) \end{array} \right\} \tag{2.32}$$

where P^* is the uniform distribution.

2.3.6 Alternative Expression for The Lower Bound

The minimization given in (2.18) for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ 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 (2.18) in terms of solutions of lower dimensional optimization problems.

Let $\zeta(R, P, Q)$ 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.1 that for a given P, $\zeta(R, P, Q)$ is convex in (R, Q) pair. Evidently for a given (P, Q) pair $\zeta(R, P, Q)$ is a non-increasing in R. Thus for a given (P, Q) pair $\zeta(R, P, Q)$ is strictly decreasing on a closed interval and is an extended real valued function of the form:

$$\zeta(R, P, Q) = \left\{ \begin{array}{cc} \infty & R < R_l^*(P, Q) \\ \min_{V: \substack{\mathsf{I}(P,V) \leq R \\ (PV)_Y = Q}} \mathsf{D}\left(V \| W | P\right) & R \in [R_l^*(P, Q), R_h^*(P, Q)] \\ \min_{V:(PV)_Y = Q} \mathsf{D}\left(V \| W | P\right) & R > R_h^*(P, Q) \end{array} \right\} \tag{2.33a}$$

$$R_l^*(P, Q) = \min_{V: \substack{PV \gg PW \\ (PV)_Y = Q}} \mathsf{I}\left(P, V\right) \tag{2.33b}$$

$$R_h^*(P, Q) = \min_R \left\{ R : \min_{V: \substack{\mathsf{I}(P,V) \leq R \\ (PV)_Y = Q}} \mathsf{D}\left(V \| W | P\right) = \min_{V:(PV)_Y = Q} \mathsf{D}\left(V \| W | P\right) \right\} \tag{2.33c}$$

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 $\Gamma\left(T, \Pi\right)$ 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.

$$\Gamma\left(T, \Pi\right) \triangleq \min_{U: \mathsf{D}(U \| W_a | \Pi) \leq T} \mathsf{D}\left(U \| W_r | \Pi\right) \tag{2.34}$$

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

For a given Π, $\Gamma(T,\Pi)$ is non-increasing and convex in T, thus $\Gamma(T,\Pi)$ is strictly decreasing in T on a closed interval. An equivalent expressions for $\Gamma(T,\Pi)$ and boundaries of this closed interval is derived in Appendix A,

$$\Gamma(T,\Pi) = \left\{ \begin{array}{ccl} \infty & \text{if} & T < \mathsf{D}(U_0\|W_a|\Pi) \\ \mathsf{D}(U_s\|W_r|\Pi) & \text{if} & T = \mathsf{D}(U_s\|W_a|\Pi) \quad \text{for some } s\in[0,1] \\ \mathsf{D}(U_1\|W_r|\Pi) & \text{if} & T > \mathsf{D}(U_1\|W_a|\Pi) \end{array} \right\} \tag{2.35}$$

where

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

For a $(R,E_{\mathbf{x}},\alpha,P,\Pi)$ such that $E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha},P)$, using the definition of $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ in (2.18) together with the equations (2.13), (2.33) and (2.35) we get

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \min_{\substack{Q,T,R_1,R_2: \\ R_1\geq R_2\geq 0,\ T\geq 0 \\ \alpha\zeta(\frac{R_1}{\alpha},P,Q)+|R_2-R|^+ +T\leq E_{\mathbf{x}}}} \alpha\zeta(\tfrac{R_2}{\alpha},P,Q)+|R_1-R|^+ +(1-\alpha)\Gamma\left(\tfrac{T}{1-\alpha},\Pi\right)$$

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

(a) $R_1^*\geq R_2^*\geq R$ claim holds trivially.

(b) $R_1^*\geq R > R_2^*$, since $\zeta(\frac{R_2}{\alpha},P,Q)$ is non-increasing function (Q^*,R_1^*,R,T^*), is also minimizing, thus claim holds.

(c) $R > R_1^* > R_2^*$, since $\zeta(\frac{R}{\alpha},P,Q)$ is non-increasing function (Q^*,R,R,T^*), is also minimizing, thus claim holds.

Thus we obtain the following expression for $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$,

$$E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi) = \left\{ \begin{array}{ll} \alpha E_r(\frac{R}{\alpha},P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha},P) \\ \min\limits_{\substack{Q,T,R_1,R_2: \\ R_1\geq R_2\geq R,\ T\geq 0 \\ \alpha\zeta(\frac{R_1}{\alpha},P,Q)+R_2-R+T\leq E_{\mathbf{x}}}} \alpha\zeta(\frac{R_2}{\alpha},P,Q)+R_1-R+(1-\alpha)\Gamma\left(\frac{T}{1-\alpha},\Pi\right) & \text{if } E_{\mathbf{x}}\leq \alpha E_r(\frac{R}{\alpha},P) \end{array} \right\} \tag{2.36}$$

Equation (2.36) is simplified further for symmetric channels. For symmetric channels,

$$E_{sp}(R) = \zeta(R,P^*,Q^*) = \min_{Q} \zeta(R,P^*,Q) \tag{2.37}$$

where P^* is the uniform input distribution and Q^* is the corresponding output distribution under W.

Using alternative expression for $E_{\mathbf{e}}(R,E_{\mathbf{x}},\alpha,P,\Pi)$ given in (2.36) together with equations (2.32) and (2.37) for symmetric channels we get,

$$E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{ll} E_r(R) & \text{if } E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\Pi} \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R \;\; T \geq 0 \\ \alpha^* E_{sp}(\frac{R''}{\alpha^*}) + R' - R + T \leq E_{\mathbf{x}}}} \alpha^* E_{sp}\left(\frac{R'}{\alpha^*}\right) + R'' - R + (1-\alpha^*)\Gamma\left(\frac{T}{1-\alpha^*}, \Pi\right) & \text{if } E_{\mathbf{x}} \leq E_r(R) \end{array} \right\} \tag{2.38}$$

where $\alpha^*(R, E_{\mathbf{x}})$ is given in equation (2.27).

Although (2.37) does not hold in general using definition of $\zeta(R, P, Q)$ and $E_{sp}(R, P)$ one can assert that

$$\zeta(R, P, Q) \geq \min_{\tilde{Q}} \zeta(R, P, \tilde{Q}) = E_{sp}(R, P). \tag{2.39}$$

Note that (2.39) can be used to bound the minimized expression in (2.36) from below. In addition recall that if the set that a minimization is done over is enlarged resulting minimum can not increase. One can use (2.36) also to enlarge the set that minimization is done over in (2.39). Thus we get an exponent $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ which is smaller than or equal to $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in all channels and for all $\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$'s:

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) = \left\{ \begin{array}{ll} \alpha E_r(\frac{R}{\alpha}, P) & \text{if } E_{\mathbf{x}} > \alpha E_r(\frac{R}{\alpha}, P) \\ \min\limits_{\substack{R'', R', T: \\ R'' \geq R' \geq R \;\; T \geq 0 \\ \alpha E_{sp}(\frac{R''}{\alpha}, P) + R' - R + T \leq E_{\mathbf{x}}}} \alpha E_{sp}\left(\frac{R'}{\alpha}, P\right) + R'' - R + (1-\alpha)\Gamma\left(\frac{T}{1-\alpha}, \Pi\right) & \text{if } E_{\mathbf{x}} \leq \alpha E_r(\frac{R}{\alpha}, P) \end{array} \right\} \tag{2.40}$$

After an investigation very similar to the one that has been done for $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in Section 2.3.5, one obtains the below expression for the optimal error exponent for reliable sequences emerging from (2.40):

$$\tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) = \left\{ \begin{array}{ll} E_r(R) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} > E_r(R) \\ \max\limits_{\alpha \in [\alpha^*(R, E_{\mathbf{x}}), 1]} \max\limits_{P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)} \max\limits_{\Pi} \tilde{E}_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) & \forall R \geq 0 \quad \forall E_{\mathbf{x}} \leq E_r(R) \end{array} \right\} \tag{2.41}$$

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 equations (2.27), (2.29) and (2.40), respectively.

2.3.7 Special Cases

Zero Erasure Exponent Case, $\mathcal{E}_{\mathbf{e}}(R, 0)$

Using a simple repetition-at-erasures scheme, fixed-length block codes with errors and erasures decoding, 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 in [4] is an upper bound to the error exponent of fixed-length block codes with erasures:

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

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 for $E_{\mathbf{x}} = 0$ for all rates $\tilde{E}_{\mathbf{e}}(R, 0) = \mathcal{E}_{\mathbf{e}}(R, 0) = (1 - \frac{R}{\mathcal{C}})\mathcal{D}$. In other words the two phase encoding scheme discussed above is optimal for $E_{\mathbf{x}} = 0$.

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

$$\mathcal{P}(R, 0, \alpha) = \{P : \mathsf{I}(P, W) \geq \tfrac{R}{\alpha}\}$$

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

$$\tilde{E}_{\mathbf{e}}(R, 0, \alpha, P, \Pi) = \alpha \left[E_{sp}\left(\tfrac{R}{\alpha}, P\right) + \mathsf{I}(P, W) - \tfrac{R}{\alpha} \right] + (1 - \alpha) \mathsf{D}\left(W_r \| W_a | \Pi\right) \tag{2.42}$$

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

$$\tilde{E}_{\mathbf{e}}(R, 0, \alpha) = \max_{P \in \mathcal{P}(R,0,\alpha)} \alpha E_{sp}\left(\tfrac{R}{\alpha}, P\right) + \alpha \mathsf{I}(P, W) - R + (1 - \alpha)\mathcal{D} \qquad \forall \alpha \in [\tfrac{R}{\mathcal{C}}, 1]. \tag{2.43}$$

The value of $\tilde{E}_{\mathbf{e}}(R, 0, \alpha)$ at any particular value of α is a lower bound on $\tilde{E}_{\mathbf{e}}(R, 0)$. Thus,

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

Indeed one need not rely on the converse for 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 probability of 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 non-zero Zero Error Capacity

For channels with a non-zero zero-error capacity, one can use equation (2.18) to prove that, for any $E_{\mathbf{x}} < E_r(R)$, $E_{\mathbf{e}}(R, E_{\mathbf{x}}) = \infty$. This implies that we can get error-free block codes with this two phase coding scheme for any rate $R < \mathcal{C}$ and any erasure exponent $E_{\mathbf{x}} \leq E_r(R)$. As we discuss in Section 2.5 in more detail, this is the best erasure exponent for rates over the critical rate.

2.4 An Outer Bound for Error Exponent Erasure Exponent Tradeoff

In this section we derive an upper bound on $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$ using previously known results on erasure free block codes with feedback and a generalization of the straight line bound of Shannon, Gallager and Berlekamp [38]. 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 2.4.1. Then in Section 2.4.2 we give a brief overview of the outer bounds on the error exponents of erasure free block codes with feedback. Finally in Section 2.4.3, we use the relation derived in Section 2.4.1 to tie the previously known result summarized in Section 2.4.2 to bound $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$.

2.4.1 A Property of Minimum Error Probability for Block Codes with Erasures

Shannon, Gallager and Berlekamp [38] considered fixed-length block codes, with 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, [38, Theorem 1], that for fixed-length block codes with list decoding and without feedback

$$\tilde{\mathcal{P}}_{\mathbf{e}}(M,\mathsf{n},L) \geq \tilde{\mathcal{P}}_{\mathbf{e}}(M,\mathsf{n}_1,L_1)\tilde{\mathcal{P}}_{\mathbf{e}}(L_1+1,\mathsf{n}-\mathsf{n}_1,L) \tag{2.44}$$

where $\tilde{\mathcal{P}}_{\mathbf{e}}(M,\mathsf{n},L)$ 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 pointed out in [38], this theorem continues to hold in the case when a feedback link is available from receiver to the transmitter; although $\tilde{\mathcal{P}}_{\mathbf{e}}$ is different when feedback is available, the relation given in equation (2.44) still holds. They were interested in erasure free codes. We, on the other hand, are interested in block codes which might have non-zero erasure probability. Accordingly we will incorporate erasure probability as one of the parameters of the optimal error probability.

The decoded set $\hat{\mathsf{M}}$ of a size L list decoder with erasures is either a subset[12] of $\mathcal{M}$ whose size is at most L, like the erasure-free case, or a set which only includes the erasure symbol,[13] i.e. either $\hat{\mathsf{M}} \subset \mathcal{M}$ such that $|\hat{\mathsf{M}}| \leq L$ or $\hat{\mathsf{M}} = \{\mathbf{x}\}$. The minimum error probability of length n block codes, with M equally probable messages, decoding list size L and erasure probability $P_{\mathbf{x}}$ is denoted by $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},L,P_{\mathbf{x}})$.

Theorem 2 below bounds the error probability of block codes with erasures and list decoding using the error probabilities of shorter codes with erasures and list decoding, like [38, Theorem 1] does in the erasure free case. Like its counter part in the erasure free case Theorem 2 is later used to establish outer bounds to error exponents.

Theorem 2 *For any 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 satisfy*

$$\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},L,P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,s)\mathcal{P}_{\mathbf{e}}\left(L_1+1,\mathsf{n}-\mathsf{n}_1,L,\tfrac{(1-s)P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,s)}\right) \tag{2.45}$$

Note that given a (M,n,L) triple if the error probability erasure probability pairs $(P_{\mathbf{e}a},P_{\mathbf{x}a})$ and $(P_{\mathbf{e}b},P_{\mathbf{x}b})$ are achievable then for any $\gamma \in [0,1]$ we can use the code achieving $(P_{\mathbf{e}a},P_{\mathbf{x}a})$ with probability γ the code achieving $(P_{\mathbf{e}b},P_{\mathbf{x}b})$ with probability $(1-\gamma)$ and achieve error probability erasure probability pair $(\gamma P_{\mathbf{e}a}+(1-\gamma)P_{\mathbf{e}b},\gamma P_{\mathbf{x}a}+(1-\gamma)P_{\mathbf{x}b})$. As a result for any (M,n,L) triple the set set of achievable error probability erasure probability pairs is convex. We will 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*

$$P_{\mathbf{e}} \geq \Omega(\varphi,K)\,\mathcal{P}_{\mathbf{e}}\left(K+1,\mathsf{n},L,\tfrac{P_{\mathbf{x}}}{\Omega(\varphi,K)}\right) \quad \text{where} \quad \Omega(\varphi,K) = \min_{\mathcal{S}\subset\mathcal{M}:|\mathcal{S}|=|\mathcal{M}|-K} \varphi(\mathcal{S}). \tag{2.46}$$

where $\mathcal{P}_{\mathbf{e}}(K+1,\mathsf{n},L,P_{\mathbf{x}})$ 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 has feedback and without feedback if the original code has not.

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

[13]Note that $\emptyset \subset \mathcal{M}$ thus in our convention decoding to an empty list is an error rather than an erasure.

Note that $\Omega(\varphi, K)$ is the error probability of a decoder which decodes to K messages with the largest probability i.e. 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 theorem 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.

$$\Phi'_t(m, y^{t-1}) = \Phi_t(m, y^{t-1}) \qquad \forall m \in \mathcal{M}', \quad t \in [1, \mathsf{n}], \quad y^{t-1} \in \mathcal{Y}^{t-1}$$

- **Decoder:** $\forall y^{\mathsf{n}} \in \mathcal{Y}^{\mathsf{n}}$ if the original decoding rule declares erasure, declare erasure, else the 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 $P'_{\mathbf{e}|m}$ and the conditional erasure probability $P'_{\mathbf{x}|m}$ are equal to the conditional error probability $P_{\mathbf{e}|m}$ and the conditional erasure probability $P_{\mathbf{x}|m}$ in the original code, respectively. Thus

$$\tfrac{1}{K+1} \sum\nolimits_{m \in \mathcal{M}'} \left(P_{\mathbf{e}|m}, P_{\mathbf{x}|m}\right) \in \Psi(K+1, \mathsf{n}, L) \qquad \forall \mathcal{M}' \subset \mathcal{M} \text{ such that } |\mathcal{M}'| = K+1 \quad (2.47)$$

where $\Psi(K+1, \mathsf{n}, L)$ is the set of 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 non-zero element of $\{\varphi(1), \varphi(2), \dots \varphi(|\mathcal{M}|)\}$ be $\varphi(\xi_1)$. For any size $(K+1)$ subset of $\mathcal{M}$ which includes ξ_1 and all whose elements have non-zero probabilities, say $\mathcal{M}_1$, we have,

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

Equation (2.47) and the definition of $\Psi(K+1, \mathsf{n}, L)$, implies that $\exists \psi_1 \in \Psi(K+1, \mathsf{n}, L)$ such that

$$(P_{\mathbf{e}}, P_{\mathbf{x}}) = \sum\nolimits_{m \in \mathcal{M}} \varphi^{(1)}(m)(P_{\mathbf{e}|m}, P_{\mathbf{x}|m}) + \varphi(\psi_1)\psi_1 \quad (2.48)$$

$$1 = \varphi(\psi_1) + \sum\nolimits_{m \in \mathcal{M}} \varphi^{(1)}(m) \quad (2.49)$$

where $\varphi(\psi_1) = (K+1)\varphi(\xi_1)$ and $\varphi^{(1)}(m) = \varphi(m) - \varphi(\xi_1)\mathbb{1}_{\{m \in \mathcal{M}_1\}}$. Furthermore the number of non-zero $\varphi^{(1)}(m)$'s is at least one less than that of non-zero $\varphi(m)$'s. The remaining probabilities, $\varphi^{(1)}(m)$, have a minimum, $\varphi^{(1)}(\xi_2)$ among its non-zero elements. One can repeat the same argument once more using that element and reduce the number of non-zero elements at least one more. After at most $|\mathcal{M}| - K$ such iterations one reaches to a $\varphi^{(\ell)}$ which is non-zero for K or fewer messages:

$$(P_{\mathbf{e}}, P_{\mathbf{x}}) = \sum\nolimits_{j=1}^{\ell} \varphi(\psi_j)\psi_j + \sum\nolimits_{m\in\mathcal{M}} \varphi^{(\ell)}(m)(P_{\mathbf{e}|m}, P_{\mathbf{x}|m}) \tag{2.50}$$

where $\varphi^{(\ell)}(m) \leq \varphi(m)$ for all m in $\mathcal{M}$ and $\sum_{m\in\mathcal{M}} \mathbb{1}_{\left\{\varphi^{(\ell)}(m)>0\right\}} \leq K$.

In equation (2.50), 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 non-negative entries. As a result of definition of $\Omega(\varphi, K)$ given in equation (2.46),

$$\Omega(\varphi, K) \leq \sum\nolimits_{j=1}^{\ell} \varphi^{(j-1)}(\psi_j). \tag{2.51}$$

Then as a result of convexity of $\Psi(K+1, \mathsf{n}, L)$ for some $a \geq 1$, $b_1 \geq 0$ and $b_2 \geq 0$ we have,

$$\exists \tilde{\psi} \in \Psi(K+1, \mathsf{n}, L) \text{ such that } (P_{\mathbf{e}}, P_{\mathbf{x}}) = a\Omega(\varphi, K)\,\tilde{\psi} + (b_1, b_2) \tag{2.52}$$

Thus from equation (2.52) and the definition of $\Psi(K+1, \mathsf{n}, L)$ we have

$$\exists \psi \in \Psi(K+1, \mathsf{n}, L) \text{ such that } \left(\tfrac{P_{\mathbf{e}}}{\Omega(\varphi,K)}, \tfrac{P_{\mathbf{x}}}{\Omega(\varphi,K)}\right) = \psi \tag{2.53}$$

Then the lemma follows from equation (2.53) and the fact that $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, s_{\mathbf{x}})$ is uniquely determined by $\Psi(M, \mathsf{n}, L)$ for $s_{\mathbf{x}} \in [0, 1]$ as follows

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, \psi_{\mathbf{x}}) = \min_{\psi_{\mathbf{x}}:(\psi_{\mathbf{e}},\psi_{\mathbf{x}})\in\Psi(M,\mathsf{n},L)} \psi_{\mathbf{e}} \qquad \forall (M, \mathsf{n}, L, \psi_{\mathbf{x}}). \tag{2.54}$$

QED

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 $\varphi_{y^{\mathsf{n}_1}}(m) = \mathbf{P}[m|\, y^{\mathsf{n}_1}]$ over the messages and apply Lemma 2 together with convexity arguments similar to the ones above.

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

$$\mathcal{G}(m) \triangleq \{y^{\mathsf{n}} : m \in \hat{\mathsf{M}}\} \qquad \mathcal{G}(\mathbf{x}) \triangleq \{y^{\mathsf{n}} : \mathbf{x} \in \hat{\mathsf{M}}\} \qquad \tilde{\mathcal{G}}(m) \triangleq \mathcal{G}(m)^c \cap \mathcal{G}(\mathbf{x})^c. \tag{2.55}$$

Then for all $m \in \mathcal{M}$,

$$(P_{\mathbf{e}|m}, P_{\mathbf{x}|m}) = \left(\mathbf{P}\left[\tilde{\mathcal{G}}(m)\middle|\, m\right], \mathbf{P}[\mathcal{G}(\mathbf{x})|\, m]\right). \tag{2.56}$$

Note that

$$\begin{aligned} P_{\mathbf{x}|m} &= \sum\nolimits_{y^{\mathsf{n}}:y^{\mathsf{n}}\in\mathcal{G}(\mathbf{x})} \mathbf{P}[y^{\mathsf{n}}|\, m] \\ &= \sum\nolimits_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}|\, m] \sum\nolimits_{y_{\mathsf{n}_1+1}^{\mathsf{n}}:(y^{\mathsf{n}_1}, y_{\mathsf{n}_1+1}^{\mathsf{n}})\in\mathcal{G}(\mathbf{x})} \mathbf{P}\left[y_{\mathsf{n}_1+1}^{\mathsf{n}}\middle|\, m, y^{\mathsf{n}_1}\right]. \end{aligned}$$

Then the erasure probability is

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

Note that for every y^{n_1}, $P_{\mathbf{x}}(y^{\mathsf{n}_1})$ is the erasure probability of a code of length $(\mathsf{n}-\mathsf{n}_1)$ with a priori probability distribution is $\varphi_{y^{\mathsf{n}_1}}(m) = \mathbf{P}[m|\, y^{\mathsf{n}_1}]$. Furthermore one can write the error probability, $P_{\mathbf{e}}$ as

$$\begin{aligned} P_{\mathbf{e}} &= \sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}] \left(\sum_{m\in\mathcal{M}} \mathbf{P}[m|\, y^{\mathsf{n}_1}] \sum_{y^{\mathsf{n}}_{\mathsf{n}_1+1}:(y^{\mathsf{n}_1},y^{\mathsf{n}}_{\mathsf{n}_1+1})\in\tilde{\mathcal{G}}(m)} \mathbf{P}\left[y^{\mathsf{n}}_{\mathsf{n}_1+1}\middle|\, m, y^{\mathsf{n}_1}\right]\right) \\ &= \sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\, P_{\mathbf{e}}(y^{\mathsf{n}_1}) \end{aligned}$$

where $P_{\mathbf{e}}(y^{\mathsf{n}_1})$ is the error probability of the very same length $(\mathsf{n}-\mathsf{n}_1)$ code. As a result of Lemma 2, the pair $(P_{\mathbf{e}}(y^{\mathsf{n}_1}), P_{\mathbf{x}}(y^{\mathsf{n}_1}))$ satisfies

$$P_{\mathbf{e}}(y^{\mathsf{n}_1}) \geq \Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right). \qquad (2.57)$$

Then for any $s \in [0,1]$.

$$\begin{aligned} P_{\mathbf{e}} &\geq \sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \\ &\geq \left(\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)P_{\mathbf{x}}(y^{\mathsf{n}_1})}{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \\ &= \left(\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)\right) \mathcal{P}_{\mathbf{e}}\left(L_1+1, (\mathsf{n}-\mathsf{n}_1), L, \tfrac{(1-s)P_{\mathbf{x}}}{\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}](1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right)}\right) \end{aligned} \qquad (2.58)$$

where the second inequality follows from the convexity of $\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L, P_{\mathbf{x}})$ in $P_{\mathbf{x}}$.
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 $\mathsf{Y}^{\mathsf{n}_1}$ (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 the below expression (2.59) is its error probability. But that probability is lower bounded by $\mathcal{P}_{\mathbf{e}}\left(M, \mathsf{n}_1, L_1, s\right)$ which is minimum error probability over all length n_1 block codes with M messages and list size L_1, i.e.

$$\sum_{y^{\mathsf{n}_1}} \mathbf{P}[y^{\mathsf{n}_1}]\,(1-s)\Omega\left(\varphi_{y^{\mathsf{n}_1}}, L_1\right) \geq \mathcal{P}_{\mathbf{e}}\left(M, \mathsf{n}_1, L_1, s\right). \qquad (2.59)$$

Then the theorem follows from the fact that $P_{\mathbf{e}}(M, \mathsf{n}, L_1, P_{\mathbf{x}})$ is decreasing function of $P_{\mathbf{x}}$ and the equations (2.58) and (2.59).

QED

Like the result of Shannon, Gallager and Berlekamp in [38, Theorem 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 equation (2.45) holds in both cases.

2.4.2 Outer Bounds for 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 result are used in Section 2.4.3 together with Theorem 2 to bound $\mathcal{E}_{\mathsf{e}}(R, E_{\mathsf{x}})$ 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 combine various bounds on the error probability.

First bound we consider is on the error exponent of erasure free block codes with feedback. Haroutunian proved in [20] that, for any $(M_{\mathsf{n}}, \mathsf{n}, L_{\mathsf{n}})$ sequence of triples, such that $\lim_{\mathsf{n}\to\infty} \frac{\ln M_{\mathsf{n}} - \ln L_{\mathsf{n}}}{\mathsf{n}} = R$,

$$\lim_{\mathsf{n}\to\infty} \frac{-\ln \mathcal{P}_{\mathsf{e}}(M_{\mathsf{n}},\mathsf{n},L_{\mathsf{n}},0)}{\mathsf{n}} \leq E_H(R) \tag{2.60}$$

where

$$E_H(R) = \min_{V:\mathcal{C}(V)\leq R} \max_{P} \mathsf{D}\left(V \| W|P\right) \qquad \text{and} \qquad \mathcal{C}(V) = \max_{P} \mathsf{I}\left(P, V\right). \tag{2.61}$$

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 while passing in [1] and attributes it to Gallager and Shannon.

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

$$\textit{either} \qquad P_{\mathbf{e}1} \geq \tfrac{1}{4} e^{-\mathsf{n}T + \sqrt{\mathsf{n}} 4 \ln P_{min}} \qquad \textit{or} \qquad P_{\mathbf{e}2} > \tfrac{1}{4} e^{-\mathsf{n}\Gamma(T) + \sqrt{\mathsf{n}} 4 \ln P_{min}}. \tag{2.62}$$

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

$$T_0 \triangleq \max_{x,\tilde{x}} - \ln \sum\nolimits_{y:W(y|\tilde{x})>0} W(y|x) \tag{2.63}$$

$$\Gamma\left(T\right) \triangleq \max_{\Pi} \Gamma\left(T, \Pi\right). \tag{2.64}$$

Result is old and somewhat intuitive to those who are familiar with the calculations in the non-feedback case; thus probably it 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, $(T, \Gamma\left(T\right))$ is indeed the optimal tradeoff for the error exponent of two messages in an erasure free block code, both with and without feedback. Achievability of this tradeoff has already been established in [38, Theorem 5] for the case without feedback; evidently this implies the achievability 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

$$T^* \triangleq \max_{T} \min\{T, \Gamma\left(T\right)\}.$$

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,

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

Even in the cases where $E_H(0)$ is finite, $E_H(0) \geq T^*$. Following lemma uses this fact, Lemma 3, and Theorem 2 to strengthen Haroutunian bound at low rates.

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

$$\lim_{\mathsf{n}\to\infty} \frac{-\ln \mathcal{P}_{\mathbf{e}}(M_{\mathsf{n}},\mathsf{n},1,0)}{\mathsf{n}} \leq \tilde{E}_H(R) \tag{2.65}$$

where

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

and R_{ht} is the unique solution of the equation $T^* = E_H(R) - RE'_H(R)$ if it exists $\mathcal{C}$ 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 same as the tangent line, for the rates between 0 and R_{ht}, and it is 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 in [20] for $L_1 = 1$. If $R < R_{ht}$ then we apply Theorem 2.

$$\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, L_1, P_{\mathbf{x}}) \geq \mathcal{P}_{\mathbf{e}}(M, \tilde{\mathsf{n}}, L_1, s)\mathcal{P}_{\mathbf{e}}\left(L_1 + 1, \mathsf{n} - \tilde{\mathsf{n}}, \tilde{L}, \frac{(1-s)P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n},L_1,s)}\right) \tag{2.66}$$

with[14] $s = 0$, $P_{\mathbf{x}} = 0$, $L_1 = 1$ and $\tilde{\mathsf{n}} = \lfloor \frac{R}{R_{ht}} \mathsf{n} \rfloor$. On the other hand as a result of Lemma 3 and definition of T^*,

$$\mathcal{P}_{\mathbf{e}}(2, \mathsf{n} - \tilde{\mathsf{n}}, L, 0) \geq \frac{e^{-(\mathsf{n}-\tilde{\mathsf{n}})T^* + \sqrt{\mathsf{n}-\tilde{\mathsf{n}}} \ln P_{min}}}{8}. \tag{2.67}$$

Using equations (2.66) and (2.67) one gets,

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

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

QED

2.4.3 Generalized Straight Line Bound and Upper Bounds to $\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}})$

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 constraint the rates and erasure probabilities

[14] Or [38, Theorem 1] with $L_1 = 1$ and $\mathsf{n}_1 = \lfloor \frac{R}{R_{ht}} \mathsf{n} \rfloor$.

of the shorter ones, but do 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 $E_{\mathbf{x}} \in [0, \tilde{E}_H(R)]$ and for any $r \in [r_h(R, E_{\mathbf{x}}), \mathcal{C}]$*

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

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 $(R, E_{\mathbf{x}}, E_{\mathbf{e}}) = (0, E_{\mathbf{x}}, \Gamma(E_{\mathbf{x}}))$ lays 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 it is shown in Section 2.5 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.

- $r \in (r_h(R, E_{\mathbf{x}}), \mathcal{C}]$: Apply Theorem 2, for $s = 0$, $L = 1$, $L_1 = 1$, take the logarithm of both sides of equation (2.45) and divide by n,

$$\frac{-\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n},1,P_{\mathbf{x}})}{\mathsf{n}} \leq \left(\frac{\mathsf{n}_1}{\mathsf{n}}\right)\frac{-\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}{\mathsf{n}_1} + \left(1 - \frac{\mathsf{n}_1}{\mathsf{n}}\right)\frac{-\ln \mathcal{P}_{\mathbf{e}}\left(2,\mathsf{n}-\mathsf{n}_1,1,\frac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}\right)}{\mathsf{n}-\mathsf{n}_1}. \quad (2.68)$$

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 $\mathsf{n}_1 = \lfloor \frac{R}{r}\mathsf{n} \rfloor$ since $r > r_h(R, E_{\mathbf{x}})$ we have,

$$\begin{aligned}\liminf_{\mathsf{n}\to\infty} \tfrac{-1}{\mathsf{n}-\mathsf{n}_1} \ln \tfrac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)} &= \tfrac{1}{1-R/r}(E_{\mathbf{x}} + \tfrac{R}{r}\liminf_{\mathsf{n}_1\to\infty} \tfrac{\ln \mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)}{\mathsf{n}_1}) \\ &\geq \tfrac{1}{1-R/r}(E_{\mathbf{x}} - \tilde{E}_H(r) \\ &> 0.\end{aligned}$$

where the last step follows from our assumption that $r > r_h(R, E_{\mathbf{x}})$ and the fact that $\tilde{E}_H(r)$ is strictly decreasing.

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

$$\liminf_{\mathsf{n}\to\infty} \frac{-\ln \mathcal{P}_{\mathbf{e}}(2,\mathsf{n},1,{P_{\mathbf{x}}}^{(\mathsf{n})})}{\mathsf{n}} \leq \Gamma(T). \quad (2.69)$$

Using equation (2.68), taking the limit as n goes to infinity and using Theorem 3 we get

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

as long as $\frac{rE_{\mathbf{x}} - R\tilde{E}_H(r)}{r-R} < T^*$. But this constraint holds for all $r \in (r_h(R, E_{\mathbf{x}}), \mathcal{C}]$ because of the convexity of $\tilde{E}_H\cdot$ and the fact that $\tilde{E}_H(0) = T^*$.

In order to establish equation (2.69); note that if $T_0 > 0$ and $T \leq T_0$ then $\Gamma(T) = \infty$. Thus equation (2.69) holds trivially when $t \leq T_0$. For the case $T > T_0$ we prove equation (2.69) by contradiction. Assume that (2.69) is wrong. Then there exists a block code with erasures that satisfies

$$\mathbf{P}\left[\tilde{\mathcal{G}}(\tilde{\tilde{m}})\middle|\, \tilde{m}\right] \leq e^{-\mathsf{n}(\Gamma(T)+\mathsf{o}(1))} \qquad \mathbf{P}[\mathcal{G}(\mathbf{x})|\, \tilde{m}] \leq e^{-\mathsf{n}(T+\mathsf{o}(1))}$$
$$\mathbf{P}\left[\tilde{\mathcal{G}}(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] \leq e^{-\mathsf{n}(\Gamma(T)+\mathsf{o}(1))} \qquad \mathbf{P}\left[\mathcal{G}(\mathbf{x})|\, \tilde{\tilde{m}}\right] \leq e^{-\mathsf{n}(T+\mathsf{o}(1))}$$

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

$$\mathcal{G}'(\tilde{m}) = \mathcal{G}(\tilde{m}) \cup \mathcal{G}(\mathbf{x}) \qquad \mathcal{G}'(\tilde{\tilde{m}}) = \mathcal{G}(\tilde{\tilde{m}}) \qquad \mathcal{G}'(\mathbf{x}) = \emptyset.$$

The resulting code is an erasure free code with

$$\mathbf{P}\left[\mathcal{G}'(\tilde{\tilde{m}})\middle|\, \tilde{m}\right] \leq e^{-\mathsf{n}(\Gamma(T)+\mathsf{o}(1))} \qquad \text{and} \qquad \mathbf{P}\left[\mathcal{G}'(\tilde{m})\middle|\, \tilde{\tilde{m}}\right] \leq e^{-\mathsf{n}(\min\{\Gamma(T),T\}+\mathsf{o}(1))}$$

Since $T_0 < T \leq T^*$, $\Gamma(T) \geq T$, this contradicts with Lemma 3 thus equation (2.69) holds.

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

$$\begin{aligned}\mathcal{P}_{\mathbf{e}}(M, \mathsf{n}, 1, P_{\mathbf{x}}) &\geq \mathcal{P}_{\mathbf{e}}(M, \mathsf{n}_1, 1, 0)\mathcal{P}_{\mathbf{e}}(2, \mathsf{n}-\mathsf{n}_1, 1, \tfrac{P_{\mathbf{x}}}{\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,1,0)})\\ &\geq P_{\mathbf{x}} \ln \tfrac{1}{P_{\mathbf{x}}} \mathcal{P}_{\mathbf{e}}(2, \mathsf{n}-\mathsf{n}_1, 1, \tfrac{1}{-\ln P_{\mathbf{x}}})\end{aligned} \tag{2.70}$$

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

$$\liminf_{\mathsf{n}\to\infty} \tfrac{\mathsf{n}_1}{\mathsf{n}} \mathcal{E}(\tfrac{R\mathsf{n}}{\mathsf{n}_1}) = E_{\mathbf{x}}$$

Then as a result of Lemma 4 we have,

$$\liminf_{\mathsf{n}\to\infty} \tfrac{\mathsf{n}_1}{\mathsf{n}} \tilde{E}_H(\tfrac{R\mathsf{n}}{\mathsf{n}_1}) \geq E_{\mathbf{x}}$$

Then

$$\liminf_{\mathsf{n}\to\infty} \tfrac{\mathsf{n}_1}{\mathsf{n}} \geq \tfrac{R}{r_h(R,E_{\mathbf{x}})} \tag{2.71}$$

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

$$\liminf_{\mathsf{n}\to\infty} \tfrac{-\ln \mathcal{P}_{\mathbf{e}}(2,\mathsf{n},1,\epsilon_{\mathsf{n}})}{\mathsf{n}} \leq \Gamma(0) \tag{2.72}$$

Then taking the logarithm of both sides of the equation (2.70), dividing both sides by n, taking the limit as n tends to infinity and substituting equations (2.71) and (2.72) we get,

$$\mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) \leq E_{\mathbf{x}} + (1 - \tfrac{E_{\mathbf{x}}}{\tilde{E}_H(r_h(R,E_{\mathbf{x}}))})\Gamma(0) \tag{2.73}$$

For $r = r_h(R, E_{\mathbf{x}})$ Theorem 3 is equivalent to (2.73). Identity given in (2.72) follows from an analysis similar to the one used for establishing (2.69), in which instead of Lemma 3, we use a simple typicality argument like [12, Corollary 1.2].

QED

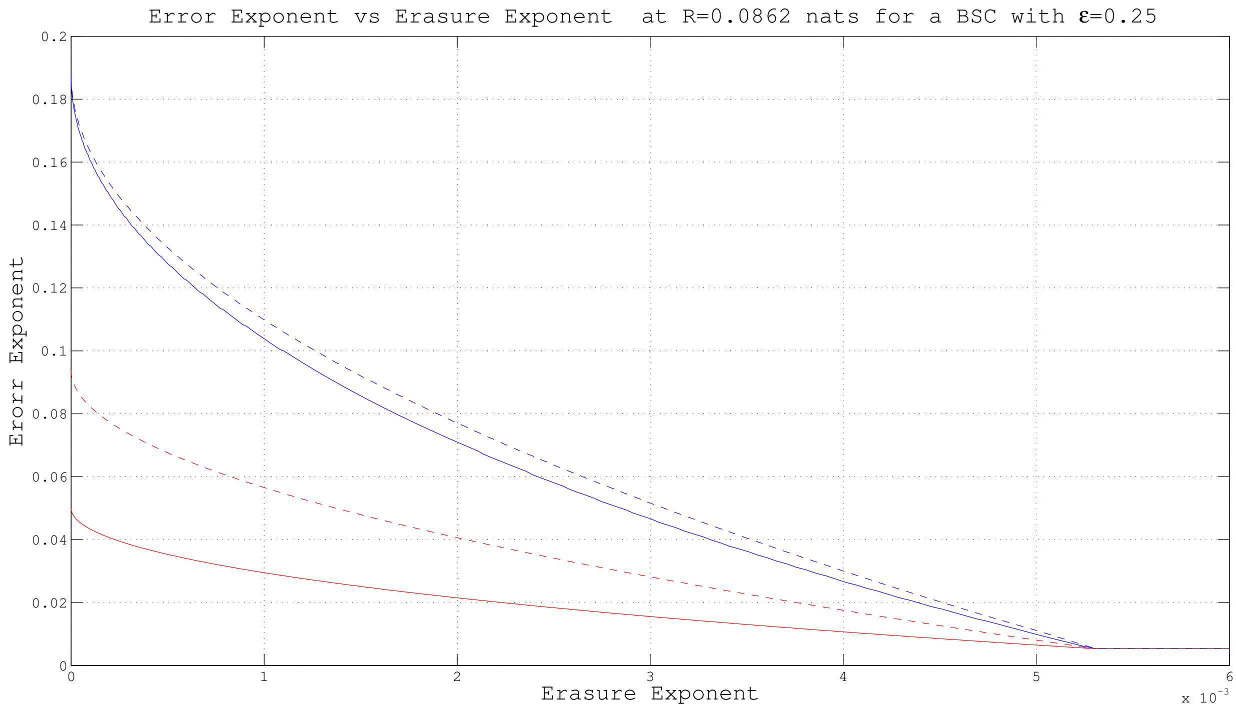

Figure 2-1: Error Exponent vs Erasure Exponent: Upper and lower bounds on error exponents as a function of erasure exponent for a binary symmetric channel with cross over probability $\epsilon = 0.25$ at rate $R = 8.62 \times 10^{-2}$ nats per channel use. Solid lines are lower bounds to the error exponent for block codes with feedback established in Section 2.3, and without feedback established previously [15], [12], [41]. Dashed lines are the upper bounds obtained using Theorem 3.

We have set $L_1 = 1$ in the proof but we could have set it to any subexponential function of block length which diverges as n diverges. By doing so 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 $E_{\mathbf{x}} \leq T^*$ final result is same for case with feedback.[15] 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 subexponential function of block length one can use Telatar's converse result [41, Theorem 4.4] on the error exponent at zero rate and zero erasure exponent without feedback.

In Figure 2-1, the upper and lower bounds we have derived for error exponent are plotted as a function of erasure exponent for a binary symmetric channel with cross over probability $\epsilon = 0.25$ at rate $R = 8.62 \times 10^{-2}$ nats per channel use. Note that all four curves meet at a point on bottom right, this is the point that corresponds to the error exponent of block codes at rate $R = 8.62 \times 10^{-2}$ nats per channel use and its values are the same with and without feedback since the channel is symmetric and the rate is over the critical rate. Any point to the lower right of this point is achievable both with and without feedback.

[15] In binary symmetric channels these result can be strengthened using the value of $\mathcal{E}(0)$, [45]. However those changes will improve the upper bound on error exponent only at low rates and high erasure exponents.

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

For all DMC's which has one or more zero probability transitions, for all rates below capacity, $R \leq \mathcal{C}$ 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 described in Section 2.3 gives an error free code. The connection between the erasure exponent of error free block codes, and 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 formally.

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

$$P_{\mathbf{e}}^{(\mathsf{n})} = 0 \quad \forall \mathsf{n}, \qquad \textit{and} \qquad \limsup_{\mathsf{n}\to\infty}(P_{\mathbf{x}}^{(\mathsf{n})} + \tfrac{1}{|\mathcal{M}^{(\mathsf{n})}|}) = 0.$$

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 $\mathbf{P}[y^{\mathsf{n}}|\, m] \geq \delta^{n}$ for all $m \in \mathcal{M}$ and $y^{\mathsf{n}} \in \mathcal{Y}^{\mathsf{n}}$. Consequently $\mathcal{C}_{\mathbf{x},0}$ is zero. On the other hand as an immediate consequence of the encoding scheme suggested by Yamamoto and Itoh in [43], if there is one or more zero probability transitions, $\mathcal{C}_{x,0}$ is equal to channel capacity $\mathcal{C}$.

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

$$\mathcal{E}_{\mathbf{x}}(R) \triangleq \sup_{\mathcal{Q}_0 : R(\mathcal{Q}_0) \geq R} E_{\mathbf{x}}(\mathcal{Q}_0). \tag{2.74}$$

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 infinite error exponent:

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

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 non-zero transition probability. Note that if $\mathbf{P}[y^{\mathsf{n}}|\, m] \neq 0$ then $\mathbf{P}[y^{\mathsf{n}}|\, m] \geq \delta^{\mathsf{n}}$. Thus if $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] \neq 0$, then $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] \geq \delta^{\mathsf{n}} e^{-\mathsf{n}R}$, i.e. $\frac{-\ln P_{\mathbf{e}}^{(\mathsf{n})}}{\mathsf{n}} \leq R - \ln \delta$. However if $E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R)$ then there is no error free reliable sequence at rate R with erasure exponent $E_{\mathbf{x}}$. Thus $P_{\mathbf{e}}^{(\mathsf{n})} > 0$ for infinitely many n in any reliable sequence and error exponent of all of those codes are bounded above by a finite number. Consequently,

$$E_{\mathbf{x}} > \mathcal{E}_{\mathbf{x}}(R) \Rightarrow \mathcal{E}_{\mathbf{e}}(R, E_{\mathbf{x}}) < \infty.$$

In a sense like the error exponent of erasure free block codes, $\mathcal{E}(R)$, 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 erasure exponent can be pushed to infinity and $\mathcal{E}_{\mathbf{x}}(R)$ gives the value of erasure exponent below which error exponent can be pushed to infinity.

Below 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. $\mathcal{C}_0 = 0$.

2.5.1 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 then 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 transmitter uses a length $\mathsf{n}_1 = \lceil e^{\mathsf{n}_1 R} \rceil$ block code without feedback with a size $L = \lceil \frac{\partial}{\partial R} E_{sp}(R, P_R^*) \rceil$ list decoder, where P_R^* is an optimal input distribution.[16] At rate R with this list size is sphere packing exponent is achievable.[17] Thus correct message is in the list with at least probability $(1 - e^{-\mathsf{n}_1 E_{sp}(R)})$, see [12, Page 196]. In the second phase transmitter uses a zero error code, of length[18] $\mathsf{n}_2 = \lceil \frac{\ln(L+1)}{\mathcal{C}_0} \rceil$ with $L+1$ messages, to tell the receiver whether the correct message is in that list or not, and to specify 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 over all 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, diverges, 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 error probability of the erasure free block codes with feedback decreases with an exponent no larger than $E_H(R)$, [20]. Thus,

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

This upper bound on the erasure exponent also follows from the outer bound we present in the next section, Theorem 6.

QED

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

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

This case is more involved than the previous one. We first 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.

---

[16] P_R^* is an input distribution satisfying $E_{sp}(R) = E_{sp}(R, P_R^*)$

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

[18] 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 that transmissions, which does not change with n_1. Thus above argument holds as is even in those cases.

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 $\mathcal{E}_{\mathbf{x},2}$ can be written as

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

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*

$$\mathcal{P}_{0,\mathbf{x}}(M,\mathsf{n},L)\geq\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)\mathcal{P}_{0,\mathbf{x}}\left(L_1+1,\mathsf{n}-\mathsf{n}_1,L\right). \tag{2.75}$$

Like Theorem 2, Theorem 5 is correct both with and without feedback. Although $\mathcal{P}_{0,\mathbf{x}}$'s and $\mathcal{P}_{\mathbf{e}}$ are different in each case, the relationship between them given in equation (2.75) holds in both cases.

Proof:
If $\mathcal{P}_{\mathbf{e}}(M,\mathsf{n}_1,L_1,0)=0$ 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

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

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,

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

Thus

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

QED

As it was done in the errors and erasures case, one can convert this into a bound on exponents. Using the improved version of Haroutunian's bound, i.e. Lemma 4, as an upper bound on the error exponent of erasure free block codes one gets the following.

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

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

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

Lemma 5 *For the channels which has zero zero-error capacity, i.e. $\mathcal{C}_0=0$, 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, in the zero rate case the number of messages are increasing with block length to infinity, thus their equality is not an immediate consequence of the definitions of $\mathcal{E}_{\mathbf{x},2}$ and $\mathcal{E}_{\mathbf{x}}(0)$.

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

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

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 block code of length $\ell(\epsilon)$:

$$\epsilon \geq \mathcal{P}_{0,\mathbf{x}}(2, \ell(\epsilon), 1). \tag{2.76}$$

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 then $k - r$. Let $\#\mathbf{x}$ denote the number of erasures out of k trials then

$$\mathbf{P}[\#\mathbf{x} = l] = \tfrac{k!}{(k-l)!l!}(1-\epsilon)^{k-l}\epsilon^{l} \qquad \text{and} \qquad P_{\mathbf{x}} = \sum\nolimits_{l=k-r+1}^{k} \mathbf{P}[\#e = l]\,.$$

Thus

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

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

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

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

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

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

QED

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

$$\mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1) \geq \tfrac{1}{2}\big(\sup_{s\in(0,.5)} \beta(s)\big)^{\mathsf{n}} \quad \text{where} \quad \beta(s) = \min\nolimits_{x,\tilde{x}} \sum\nolimits_{y} W(y|x)^{(1-s)} W(y|\tilde{x})^{s}.$$

Thus as a result of Lemma 5 we have

$$\mathcal{E}_{\mathbf{x}}(0) = \mathcal{E}_{\mathbf{x},2} \leq -\ln \sup_{s\in(0,.5)} \beta(s).$$

2.5.3 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 can not to 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,

$$\mathcal{P}_{0,\mathbf{x}}(2, \mathsf{n}, 1) \geq \left(\min\nolimits_{x,\tilde{x}} \sum\nolimits_{y:W(y|\tilde{x})} W(y|x)\right)^{\mathsf{n}}. \tag{2.77}$$

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

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

Let us denote the output symbol reachable from both of the input letters by $\tilde{y}$. If Y^{n} is a sequence of $\tilde{y}$'s then the receiver can not decode without errors, i.e. it has to declare and erasure. Thus

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

where (a) hods because arithmetic mean is larger than the geometric mean, and (b) holds because

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

Indeed this is bound is tight.[19] If the encoder assigns first message to the input letter that always leads to $\tilde{y}$ and the second message to the other input letter in first $\lfloor\frac{\mathsf{n}}{2}\rfloor$ time instances, and does the flipped assignment in the last $\lceil\frac{\mathsf{n}}{2}\rceil$ time instances, then an erasure happens with a probability less than $q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$.

Note that equation (2.77) bounds $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1)$ only by q^{n}, rather than $q^{\lfloor\frac{\mathsf{n}}{2}\rfloor}$. Using the insight from this example one can establish the following lower bound,

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) \geq \tfrac{1}{2}\left(\min_{x,\tilde{x}}\sum\nolimits_{y}\sqrt{W(y|x)W(y|\tilde{x})}\right)^{\mathsf{n}}. \tag{2.78}$$

However the bound given in equation (2.78) is decaying exponentially in n, even when all entries of the W are positive, i.e. even when $\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) = 1$. In other words it is not superior to the bound given in equation (2.77). Following bound implies bounds given in equations (2.77) and (2.78). Furthermore for certain channels it is strictly better than both.

Lemma 6 *Minimum erasure probability of error free codes with two messages is lower bounded as*

$$\mathcal{P}_{0,\mathbf{x}}(2,\mathsf{n},1) \geq \tfrac{1}{2}(\sup_{s\in(0,.5)}\beta(s))^{\mathsf{n}} \quad \textit{where} \quad \beta(s) = \min_{x,\tilde{x}}\sum\nolimits_{y} W(y|x)^{(1-s)}W(y|\tilde{x})^{s} \tag{2.79}$$

Proof:
For any error free code and for any $s \in (0, 0.5)$

$$\begin{aligned}P_{\mathbf{x}} &= \tfrac{1}{2}\sum\nolimits_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0}\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]\left(1+\tfrac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]}{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]}\right) \\ &\geq \tfrac{1}{2}\sum\nolimits_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0}\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]\left(1+\tfrac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]}{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]}\right)^{s} \\ &\geq \tfrac{1}{2}\sum\nolimits_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=2]>0}\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]^{1-s}\,\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=2]^{s}.\end{aligned} \tag{2.80}$$

[19] I would like to thank Emre Telatar for sharing these observations about z-channel.

Furthermore

$$\begin{aligned}
\sum_{y^{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=2]>0} & \mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]^{1-s}\,\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=2]^{s} \\
&= \sum_{y^{\mathsf{n}-1}:\mathbf{P}[y^{\mathsf{n}-1}|\,\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}-1}|\,\mathsf{M}=2]>0} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=1\right]^{1-s} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=2\right] \sum_{y_{\mathsf{n}}:\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=2]>0} \mathbf{P}\left[y_{\mathsf{n}}\middle|\,\mathsf{M}=1,y^{\mathsf{n}-1}\right]^{1-s} \mathbf{P}\left[y_{\mathsf{n}}\middle|\,\mathsf{M}=2,y^{\mathsf{n}-1}\right] \\
&\geq \sum_{y^{\mathsf{n}-1}:\mathbf{P}[y^{\mathsf{n}-1}|\,\mathsf{M}=1]\mathbf{P}[y^{\mathsf{n}-1}|\,\mathsf{M}=2]>0} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=1\right]^{1-s} \mathbf{P}\left[y^{\mathsf{n}-1}\middle|\,\mathsf{M}=2\right] \beta(s) \\
&\geq \frac{1}{2}(\beta(s))^{\mathsf{n}} \qquad (2.81)
\end{aligned}$$

Lemma 6 follows equations (2.80) and (2.81) by taking the supremum over $s \in (0, 0.5)$.

QED

The bound in equation (2.77) is implied by $\lim_{s\to 0^+}\beta(s)$ and bound in equation (2.78) is implied by $\lim_{s\to 0.5^-}\beta(s)$. However bound given by Lemma 6 does not follow from the bounds given in equation (2.77) and (2.78). In order to see this note that; although $\sum_y W(y|x)^s W(y|\tilde{x})^{1-s}$ 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 and there exist channels for which bound given in Lemma 6 is strictly better than the bounds given in (2.77) and (2.78). Following is the transition probabilities of one such channel.

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

$\lim_{s\to 0^+}\beta(s) = 0.7$, $\lim_{s\to 0.5^-}\beta(s) = 0.7027$, $\beta(0.18) = 0.7299$.

2.6 Conclusions

In the erasure-free case, the error exponent is not known for a general DMC with feedback. It is not even known if it is still upper bounded by sphere packing exponent for non-symmetric DMCs. However for the case with erasures, at zero erasure exponent, the value of error exponent has been known for a long time, [4], [43]. The main aim in this chapter was establishing upper and lower bounds that will extend the bounds at the zero erasure exponent case gracefully and non-trivially to the positive exponents. Results of this chapter are best understood in this framework and should be interpreted accordingly.

We derived inner bounds using a two phase encoding schemes, which are known to be optimal at zero-erasure exponent case. We have improved their performance at positive erasure exponent values by choosing relative durations of the phases properly and by using an appropriate decoder. However within each phase the assignment of messages to input letters is fixed. In a general feedback encoder, on the other hand, 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, [44], [14], [7], [31] 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 communication phase to the control phase. However in order to apply those ideas effectively for a general DMC, it seems one first needs to solve the problem for the erasure-free block codes for a general DMC. The analysis presented in Chapter 4 is a step in that direction.

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 and erasure exponent points by taking their convex combinations. This approach can be interpreted as a generalization of the outer bounds used for variable-length block codes, [4], [2]. As it was the case for the inner bounds, it seems in order to improve the outer bounds one needs establish outer bounds on some related problems: on the error exponents of erasure free block codes with feedback and on the error exponent erasure exponent trade of at zero rate.

Chapter 3

Bit-Wise Unequal Error Protection for Variable-Length Block Codes[1]

In the classic framework for point to point communication, [36], there is a tacit assumption that any error event is as undesirable as any other error event. Consequently one is only concerned with the average error probability or maximum conditional error probability among the messages. In many applications, however, not all error events are equally undesirable. In those situations one can group the error events into different classes and analyze probabilities of these classes of error events separately. In order to prioritize protection against one or the other class of error events, corresponding error exponent is increased at the expense of the other error exponents.

This information theoretic perspective on unequal error protection (*UEP*) was first pursued by Csiszar in his work on joint source channel coding [11]. He showed that for any integer k one can have a length n code such that $\mathcal{M} = \cup_{j=1}^{k} \mathcal{M}_j$ where $|\mathcal{M}_j| \approx e^{\mathsf{n} R_j}$ and conditional error probability of each message in each subset $|\mathcal{M}_j|$ is $\approx e^{-\mathsf{n} E_r(R_j)}$. The problem considered by Csiszar in [11] is a *message-wise UEP* problem, because the probability of the class of error events in consideration can be expressed solely in terms of the conditional error probabilities of the messages.

Bit-wise UEP problems are the other canonical form of *UEP* problems. In *Bit-wise UEP* problems the error events in consideration can be expressed solely in terms of error probabilities of different groups of bits. Consider for example the situation where the message set is of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_k$ and $|\mathcal{M}_j| \approx e^{\mathsf{n} R_j}$. Then messages are of the form $\mathsf{M} = (\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_k)$ and one can study the exponential decay rate of the error probabilities associated with different groups, $\mathbf{P}\left[\mathsf{M}_j \neq \hat{\mathsf{M}}_j\right]$ for $j = 1, 2, \ldots, k$, separately.

In order to demonstrate the stark difference between the *bit-wise* and the *message-wise UEP* problems it is shown in [3] that if $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$, $\mathcal{M}_1 = \{0, 1\}$, $\mathcal{M}_2 = \{0, 1, \ldots, e^{\mathsf{n}\mathcal{C}}\}$ and $\mathbf{P}\left[\mathsf{M}_2 \neq \hat{\mathsf{M}}_2\right] \approx 0$ then[2] $\frac{-\ln \mathbf{P}\left[\mathsf{M}_1 \neq \hat{\mathsf{M}}_1\right]}{\mathsf{n}} \approx 0$. Thus in the *bit-wise UEP* problem even a bit can not have a positive error exponent. But as result of [11] we know that in the *message-wise UEP* problem a two-element subset of the message set can have an error exponent as high as $E_r(0) > 0$, i.e. $\frac{-\ln \mathbf{P}\left[\hat{\mathsf{M}} \neq m \middle| \mathsf{M} = m\right]}{\mathsf{n}} \approx E_r(0)$ for $m = 1, 2$. As noted in [3], there are many *UEP* problems of practical importance that are neither *message-wise UEP* nor *bit-wise UEP* problems. Yet studying these special cases is a good starting point.

[1] The results presented in this chapter are part of a joint work with Siva K. Gorantla of UIUC, [19].

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

In this chapter two closely related *UEP* problems, the multi-layer *bit-wise UEP* problem and the single message *message-wise UEP* problem are considered. In the *bit-wise UEP* problem there are multiple groups of bits each with different priorities and rates; the aim is characterizing the tradeoff between the rates and error exponents of the layers, by revealing the volume of achievable rate vector, exponent vector pairs. In the single message *message-wise UEP* problem, failing to detect the special message when it is transmitted is much more costly than failing to detect ordinary messages and we characterize the tradeoff between best missed detection exponent of a message of the code and over all error exponent of the code. Both of these problems were first considered in [3], for the case when overall rate is (very close to) the channel capacity. For many of the *UEP* problems not only the channel model but also the particular family of codes in consideration makes a big difference and problems considered in this chapter are no exception. We will be investigating these problems for variable-length block codes on DMCs, like the ones in [4].

First we formally define variable-length block codes and the two unequal error protection problems of interest. Then we present the main results of the chapter by giving the achievable region of rate and error exponent vectors for the two problems in consideration. For both of the problems the proofs of achievability and converse are presented in Section 3.2 and Section 3.3 respectively. We will finish our discussion of *UEP* problems for variable-length block codes by summarizing our qualitative conclusions in Section 3.4.

3.1 Model and Main Results

3.1.1 Variable-Length Block Codes

A variable-length block code on a DMC is given by a decoding time τ, an encoding scheme Φ and a decoding rule Ψ. Decoding time τ is a stopping time with respect to the receivers observation.[3] For each $y^t \in \mathcal{Y}^t$ such that $t < \tau$, encoding scheme $\Phi(\cdot, y^t)$ determines the input letter at time $(t+1)$ for each message in the message set $\mathcal{M}$,

$$\Phi(\cdot, y^t) : \mathcal{M} \to \mathcal{X} \qquad \forall y^t : t < \tau.$$

The decoding rule is a mapping from the set of output sequences at τ to the message set $\mathcal{M}$ which determines the decoded message,

$$\Psi(\cdot) : \mathcal{Y}^\tau \to \mathcal{M}.$$

At time zero a message M chosen uniformly at random from $\mathcal{M}$ is given to the transmitter; transmitter uses the codeword associated with this message, i.e. $\Phi(\mathsf{M}, \cdot)$, to convey the message until the decoding time τ. Then the receiver decodes a message $\hat{\mathsf{M}} = \Psi(\mathsf{Y}^\tau)$. The error probability and the rate of a variable-length block code are given by

$$P_{\mathbf{e}} = \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] \qquad\qquad R = \tfrac{\ln |\mathcal{M}|}{\mathbf{E}[\tau]}.$$

One can interpret the variable-length block codes on DMCs as trees; a detailed discussion of this perspective is given in [2, Section II].

[3]In other words given Y^t the value of $\mathbb{1}_{\{\tau > t\}}$ is known, where $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

3.1.2 Reliable Sequences for Variable-Length Block Codes

In order to suppress the secondary terms while discussing the main results we use the concept of reliable sequences. In a sequence of codes we denote the error probability and the message set of the n^{th} code of the sequence by $P_{\mathbf{e}}{}^{(n)}$ and $\mathcal{M}^{(n)}$, respectively. Similarly, whenever necessary in order to avoid confusion, we denote the probability of an event under the probability measure resulting from the n^{th} code of the code sequence by $\mathbf{P}^{(n)}[\cdot]$ instead of $\mathbf{P}[\cdot]$.

A sequence of variable-length block codes $\mathcal{Q}$ is reliable iff its error probability vanishes and the size of its message set diverges:

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

The rate and the error exponent of a reliable sequence $\mathcal{Q}$ is defined as,

$$R_{\mathcal{Q}} \triangleq \liminf_{n\to\infty} \tfrac{\ln |\mathcal{M}^{(n)}|}{\mathbf{E}[\tau^{(n)}]} \qquad E_{\mathcal{Q}} \triangleq \liminf_{n\to\infty} \tfrac{-\ln P_{\mathbf{e}}{}^{(n)}}{\mathbf{E}[\tau^{(n)}]}.$$

Furthermore the reliability function of the variable-length block codes is defined as,

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

Burnashev [4] determined the reliability function of variable-length block codes on DMCs for all rates. According to [4]:

- If all entries of $W(\cdot|\cdot)$ are positive then

  $$E(R) = \left(1 - \tfrac{R}{\mathcal{C}}\right) \mathcal{D}$$

  where $\mathcal{D}$ is maximum Kullback Leibler divergence between the output distributions of any two input letters:

  $$\mathcal{D} = \max_{x,\tilde{x} \in \mathcal{X}} \mathsf{D}\left(W_x \| W_{\tilde{x}}\right).$$

- If there are one or more zero entries[4] in $W(\cdot|\cdot)$, i.e. if there are two input letters x_1, x_2 and an output letter y such that, $W(y|x_1) = 0$ and $W(y|x_2) > 0$, then for all $R < \mathcal{C}$, for large enough $\mathbf{E}[\tau]$ there are rate R variable-length block codes which are error free, i.e. $P_{\mathbf{e}} = 0$.

When $P_{\mathbf{e}} = 0$, all of the messages and bits can have zero error probability simultaneously. Consequently all the *UEP* problems are answered trivially when there is a zero probability transition. Thus we assume hence forth that $W(y|x) > 0$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ and denote the smallest transition probability by λ, i.e. $\min_{x,y} W(y|x) = \lambda > 0$. Furthermore we denote the input letters that get this maximum value of Kullback Leibler divergence by x_a and x_r:

$$\mathcal{D} = \mathsf{D}\left(W_{x_a} \| W_{x_r}\right) \tag{3.1}$$

where $W_x(\cdot) = W(\cdot|x)$.

[4] Note that in this situation $\mathcal{D} = \infty$.

3.1.3 Exponents for *UEP* in Variable-Length Block Codes with Feedback

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

$$P_{\mathbf{e}|m} \triangleq \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathsf{M} = m\right]. \tag{3.2}$$

In the conventional setting, either the average or the maximum of the conditional error probabilities of the messages is studied. However, in many situations different kinds of error events have different costs and the subtlety between different error events can not be captured by the conditional error probabilities of the messages. In particular when different groups of bits of the message have different importance; we need to study the error events associated with each group separately. For doing that we consider message sets of the form $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_k$ and analyze the probabilities of the events $\hat{\mathsf{M}}_i \neq \mathsf{M}_i$, separately. To that end let $P_{\mathbf{e},i}$ for each $i \in \{1, 2, \ldots, k\}$ be

$$P_{\mathbf{e},i} \triangleq \mathbf{P}\left[\hat{\mathsf{M}}_i \neq \mathsf{M}_i\right] \tag{3.3}$$

We assume, without loss of generality, that indexing order is also the importance order for different groups of bit, i.e.

$$P_{\mathbf{e},1} \leq P_{\mathbf{e},2} \leq P_{\mathbf{e},3} \leq \ldots \leq P_{\mathbf{e},k}. \tag{3.4}$$

Definition 5 *For any reliable sequence $\mathcal{Q}$ whose message sets $\mathcal{M}^{(n)}$ are of the form*

$$\mathcal{M}^{(n)} = \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \times \ldots \times \mathcal{M}_k^{(n)} \tag{3.5}$$

the rate and the error exponents of the layers are defined as,

$$R_{\mathcal{Q},i} \triangleq \liminf_{n\to\infty} \frac{\ln |\mathcal{M}_i^{(n)}|}{\mathbf{E}\left[\tau^{(n)}\right]} \qquad \forall i \in \{1, 2, \ldots, k\}$$
$$E_{\mathcal{Q},i} \triangleq \liminf_{n\to\infty} \frac{-\ln \mathbf{P}^{(n)}\left[\hat{\mathsf{M}}_i \neq \mathsf{M}_i\right]}{\mathbf{E}\left[\tau^{(n)}\right]} \qquad \forall i \in \{1, 2, \ldots, k\}.$$

Closure of the points of the form $(\vec{R}_{\mathcal{Q}}, \vec{E}_{\mathcal{Q}})$ is the achievable volume of rate vector exponent vector pairs.

In characterizing the tradeoff for the *bit-wise UEP* problem, the *message wise UEP* problem with a single special message plays a key role. In the single message *message wise UEP* problem the tradeoff between exponential decay rates of $P_{\mathbf{e}}$ and $\min_{m\in\mathcal{M}} P_{\mathbf{e}|m}$ is studied. The operational definition of the problem in terms of reliable sequences is as follows.

Definition 6 *For any reliable sequence $\mathcal{Q}$ missed detection exponent is defined as:*

$$E_{md\mathcal{Q}} = \liminf_{n\to\infty} \frac{-\ln \mathbf{P}^{(n)}\left[\hat{\mathsf{M}} \neq 1 \middle| \mathsf{M} = 1\right]}{\mathbf{E}\left[\tau^{(n)}\right]}. \tag{3.6}$$

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

$$E_{md}(R, E) \triangleq \sup_{\mathcal{Q}: R_{\mathcal{Q}} \geq R,\ E_{\mathcal{Q}} \geq E} E_{md\mathcal{Q}}. \tag{3.7}$$

3.1.4 Main Results

The results of both the multi-layer *bit-wise UEP* problem and the single message *message-wise UEP* problem are give in terms of the function $\mathcal{J}(R)$ defined below.

Definition 7

$$\mathcal{J}(R) \triangleq \max_{\substack{\alpha, x_1, x_2, \mathsf{P}_{X,1}, \mathsf{P}_{X,2}: \\ \alpha \mathsf{I}(\mathsf{P}_{X,1}, W) + (1-\alpha)\mathsf{I}(\mathsf{P}_{X,2}, W) \geq R}} \alpha \mathsf{D}\left(\mathsf{P}_{Y,1} \| W_{x_1}\right) + (1-\alpha)\mathsf{D}\left(\mathsf{P}_{Y,2} \| W_{x_2}\right) \tag{3.8a}$$

$$j(R) \triangleq \max_{x, \mathsf{P}_X: \mathsf{I}(\mathsf{P}_X, W) \geq R} \mathsf{D}\left(\mathsf{P}_Y \| W_x\right) \tag{3.8b}$$

where $\mathsf{P}_{Y,i}(y) = \sum_x W(y|x)\mathsf{P}_{X,i}(x)$ *for* $i = 1, 2$ *and* $\mathsf{P}_Y(y) = \sum_x \mathsf{P}_X(x)W(y|x)$.

Note that the function $\mathcal{J}(R)$ is the minimum concave function upper bounding the function $j(R)$ and $\mathcal{J}(0) = j(0) = \mathcal{D}$.

Theorem 7 *A rate vector error exponent vector pair* $(\vec{R}, \vec{E})$ *is achievable if and only if there exists a time sharing vector* $\vec{\eta}$ *such that,*

$$E_i \leq (1 - \sum_{j=1}^{k} \eta_j)\mathcal{D} + \sum_{j=i+1}^{k} \eta_j \mathcal{J}\left(\tfrac{R_j}{\eta_j}\right) \qquad \forall i \in \{1, 2, \ldots, k\} \tag{3.9a}$$

$$\tfrac{R_i}{\eta_i} \leq \mathcal{C} \qquad \forall i \in \{1, 2, \ldots, k\} \tag{3.9b}$$

$$\sum_{j=1}^{k} \eta_j \leq 1. \tag{3.9c}$$

Proof:
Use Lemma 12 for $\delta = \frac{-1}{\ln P_{\mathbf{e}}}$ together with Lemma 9 and equation (3.27).

QED

The region of achievable $(\vec{R}, \vec{E})$ pairs is convex. In order to see this, note that if $(\vec{R}_a, \vec{E}_a)$ and $(\vec{R}_b, \vec{E}_b)$ are achievable then there exist $(\vec{R}_a, \vec{E}_a, \vec{\eta}_a)$ and $(\vec{R}_b, \vec{E}_b, \vec{\eta}_b)$ triples satisfying (3.9). Thus as a result of concavity of $\mathcal{J}(\cdot)$, $(\alpha(\vec{R}_a, \vec{E}_a, \vec{\eta}_a) + (1-\alpha)(\vec{R}_b, \vec{E}_b, \vec{\eta}_b))$ also satisfies (3.9). Hence $(\alpha(\vec{R}_a, \vec{E}_a) + (1-\alpha)(\vec{R}_b, \vec{E}_b))$ is also achievable because of Theorem 7 .

For the case with two priority levels the condition given in Theorem 7 leads to an analytical expression for the optimal value of E_1 in terms of R_1, R_2 and E_2.

Corollary 2 *For any rate pair* (R_1, R_2) *such that* $R_1 + R_2 \leq \mathcal{C}$ *and error exponent* E_2 *such that* $E_2 \leq (1 - \frac{R_1 + R_2}{\mathcal{C}})\mathcal{D}$, *the optimal value of* E_1 *is given by*

$$\mathcal{E}_1(R_1, R_2, E_2) = E_2 + \left(1 - \tfrac{R_1}{\mathcal{C}} - \tfrac{E_2}{\mathcal{D}}\right) \mathcal{J}\left(\frac{R_2}{1 - \frac{R_1}{\mathcal{C}} - \frac{E_2}{\mathcal{D}}}\right). \tag{3.10}$$

Theorem 8 *For any rate* $0 \leq R \leq \mathcal{C}$ *and error exponent* $E \leq (1 - \frac{R}{\mathcal{C}})\mathcal{D}$ *missed detection exponent* $E_{md}(R, E)$ *is given by*

$$E_{md}(R, E) = E + \left(1 - \tfrac{E}{\mathcal{D}}\right) \mathcal{J}\left(\frac{R}{1 - \frac{E}{\mathcal{D}}}\right). \tag{3.11}$$

Proof:
Use Lemma 11 for $\delta = \frac{-1}{\ln P_{\mathbf{e}}}$ together with Lemma 8 and equation (3.19).

QED

3.2 Achievablity

The design of the codes that achieve the optimal performance use multiple ideas at the same time. In order to introduce those ideas one by one in a piecemeal fashion, we introduce various families of codes and discuss their properties.

We first consider a family of fixed-length block codes without feedback or erasures and establish an inner bound to the achievable rate, missed detection exponent pairs. The codes achieving this tradeoff have a positive missed detection exponent but their overall error exponent is zero, i.e. as we consider longer and longer codes the average error probability decays to zero but subexponentially in the block length. Then we append a control phase, like the one used by Yamamoto and Itoh in [43], to these codes to obtain a positive error exponent. These fixed-length block codes with feedback and erasures are then used as the building block for the variable-length block codes for the two *UEP* problems we are interested in. This encoding scheme can be seen as a generalization of an encoding scheme first suggested by Kudrayshov [22]. The key feature of the encoding scheme in [22] is the tacit acceptance and explicit rejection strategy, which was also used in [3]. We combine this strategy with a classic control phase with explicit acknowledgments to get a positive error exponent for all messages/bits. The outer bounds we derive in Section 3.3 reveal that such schemes are optimal, in terms of the decay rate of error probability with expected decoding time, $\mathbf{E}[\tau]$.

3.2.1 An Achievable Scheme without Feedback

Let us first consider a parametric family of codes in terms of two input-letter-input-distribution pairs $(x_1, \mathsf{P}_{X,1})$ and $(x_2, \mathsf{P}_{X,2})$ and a time sharing constant α. In these codes the codeword of the special message is concatenation of $\alpha\mathsf{n}$ x_1's and $(1-\alpha)\mathsf{n}$ x_2's and the codewords for the ordinary messages are generated via random coding with time sharing. A joint typicality decoding is used for the ordinary messages, and pretty much all of the output sequences that are not decoded to an ordinary message is decoded as the special message. The error analysis of these codes and resulting performance are presented below.

Lemma 7 *For any block length* n*, time sharing constant* $\alpha \in [0,1]$*, input distribution input letter pairs* $(x_1, \mathsf{P}_{X,1})$ *and* $(x_2, \mathsf{P}_{X,2})$ *there exists a fixed-length block code such that*

$$\begin{aligned} |\mathcal{M}|-1 &\geq e^{\mathsf{n}(\alpha \mathsf{I}(\mathsf{P}_{X,1},W)+(1-\alpha)\mathsf{I}(\mathsf{P}_{X,2},W)-\epsilon_1(\mathsf{n}))} \\ P_{\mathbf{e}|m} &\leq \epsilon_2(\mathsf{n}) \qquad\qquad m=1,2,3,\ldots,|\mathcal{M}|-1 \\ P_{\mathbf{e}|0} &\leq e^{-\mathsf{n}(\alpha \mathsf{D}(\mathsf{P}_{Y,1}\|W_{x_1})+(1-\alpha)\mathsf{D}(\mathsf{P}_{Y,2}\|W_{x_2})-\epsilon_3(\mathsf{n}))} \end{aligned}$$

where $\epsilon_i(\mathsf{n}) \geq 0$ *and* $\lim_{\mathsf{n}\to\infty} \epsilon_i(\mathsf{n}) = 0$ *for* $i = 1,2,3$.

Proof:
Recall that the output distributions resulting from $\mathsf{P}_{X,1}$ and $\mathsf{P}_{X,2}$ under $W(\cdot|\cdot)$ are denoted by $\mathsf{P}_{Y,1}$ and $\mathsf{P}_{Y,2}$:

$$\mathsf{P}_{Y,i}(y) = \sum\nolimits_x W(y|x)\mathsf{P}_{X,i}(x) \quad i=1,2.$$

Let n_1 be $\mathsf{n}_1 = \lfloor \mathsf{n}\alpha \rfloor$. The codeword of the message $\mathsf{M}=0$ is the concatenation of n_1 x_1's and $(\mathsf{n}-\mathsf{n}_1)$ x_2's. The decoding region of the first message is given by

$$\mathcal{G}(0)=\left\{y^{\mathsf{n}}:\mathsf{n}_1\Delta_{(\mathsf{Q}_{(y_1^{\mathsf{n}_1})};\mathsf{P}_{Y,1})}+(\mathsf{n}-\mathsf{n}_1)\Delta_{(\mathsf{Q}_{(y_{\mathsf{n}_1}^{\mathsf{n}})};\mathsf{P}_{Y,2})}\geq \mathsf{n}^{3/4}\right\}$$

where $\mathsf{Q}_{(y_1^{\mathsf{n}_1})}$ denotes the empirical distribution of $y_1^{\mathsf{n}_1}$ and $\Delta_{(\cdot;\cdot)}$ denotes the total variation between the two distributions. Then

$$\begin{aligned}P_{\mathsf{e}|0}&=\mathbf{P}\left[\hat{\mathsf{M}}\neq 0\middle|\,\mathsf{M}=0\right]\\&=\sum\nolimits_{y^{\mathsf{n}}\notin\mathcal{G}(0)}\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M}=0]\,.\end{aligned}$$

For all $y^{\mathsf{n}}\notin\mathcal{G}(0)$ we have,

$$\begin{aligned}&\mathsf{n}_1\mathsf{D}\left(\mathsf{Q}_{(y_1^{\mathsf{n}_1})}\middle\|\,W_{x_1}\right)+(\mathsf{n}-\mathsf{n}_1)\mathsf{D}\left(\mathsf{Q}_{(y_{\mathsf{n}_1}^{\mathsf{n}})}\middle\|\,W_{x_2}\right)\\&\qquad=\mathsf{n}_1\mathsf{D}\left(\mathsf{Q}_{(y_1^{\mathsf{n}_1})}\middle\|\,\mathsf{P}_{Y,1}\right)+(\mathsf{n}-\mathsf{n}_1)\mathsf{D}\left(\mathsf{Q}_{(y_{\mathsf{n}_1}^{\mathsf{n}})}\middle\|\,\mathsf{P}_{Y,2}\right)\\&\qquad\quad+\mathsf{n}_1\sum\nolimits_y\mathsf{Q}_{(y_1^{\mathsf{n}_1})}(y)\ln\tfrac{\mathsf{P}_{Y,1}(y)}{W(y|x_1)}+(\mathsf{n}-\mathsf{n}_1)\sum\nolimits_y\mathsf{Q}_{(y_{\mathsf{n}_1}^{\mathsf{n}})}(y)\ln\tfrac{\mathsf{P}_{Y,2}(y)}{W(y|x_2)}\\&\qquad\geq\mathsf{n}_1\mathsf{D}\left(\mathsf{P}_{Y,1}\|\,W_{x_1}\right)+(\mathsf{n}-\mathsf{n}_1)\mathsf{D}\left(\mathsf{P}_{Y,2}\|\,W_{x_2}\right)+2\mathsf{n}^{3/4}\ln\lambda.\end{aligned}$$

Furthermore, there are less than $(\mathsf{n}_1+1)^{|\mathcal{Y}|}$ distinct empirical distributions in the first phase and there are less than $(\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{Y}|}$ distinct empirical distributions in the second phase. Thus

$$\begin{aligned}P_{\mathsf{e}|0}&\leq(\mathsf{n}_1+1)^{|\mathcal{Y}|}(\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{Y}|}e^{-\mathsf{n}_1\mathsf{D}\left(\mathsf{P}_{Y,1}\|W_{x_1}\right)+(\mathsf{n}-\mathsf{n}_1)\mathsf{D}\left(\mathsf{P}_{Y,2}\|W_{x_2}\right)-2\mathsf{n}^{3/4}\ln\lambda}\\&\leq e^{-\mathsf{n}(\alpha\mathsf{D}\left(\mathsf{P}_{Y,1}\|W_{x_1}\right)+(1-\alpha)\mathsf{D}\left(\mathsf{P}_{Y,2}\|W_{x_2}\right)-\epsilon_3(\mathsf{n}))}\end{aligned}$$

where $\epsilon_3(\mathsf{n})=\frac{-2\mathsf{n}^{3/4}\ln\lambda+\mathcal{D}+2|\mathcal{Y}|\ln(\mathsf{n}+1)}{\mathsf{n}}$.

The codewords of the remaining messages are specified using a random coding argument with the empirical typicality as follows. Consider an ensemble of codes in which first n_1 entries of all the codewords are independent and identically distributed (i.i.d.) with input distribution $\mathsf{P}_{X,1}$ and the rest of the entries are i.i.d. with the input distribution $\mathsf{P}_{X,2}$. For each message $m\in\mathcal{M}$ and codeword $x^{\mathsf{n}}(m)\in\mathcal{X}^{\mathsf{n}}$ let $\Gamma(x^{\mathsf{n}}(m))$ be the set of y^{n} for which $(x^{\mathsf{n}}(m),y^{\mathsf{n}})$ is typical with $(\alpha,\mathsf{P}_{X1}W,\mathsf{P}_{X2}W)$:

$$\Gamma(x^{\mathsf{n}}(m))=\{y^{\mathsf{n}}:\mathsf{n}_1\Delta_{(\mathsf{Q}_{(x^{\mathsf{n}_1}(m),y^{\mathsf{n}_1})};\mathsf{P}_{X1}W)}+(\mathsf{n}-\mathsf{n}_1)\Delta_{(\mathsf{Q}_{(\mathsf{x}_{\mathsf{n}_1+1}^{\mathsf{n}}(m),y_{\mathsf{n}_1+1}^{\mathsf{n}})};\mathsf{P}_{X2}W)}\leq\mathsf{n}^{3/4}\}.$$

Note that $\Gamma(x^{\mathsf{n}}(m))\cap\mathcal{G}(0)=\emptyset$ by definition for $m=1,2,3,\ldots|\mathcal{M}|-1$. Let the decoding region of the ordinary messages be

$$\mathcal{G}(m)=\Gamma(x^{\mathsf{n}}(m))\bigcap\left(\cap_{\tilde{m}\neq m}\overline{\Gamma(x^{\mathsf{n}}(\tilde{m}))}\right)\qquad\forall m\in\{1,2,3,\ldots,(|\mathcal{M}|-1)\}.$$

Then the average of the conditional error probability of m^{th} message over the ensemble is upper bounded as,

$$\mathbf{E}\left[P_{\mathsf{e}|m}\right]\leq\mathbf{P}[y^{\mathsf{n}}\notin\Gamma(x^{\mathsf{n}}(m))|\,\mathsf{M}=m]+\sum_{\tilde{m}\neq m}\mathbf{P}[y^{\mathsf{n}}\in\{\Gamma(x^{\mathsf{n}}(m))\cap\Gamma(x^{\mathsf{n}}(\tilde{m}))\}|\,\mathsf{M}=m]\,.$$

Let us start with bounding $\mathbf{P}[y^{\mathsf{n}} \notin \Gamma(x^{\mathsf{n}}(m))|\,\mathsf{M}=m]$. Note if the weighted sum of the total variations are greater than $\mathsf{n}^{3/4}$, then there exists at least one (x,y) pair such that

$$\zeta_1(x,y)+\zeta_2(x,y) \geq \tfrac{\mathsf{n}^{3/4}}{2|\mathcal{X}||\mathcal{Y}|}$$

where

$$\zeta_1(x,y)=\mathsf{n}_1|\mathsf{Q}_{(x^{\mathsf{n}_1}(m),y^{\mathsf{n}_1})}(x,y)-\mathsf{P}_{X1}(x)W(y|x)|, \tag{3.12a}$$

$$\zeta_2(x,y)=(\mathsf{n}-\mathsf{n}_1)|\mathsf{Q}_{(x^{\mathsf{n}}_{\mathsf{n}_1+1}(m),y^{\mathsf{n}}_{\mathsf{n}_1+1})}(x,y)-\mathsf{P}_{X2}(x)W(y|x)|. \tag{3.12b}$$

As a result of Chebyshev's inequality we have,

$$\begin{aligned}\mathbf{P}[y\notin\Gamma(x^{\mathsf{n}}(m))|\,\mathsf{M}=m] &\leq \mathbf{P}\left[\zeta_1(x,y)+\zeta_2(x,y)\geq \tfrac{\mathsf{n}^{3/4}}{2|\mathcal{X}||\mathcal{Y}|}\middle|\,\mathsf{M}=m\right]\\ &\leq \mathbf{E}\left[\frac{(\zeta_1(x,y)+\zeta_2(x,y))^2}{\frac{\mathsf{n}^{3/2}}{4|\mathcal{X}|^2|\mathcal{Y}|^2}}\middle|\,\mathsf{M}=m\right].\end{aligned} \tag{3.13}$$

Using Schwarz inequality we get,

$$\mathbf{E}\left[(\zeta_1(x,y)+\zeta_2(x,y))^2\middle|\,\mathsf{M}=m\right]\leq 2\mathbf{E}\left[\zeta_1(x,y)^2+\zeta_2(x,y)^2\middle|\,\mathsf{M}=m\right]. \tag{3.14}$$

Using equations(3.12), (3.13) and (3.14) we bound $\mathbf{P}[y^{\mathsf{n}}\notin\Gamma(x^{\mathsf{n}}(m))|\,\mathsf{M}=m]$ as follows,

$$\mathbf{P}[y^{\mathsf{n}}\notin\Gamma(\mathbf{x}^{\mathsf{n}}(m))|\,\mathsf{M}=m]\leq 2\frac{\frac{1}{4}\mathsf{n}}{(\frac{\mathsf{n}^{3/4}}{2|\mathcal{X}||\mathcal{Y}|})^2}=2(|\mathcal{X}||\mathcal{Y}|)^2\mathsf{n}^{-1/2}.$$

For bounding $\mathbf{P}[y^{\mathsf{n}}\in\Gamma(x^{\mathsf{n}}(m))\cap\Gamma(x^{\mathsf{n}}(\tilde{m}))|\,\mathsf{M}=m]$ terms, we use the fact that n i.i.d. trials with probability distribution P will have an empirical type Q with probability $e^{-\mathsf{n}\mathsf{D}(Q\|P)}$. Furthermore n trials on the alphabet $\mathcal{Z}$ has less than $(\mathsf{n}+1)^{|\mathcal{Z}|}$ different empirical types. Thus

$$\begin{aligned}\mathbf{P}[\Gamma_m\cap\Gamma_{\tilde{m}}|\,\mathsf{M}=m] &\leq (\mathsf{n}_1+1)^{|\mathcal{X}||\mathcal{Y}|}(\mathsf{n}-\mathsf{n}_1+1)^{|\mathcal{X}||\mathcal{Y}|}e^{-\mathsf{n}_1\mathsf{I}(\mathsf{P}_{X,1},W)-(\mathsf{n}-\mathsf{n}_1)\mathsf{I}(\mathsf{P}_{X,2},W)-2\mathsf{n}^{3/4}\ln\lambda}\\ &\leq e^{-\mathsf{n}(\alpha\mathsf{I}(\mathsf{P}_{X,1},W)+(1-\alpha)\mathsf{I}(\mathsf{P}_{X,2},W))}e^{\mathcal{C}+2|\mathcal{X}||\mathcal{Y}|\ln\mathsf{n}-2\mathsf{n}^{3/4}\ln\lambda}.\end{aligned}$$

Hence if $|\mathcal{M}|-1=\mathsf{n}^{-1}e^{\mathsf{n}(\alpha\mathsf{I}(\mathsf{P}_{X,1},W)+(1-\alpha)\mathsf{I}(\mathsf{P}_{X,2},W))}e^{-\mathcal{C}-2|\mathcal{X}||\mathcal{Y}|\ln\mathsf{n}+2\mathsf{n}^{3/4}\ln\lambda}$ then

$$\sum_{\tilde{m}\neq m}\mathbf{P}[\Gamma_m\cap\Gamma_{\tilde{m}}|\,\mathsf{M}=m]\leq\tfrac{1}{\mathsf{n}}.$$

Thus the average $P_{\mathbf{e}}$ over the ensemble is bounded as

$$\mathbf{E}[P_{\mathbf{e}}]\leq\tfrac{2(|\mathcal{X}||\mathcal{Y}|)^2}{\mathsf{n}^{1/2}}+\tfrac{1}{\mathsf{n}}.$$

But if the ensemble average of the error probability is upper bounded like this there is, at least one code that has this low error probability. Furthermore half of its messages have conditional error probabilities less then twice this average. Thus lemma holds for,

$$\epsilon_1(\mathsf{n})=\frac{\mathcal{C}+(2|\mathcal{X}||\mathcal{Y}|+1)\ln\mathsf{n}-\mathsf{n}^{3/4}\ln\lambda+\ln 2}{\mathsf{n}}\qquad\qquad \epsilon_2(\mathsf{n})=\frac{(2|\mathcal{X}||\mathcal{Y}|)^2}{\mathsf{n}^{1/2}}+\frac{2}{\mathsf{n}}.$$

QED

Given the channel, $W(\cdot|\cdot)$, and the rate $0 \leq R \leq \mathcal{C}$ one can optimize over time-sharing constant α, and the input letter input distribution pairs $(x_1, \mathsf{P}_{X,1})$ and $(x_2, \mathsf{P}_{X,2})$ to obtain the best missed detection exponent achievable for a given rate with the above architecture. As a result of Carathéodory's Theorem, we know that one need not to do time sharing between more than two input-letter-input-distribution pairs.

3.2.2 Error-Erasure Decoding

The codes described in Lemma 7 have large missed detection exponent for their first message; but their over all error exponent is zero. We append them with a control phase and allow erasures to give them a positive error exponent, like it was done in [43].

Lemma 8 *For any block length* n*, rate* $0 \leq R \leq \mathcal{C}$ *and error exponent* $0 \leq E \leq (1 - \frac{R}{\mathcal{C}})\mathcal{D}$*, there exists a block code with erasures such that,*

$$
\begin{aligned}
|\mathcal{M}| &\geq 1 + e^{\mathsf{n}(R - \epsilon_1(\mathsf{n}))} \\
P_{\mathbf{e}|m} &\leq (1 - \tfrac{E}{\mathcal{D}})^{-1} \epsilon_2(\mathsf{n}) \min\{1, e^{-\mathsf{n}(E - \epsilon_3(\mathsf{n}))}\} \qquad m = 1, 2, 3, \ldots, |\mathcal{M}| - 1 \\
P_{\mathbf{e}|0} &\leq e^{-\mathsf{n}(E + (1 - \frac{E}{\mathcal{D}})\mathcal{J}\left(\frac{R}{1 - E/\mathcal{D}}\right) + \epsilon_3(\mathsf{n}))} \\
P_{\mathbf{x}|m} &\leq 2(1 - \tfrac{E}{\mathcal{D}})^{-1} \epsilon_2(\mathsf{n})
\end{aligned}
$$

where $\epsilon_i(\mathsf{n}) \geq 0$ *and* $\lim_{\mathsf{n}\to\infty} \epsilon_i(\mathsf{n}) = 0$ *for* $i = 1, 2, 3$.

We use a two phase block code to achieve this performance. In the first phase a length n_1, rate $\frac{\mathsf{n}}{\mathsf{n}_1}R$ code with high missed detection exponent is used to convey the message, like the one described in Lemma 7. At the end of this phase a tentative decision is made. In the remaining $(\mathsf{n} - \mathsf{n}_1)$ time instances receiver either sends the accept letter x_a, if the tentative decision is correct, or sends the reject letter x_r, if the tentative decision is wrong, where accept and reject letters are the ones described in equation (3.1). At the end of the second phase an erasure is declared if the output sequence in the second phase is not typical with W_{x_a}, if it is typical with W_{x_a} tentative decision becomes the final.

Proof:
Note that we have assumed $\frac{R}{\mathcal{C}} \leq 1 - \frac{E}{\mathcal{D}}$. Consequently, if we let $\mathsf{n}_1 = \lceil (1 - \frac{E}{\mathcal{D}})\mathsf{n} \rceil$ we can conclude that $\frac{\mathsf{n}}{\mathsf{n}_1}R \leq \mathcal{C}$. Then as a result of Lemma 7 and the definition of $\mathcal{J}(\cdot)$ given in equation (3.8) there exists a length n_1 code such that,

$$|\mathcal{M}| - 1 \geq e^{\mathsf{n}_1[\frac{\mathsf{n}}{\mathsf{n}_1}R - \epsilon_1(\mathsf{n}_1)]} \tag{3.15a}$$

$$\tilde{P}_{\mathbf{e}|m} \leq \epsilon_3(\mathsf{n}_1) \qquad m = 1, 2, 3, \ldots, (|\mathcal{M}| - 1) \tag{3.15b}$$

$$\tilde{P}_{\mathbf{e}|0} \leq e^{-\mathsf{n}_1[\mathcal{J}\left(\frac{\mathsf{n}}{\mathsf{n}_1}R\right) - \epsilon_2(\mathsf{n}_1)]} \tag{3.15c}$$

We use this code for the first phase; in the second phase transmitter either accepts or rejects the tentative decision, $\tilde{\mathsf{M}}$, by using the input letters x_a or x_r for the remaining $(\mathsf{n} - \mathsf{n}_1)$ time units. Whenever the empirical distribution of the second phase is not typical with W_{x_a}, an erasure is declared. Thus decoding region for erasures is given by

$$\mathcal{G}(\mathbf{x}) = \{y^{\mathsf{n}}(\mathsf{n} - \mathsf{n}_1)\Delta_{(\mathsf{Q}_{(y^{\mathsf{n}}_{\mathsf{n}_1+1})}; W_{x_a})} \geq \mathsf{n}^{3/4}\}.$$

Then probability of erasure for correct tentative decision is upper bounded as,

$$\mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x} \middle| \tilde{\mathsf{M}} = m, \mathsf{M} = m\right] \leq 2|\mathcal{Y}|^2 \tfrac{\mathsf{n}-\mathsf{n}_1}{\mathsf{n}^{3/2}}$$
$$\leq 2|\mathcal{Y}|^2 \mathsf{n}^{-1/2} \qquad \forall m \in \mathcal{M}. \quad (3.16)$$

The probability of non-erasure decoding when tentative decision is incorrect is upper bounded as,

$$\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathbf{x} \middle| \tilde{\mathsf{M}} \neq m, \mathsf{M} = \mathsf{M}\right] \leq \min\{(\mathsf{n} - \mathsf{n}_1 + 1)^{|\mathcal{Y}|} e^{-(\mathsf{n}-\mathsf{n}_1)\mathcal{D} - \mathsf{n}^{3/4} \ln \lambda}, 1\}$$
$$\leq \min\{e^{-\mathsf{n}E + |\mathcal{Y}| \ln(\mathsf{n}+1) - \mathsf{n}^{3/4} \ln \lambda}, 1\} \qquad \forall m \in \mathcal{M}. \quad (3.17)$$

Lemma follows from the equations (3.15), (3.16), (3.17) and the following identities

$$\mathsf{n}_1 \epsilon_i(\mathsf{n}_1) \leq \mathsf{n}\epsilon_i(\mathsf{n}) \qquad i = 1, 2, 3 \qquad (3.18a)$$
$$P_{\mathbf{e}|m} \leq P_{\tilde{\mathbf{e}}|m} \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathbf{x} \middle| \tilde{\mathsf{M}} \neq m, \mathsf{M} = m\right] \qquad \forall m \in \mathcal{M} \qquad (3.18b)$$
$$P_{\mathbf{x}|m} = P_{\tilde{\mathbf{e}}|m} + \mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x} \middle| \tilde{\mathsf{M}} = m, \mathsf{M} = m\right] \qquad \forall m \in \mathcal{M}. \qquad (3.18c)$$

QED

3.2.3 Message-wise *UEP* with Single Special Message

We use a fixed-length block code with error-and-erasure decoding like the one described in Lemma 8 repetitively until a non-erasure decoding occurs. Resulting code is a variable-length block code and $\forall m \in \mathcal{M}$ the conditional values of the expected decoding time and the error probability are given by

$$\mathbf{E}[\tau | \mathsf{M} = m] = \tfrac{\mathsf{n}}{1 - P_{\mathbf{x}|m}} \qquad P_{\tilde{\mathbf{e}}|m}{}' = \tfrac{P_{\mathbf{e}|m}}{1 - P_{\mathbf{x}|m}} \qquad (3.19)$$

3.2.4 Bit-wise *UEP*

Lemma 9 *For any block length* n*, integer* k*, rate vector* $\vec{R}$*, time sharing vector* $\vec{\eta}$ *such that*

$$\tfrac{R_i}{\eta_i} \leq \mathcal{C} \quad \forall i \in \{1, 2, \ldots, k\} \qquad \text{and} \qquad \sum\nolimits_{i=1}^{k} \eta_i \leq 1 \qquad (3.20)$$

there exists a block code such that:

$$|\mathcal{M}_i| \geq e^{\mathsf{n}(R_i - \epsilon_4(\mathsf{n}))} \qquad \forall i \in \{1, 2, \ldots k\}$$
$$P_{\mathbf{x}|m} \leq \epsilon_5(\mathsf{n}) \qquad m \in \mathcal{M}$$
$$\mathbf{P}\left[\hat{\mathsf{M}}^i \notin \{\mathbf{x}, m^i\} \middle| \mathsf{M} = m\right] \leq \epsilon_5(\mathsf{n}) e^{-\mathsf{n}\left(\epsilon_6(\mathsf{n}) + \eta_{k+1}\mathcal{D} + \sum_{j=i+1}^{k} \eta_j \mathcal{J}\left(\frac{R_j}{\eta_j}\right)\right)} \qquad \forall m \in \mathcal{M}, \ i \in \{1, 2, \ldots k\}$$

where $\mathsf{M}^i = (\mathsf{M}_1, \mathsf{M}_2, \ldots \mathsf{M}_i)$*,* $\eta_{k+1} = 1 - \sum_{i=1}^{k} \eta_i$ *and* $\lim_{\mathsf{n}\to\infty} \epsilon_j(\mathsf{n}) = 0$ *for* $j = 4, 5, 6$*.*

Proof:
We use a $(k+1)$ phase block code to achieve this performance. In all of the phases a code with a special message, like the one described in Lemma 7, is used. Let us consider a

message set $\mathcal{M}$ of the form $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k$ as described in Section 3.1.3. Then each message of the message set $\mathcal{M}$ is of the form $\mathsf{M} = (\mathsf{M}_1, \mathsf{M}_2, \ldots, \mathsf{M}_k)$ where $\mathsf{M}_i \in \mathcal{M}_i$ $\forall i \in \{1, 2, \ldots, k\}$. In the first phase a length $\eta_1 \mathsf{n}$, rate $\frac{R_1}{\eta_1}$ code is used to convey M_1, i.e. $\bar{\mathsf{M}}_1 = \mathsf{M}_1$. At the end of this phase a tentative decision, $\tilde{\bar{\mathsf{M}}}_1$ is made about $\bar{\mathsf{M}}_1$. If $\tilde{\bar{\mathsf{M}}}_1 \neq \bar{\mathsf{M}}_1$ or if $\tilde{\bar{\mathsf{M}}}_1 = 0$ the special message is sent in the second phase, i.e. $\bar{\mathsf{M}}_2 = 0$. If $\tilde{\bar{\mathsf{M}}}_1 = \bar{\mathsf{M}}_1$ and $\tilde{\bar{\mathsf{M}}}_1 \neq 0$ then M_2 is sent in the second phase, i.e. $\bar{\mathsf{M}}_2 = \mathsf{M}_2$. The code in the second phase is a rate $\frac{R_2}{\eta_2}$ code of length $\eta_2 \mathsf{n}$. In 3^{rd}, 4^{th},…,k^{th} phases same signaling is repeated. In $(k+1)^{\text{th}}$ phase a length $(1 - \sum_{i=1}^{k} \eta_i)\mathsf{n}$ repetition code is used. If $\tilde{\bar{\mathsf{M}}}_k = \bar{\mathsf{M}}_k$ and $\tilde{\bar{\mathsf{M}}}_k \neq 0$ then $\bar{\mathsf{M}}_{k+1} = 1$ and the accept letter x_a is sent throughout the $(k+1)^{\text{th}}$ phase. If $\tilde{\bar{\mathsf{M}}}_k \neq \bar{\mathsf{M}}_k$ or $\tilde{\bar{\mathsf{M}}}_k = 0$ then $\bar{\mathsf{M}}_{k+1} = 0$ and the reject letter x_r is sent throughout $(k+1)^{\text{th}}$ phase. If the output distribution in the last phase is not typical with W_{x_a} an erasure is declared. Otherwise, the tentative decisions become the final, i.e $\hat{\mathsf{M}} = (\tilde{\bar{\mathsf{M}}}_1, \tilde{\bar{\mathsf{M}}}_2 \ldots, \tilde{\bar{\mathsf{M}}}_k)$. The decoding region for erasures, $\mathcal{G}(\mathbf{x})$, is given by

$$\mathcal{G}(\mathbf{x}) = \{y^{\mathsf{n}} : (\mathsf{n} - \mathsf{n}_1)\Delta_{(\mathsf{Q}_{(y^{\mathsf{n}}_{\mathsf{n}_1+1})}; W_{x_a})} \geq \mathsf{n}^{3/4}\}.$$

Then the probability of erasure, when all k tentative decisions are correct, is given by

$$\begin{aligned}\mathbf{P}\left[\hat{\mathsf{M}} = \mathbf{x} \middle| \bar{\mathsf{M}}_{k+1} = 1\right] &\leq 2|\mathcal{Y}|^2 \frac{\mathsf{n} - \sum_{j=1}^{k} \mathsf{n}_j}{\mathsf{n}^{3/2}} \\ &\leq 2|\mathcal{Y}|^2 \mathsf{n}^{-1/2}. \end{aligned} \tag{3.21}$$

The probability of not declaring an erasure when there is one or more incorrect tentative decisions in the first k phases is given by,

$$\begin{aligned}\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathbf{x} \middle| \bar{\mathsf{M}}_{k+1} = 0\right] &\leq (\mathsf{n} + 1 - \sum\nolimits_{j=1}^{k} \mathsf{n}_j)^{|\mathcal{Y}|} e^{-(\mathsf{n} - \sum_{j=1}^{k} \mathsf{n}_j)\mathcal{D} - \mathsf{n}^{3/4} \ln \lambda} \\ &\leq e^{-\mathsf{n}(\eta_{k+1}\mathcal{D} - \epsilon_3(\mathsf{n}))}. \end{aligned} \tag{3.22}$$

Furthermore, as a result of Lemma 7, we know that if $\frac{R_i}{\eta_i} \leq \mathcal{C}$, for large enough n_i there exists a code such that

$$|\bar{\mathcal{M}}_i| - 1 \geq e^{\mathsf{n}_i(\frac{R_i}{\eta_i} - \epsilon_1(\mathsf{n}_i))} \tag{3.23a}$$

$$\mathbf{P}\left[\tilde{\bar{\mathsf{M}}}_i \neq \bar{\mathsf{M}}_i \middle| \bar{\mathsf{M}}_i = m_i\right] \leq \epsilon_2(\mathsf{n}_i) \tag{3.23b}$$

$$\mathbf{P}\left[\tilde{\bar{\mathsf{M}}}_i \neq 0 \middle| \bar{\mathsf{M}}_i = 0\right] \leq e^{-\mathsf{n}_i(\mathcal{J}(R_i/\eta_i) - \epsilon_3(\mathsf{n}_i))}. \tag{3.23c}$$

In the first k phases, we use those $\mathsf{n}_i = \lfloor \eta_i \mathsf{n} \rfloor$ long codes with rate $\frac{R_i}{\eta_i}$. Then M_i is decoded incorrectly to a non-transmitted sub-message only when $\bar{\mathsf{M}}_i, \bar{\mathsf{M}}_{i+1}, \ldots, \bar{\mathsf{M}}_{k+1}$ are all decoded incorrectly:

$$\mathbf{P}\left[\hat{\mathsf{M}}^i \notin \{\mathbf{x}, m^i\} \middle| \mathsf{M} = m\right] \leq \mathbf{P}\left[\tilde{\bar{\mathsf{M}}}^i \neq \bar{\mathsf{M}}^i \middle| \mathsf{M} = m\right] \prod\nolimits_{j=i+1}^{k+1} \mathbf{P}\left[\tilde{\bar{\mathsf{M}}}_j \neq 0 \middle| \bar{\mathsf{M}}_j = 0\right] \tag{3.24}$$

An erasure, on the other hand, is declared only when one or more of the $\bar{\mathsf{M}}_1, \bar{\mathsf{M}}_2, \ldots, \bar{\mathsf{M}}_{k+1}$

are decoded incorrectly:

$$P_{\mathbf{x}|m} \leq \sum_{j=1}^{k+1} \mathbf{P}\left[\tilde{\bar{\mathsf{M}}}_j \neq \bar{\mathsf{M}}_j \middle| \bar{\mathsf{M}}_j \neq 0\right]. \tag{3.25}$$

Using equations (3.21), (3.22), (3.23), (3.24) and (3.25) we get

$$|\mathcal{M}_i| \geq e^{\mathsf{n}_i(\frac{R_i}{\eta_i} - \epsilon_1(\mathsf{n}_i))} \tag{3.26a}$$

$$P_{\mathbf{x}|m} \leq 2|\mathcal{Y}|^2 \mathsf{n}^{-1/2} + \sum_{j=1}^{k} \epsilon_2(\mathsf{n}_j) \tag{3.26b}$$

$$\mathbf{P}\left[\hat{\mathsf{M}}^i \notin \{\mathbf{x}, m^i\} \middle| \mathsf{M}=m\right] \leq [\sum_{j=1}^{i} \epsilon_2(\mathsf{n}_j)] e^{-\mathsf{n}(\eta_{k+1}\mathcal{D} - \epsilon_3(\mathsf{n}))} \prod_{j=i+1}^{k} e^{-\mathsf{n}_j(\mathcal{J}(R_j/\eta_j) - \epsilon_3(\mathsf{n}_j))}. \tag{3.26c}$$

Then the lemma follows from $\mathsf{n}_j \epsilon_i(\mathsf{n}_j) \leq \mathsf{n}\epsilon_i(\mathsf{n})$ for $i = 1, 2, 3$ and $j = 1, 2, \ldots, k$ for

$$\epsilon_4(\mathsf{n}) = \epsilon_1(\mathsf{n}) + \frac{\mathcal{C}}{\mathsf{n}} \qquad \epsilon_5(\mathsf{n}) = 2(\sum_{j=1}^{k} \frac{1}{\eta_j})\epsilon_2(\mathsf{n}) + 2|\mathcal{Y}|^2 \mathsf{n}^{-1/2} \qquad \epsilon_6(\mathsf{n}) = k\epsilon_3(\mathsf{n}).$$

QED

Note that like we did in the single special message case we can use the fixed-length block code with error-and-erasure decoding repetitively until a non-erasure decoding occurs. Then $\forall m \in \mathcal{M}$ the resulting variable-length block code will satisfy

$$\mathbf{E}[\tau|\,\mathsf{M} = m] = \frac{\mathsf{n}}{1 - P_{\mathbf{x}|m}} \qquad \mathbf{P}\left[\hat{\mathsf{M}}^i \neq m^i \middle| \mathsf{M} = m\right] \leq \frac{\mathbf{P}\left[\hat{\mathsf{M}}^i \notin \{\mathbf{x}, m^i\} \middle| \mathsf{M}=m\right]}{1 - P_{\mathbf{x}|m}}. \tag{3.27}$$

For a given rate vector $\vec{R}$ and time sharing vector $\vec{\eta}$, Lemma 9 gives us an achievable error exponent vector $\vec{E}$. In other words a $(\vec{R}, \vec{E})$ pair is achievable if there exists a time sharing vector $\vec{\eta}$ such that,

$$\frac{R_i}{\eta_i} \leq \mathcal{C} \qquad E_i \leq (1 - \sum_{j=1}^{k} \eta_j)\mathcal{D} + \sum_{j=i+1}^{k} \eta_j \mathcal{J}\left(\frac{R_j}{\eta_j}\right).$$

Thus the existence of the time sharing vector $\vec{\eta}$ is a sufficient condition for the achievablity of a $(\vec{R}, \vec{E})$ pair. We show in the following section that the existence of such a time sharing vector $\vec{\eta}$ is also a necessary condition for the achievablity of a $(\vec{R}, \vec{E})$ pair.

3.3 Converse

Berlin et. al. [2] used the error probability of a binary query posed at a stopping time in bounding the error probability of variable-length block codes. Later similar techniques have been applied in [3] for establishing outer bounds in *UEP* problems. Our approach is similar to that of [2] and [3] in using error probabilities associated with appropriately chosen queries for establishing outer bounds. The novelty of our approach is in the error events we choose to consider in our analysis.

We will start with lower bounding the expected values of certain conditional error probabilities associated with these queries, in terms of the rate of decrease of the conditional entropy of the messages in different intervals. We then use this bound, Lemma 10, to derive outer bounds for the single message *message-wise UEP* problem and the multi-layer *bit-wise UEP* problem.

3.3.1 Missed Detection Probability and Decay Rate of Entropy

Like similar lemmas in [2] and [3], Lemma 10 bounds the error probability of a query posed at a stopping time. However, instead of working with expected value of the error probability of the query and making a worst case assumption on the rate of decrease of entropy; Lemma 10 works with the expected value of a certain conditional error probability and bounds it in terms of the rate of decrease of the conditional entropy in different intervals. Above mentioned conditional error probability, is best described using the probability distribution defined below.

Definition 8 *For a stopping time, $\bar{\tau}$, smaller than or equal to τ and a subset of $\mathcal{M}$, $\mathcal{A}_{\bar{\tau}}$, determined by $\mathsf{Y}^{\bar{\tau}}$ with positive conditional probability*[5] *we define the probability distribution $\mathbf{P}_{\{\mathcal{A}_{\bar{\tau}}\}}[\cdot]$ on $\mathcal{M} \times \mathcal{Y}^{\tau}$ as follows:*

$$\mathbf{P}_{\{\mathcal{A}_{\bar{\tau}}\}}[m, y^{\tau}] \triangleq \mathbf{P}\left[y^{\bar{\tau}}\right] \frac{\mathbf{P}[m|y^{\bar{\tau}}]\mathbb{1}_{\{m \in \mathcal{A}_{\bar{\tau}}\}}}{\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\bar{\tau}}|y^{\bar{\tau}}]} \mathbf{P}\left[y_{\bar{\tau}+1}^{\tau}\middle| y^{\bar{\tau}}, m\right] \tag{3.28}$$

where $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\bar{\tau}}|\, y^{\bar{\tau}}] = \sum_{m \in \mathcal{A}_{\bar{\tau}}} \mathbf{P}[m|\, y^{\bar{\tau}}]$.

Probability distributions $\mathbf{P}_{\{\mathcal{A}_{\bar{\tau}}\}}[\cdot]$ and $\mathbf{P}[\cdot]$ have the same marginal distribution on $\mathcal{Y}^{\bar{\tau}}$ and the same conditional distribution on $\mathcal{Y}_{\bar{\tau}+1}^{\tau}$ given $(\mathsf{M}, \mathsf{Y}^{\bar{\tau}})$. For a given $y^{\bar{\tau}}$ on the other hand $\mathbf{P}_{\{\mathcal{A}_{\bar{\tau}}\}}[m|\, y^{\bar{\tau}}]$ is $\frac{\mathbb{1}_{\{m \in \mathcal{A}_{\bar{\tau}}\}}}{\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\bar{\tau}}|y^{\bar{\tau}}]}$ times $\mathbf{P}[m|\, y^{\bar{\tau}}]$. Thus $\mathbf{P}_{\{\mathcal{A}_{\bar{\tau}}\}}[\mathsf{M} = \cdot|\, y^{\bar{\tau}}]$ has probability mass only on those messages in $\mathcal{A}_{\bar{\tau}}$.

Lemma 10 *For any variable-length block code with decoding time τ, k stopping times such that $0 \le \tau_1 \le \tau_2 \le \cdots \le \tau_k \le \tau$ and k subsets of $\mathcal{M}$, $\mathcal{A}_{\tau_1}, \mathcal{A}_{\tau_2}, \ldots, \mathcal{A}_{\tau_k}$, each measurable in the corresponding $\sigma(\mathsf{Y}^{\tau_i})$ and each with positive probability,*[6]

$$(1 - P_{\mathbf{e}} - \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}]) \ln \frac{1}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}]} \le \ln 2 + \sum_{j=i}^{k} \mathbf{E}[\tau_{j+1} - \tau_j] \mathcal{J}\left(\frac{\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_j}) - \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_{j+1}})]}{\mathbf{E}[\tau_{j+1} - \tau_j]}\right) \tag{3.29}$$

where the probability distribution $\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\cdot]$ is defined in (3.28).

The bound on $\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right]$ depends only on $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}]$ and the rate of decrease of conditional entropy in the intervals $(\tau_j, \tau_{j+1}]$ for $j \ge i$. The particular choice of $\mathcal{A}_{\tau_j}$ for $j \ne i$ has no effect on the bound on $\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right]$. This property of the bound is its main merit over bounds resulting from the previously suggested techniques.

Proof:
As a result of data processing inequality for Kullback-Leibler divergence, we have

$$\mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}^{\tau}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\mathsf{Y}^{\tau}]}\right] \ge \mathbf{P}\left[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i}\right] \ln \frac{\mathbf{P}[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i}]} + \mathbf{P}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right] \ln \frac{\mathbf{P}[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}]}.$$

Since $h(x) = x \ln \frac{1}{x} - (1-x) \ln \frac{1}{1-x} \le \ln 2 \ \forall x \in [0,1]$ and $0 \le \mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i}\right] \le 1$, we have

$$\mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}^{\tau}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\mathsf{Y}^{\tau}]}\right] \ge -\ln 2 + \mathbf{P}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right] \ln \frac{1}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}]}. \tag{3.30}$$

[5] In other words $\bar{\tau} \le \tau$ and $\mathbf{P}\left[\mathsf{M} \in \mathcal{A}_{\bar{\tau}}|\, \mathsf{Y}^{\bar{\tau}}\right] > 0$ with probability one.
[6] $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}|\, \mathsf{Y}^{\tau_i}] > 0$ with probability one for $i = 1, 2, \ldots, k$.

In addition,

$$\begin{aligned}\mathbf{P}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right] &= 1 - \mathbf{P}\left[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i}\right] \\ &= 1 - \mathbf{P}\left[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i} \middle| \mathsf{M} \notin \mathcal{A}_{\tau_i}\right] \mathbf{P}[\mathsf{M} \notin \mathcal{A}_{\tau_i}] - \mathbf{P}\left[\hat{\mathsf{M}} \in \mathcal{A}_{\tau_i} \middle| \mathsf{M} \in \mathcal{A}_{\tau_i}\right] \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}] \\ &\geq 1 - P_{\mathbf{e}} - \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}]. \end{aligned} \tag{3.31}$$

Thus using equations (3.30) and (3.31) we get,

$$(1 - P_{\mathbf{e}} - \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}]) \ln \frac{1}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right]} \leq \ln 2 + \mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}^{\tau}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\mathsf{Y}^{\tau}]}\right]. \tag{3.32}$$

Assume for the moment that,

$$\mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}^{\tau}]}{\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[\mathsf{Y}^{\tau}]}\right] \leq \sum_{j=i}^{k} \mathbf{E}[\tau_{j+1} - \tau_j] \mathcal{J}\left(\frac{\mathbf{E}\left[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_j}) - \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_{j+1}})\right]}{\mathbf{E}[\tau_{j+1} - \tau_j]}\right). \tag{3.33}$$

Equation (3.29), i.e. Lemma 10, follows from equation (3.32) and (3.33).

Above, we have proved the lemma by assuming equation (3.33) is valid for all i in $\{1, 2, \ldots, k\}$. We now prove equation (3.33) for $i = 1$, which implies the validity of equation (3.33) for all $i \in \{1, 2, \ldots, k\}$. Let us consider the stochastic sequence

$$S_t = \left[-\ln \frac{\mathbf{P}\left[\mathsf{Y}_{\tau_1+1}^{t} \middle| \mathsf{Y}^{\tau_1}\right]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}\left[\mathsf{Y}_{\tau_1+1}^{t} \middle| \mathsf{Y}^{\tau_1}\right]} + \sum_{j=\tau_1+1}^{t} \mathcal{J}\left(\mathcal{I}\left(\mathsf{M}; \mathsf{Y}_j \middle| \mathsf{Y}^{j-1}\right)\right)\right] \mathbb{1}_{\{\tau_1 < t\}} \tag{3.34}$$

where $\mathcal{I}\left(\mathsf{M}; \mathsf{Y}_t \middle| \mathsf{Y}^{t-1}\right) = \mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}_t|\mathsf{M}, \mathsf{Y}^{t-1}]}{\mathbf{P}[\mathsf{Y}_t|\mathsf{Y}^{t-1}]} \middle| \mathsf{Y}^{t-1}\right]$. Then

$$S_{t+1} - S_t = \left(\mathcal{J}\left(\mathcal{I}\left(\mathsf{M}; \mathsf{Y}_{t+1} \middle| \mathsf{Y}^{t}\right)\right) - \ln \frac{\mathbf{P}[\mathsf{Y}_{t+1}|\mathsf{Y}^t]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}[\mathsf{Y}_{t+1}|\mathsf{Y}^t]}\right) \mathbb{1}_{\{\tau_1 \leq t\}}. \tag{3.35}$$

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

$$\mathcal{J}\left(\mathcal{I}\left(\mathsf{M}; \mathsf{Y}_{t+1} \middle| \mathsf{Y}^t\right)\right) \geq \mathcal{J}\left(\mathcal{I}\left(\mathsf{X}_{t+1}; \mathsf{Y}_{t+1} \middle| \mathsf{Y}^t\right)\right). \tag{3.36}$$

Furthermore because of the definition of $\mathcal{J}(\cdot)$ given in (3.8), the definition of $\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}[\cdot]$ given in (3.28) and the convexity of Kullback Leibler divergence we have

$$\mathcal{J}\left(\mathcal{I}\left(\mathsf{X}_{t+1}; \mathsf{Y}_{t+1} \middle| \mathsf{Y}^t\right)\right) \geq \mathbf{E}\left[\ln \frac{\mathbf{P}[\mathsf{Y}_{t+1}|\mathsf{Y}^t]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}[\mathsf{Y}_{t+1}|\mathsf{Y}^t]} \middle| \mathsf{Y}^t\right]. \tag{3.37}$$

Thus as a result of equations (3.35), (3.36) and (3.37) we have

$$\mathbf{E}\left[S_{t+1} \middle| \mathsf{Y}^t\right] \geq S_t. \tag{3.38}$$

Recall that $\min_{x,y} W(y|x) = \lambda$ and $|\mathcal{J}(\cdot)| \leq \mathcal{D}$. Hence as a result of equation (3.35)

$$\mathbf{E}\left[|S_{t+1} - S_t| \middle| \mathsf{Y}^t\right] \leq \ln \tfrac{1}{\lambda} + \mathcal{D}. \tag{3.39}$$

As a result of (3.38), (3.39) and the fact that $S_0 = 0$, S_t is a submartingale. Because of (3.39) we can apply a version of Doob's optional stopping theorem [40, Theorem 2,p 487] for stopping times τ_1 and τ_2 such that $\mathbf{E}[\tau_1] \leq \mathbf{E}[\tau_2] < \infty$ and get $\mathbf{E}[S_{\tau_2}] \geq \mathbf{E}[S_{\tau_1}] = 0$:

$$\mathbf{E}\left[\ln \frac{\mathbf{P}\left[\mathsf{Y}_{\tau_1+1}^{\tau_2}\middle|\mathsf{Y}^{\tau_1}\right]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}\left[\mathsf{Y}_{\tau_1+1}^{\tau_2}\middle|\mathsf{Y}^{\tau_1}\right]}\right] \leq \mathbf{E}\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{J}\left(\mathcal{I}\left(\mathsf{M};\mathsf{Y}_t\left|\mathsf{Y}^{t-1}\right.\right)\right)\right]. \tag{3.40}$$

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

$$\begin{aligned}\mathbf{E}\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{J}\left(\mathcal{I}\left(\mathsf{M};\mathsf{Y}_t\left|\mathsf{Y}^{t-1}\right.\right)\right)\right] &= \mathbf{E}[\tau_2-\tau_1]\,\mathbf{E}\left[\sum_{t\geq 1} \frac{\mathbb{1}_{\{\tau_2\geq t>\tau_1\}}\mathcal{J}\left(\mathcal{I}\left(\mathsf{M};\mathsf{Y}_t\left|\mathsf{Y}^{t-1}\right.\right)\right)}{\mathbf{E}[\tau_2-\tau_1]}\right]\\ &\leq \mathbf{E}[\tau_2-\tau_1]\,\mathcal{J}\left(\frac{\mathbf{E}\left[\sum_{t=\tau_1+1}^{\tau_2}\mathcal{I}\left(\mathsf{M};\mathsf{Y}_t\left|\mathsf{Y}^{t-1}\right.\right)\right]}{\mathbf{E}[\tau_2-\tau_1]}\right).\end{aligned} \tag{3.41}$$

In order to lower bound the sum within $\mathcal{J}(\cdot)$ in (3.41) consider the stochastic sequence

$$V_t = \mathcal{H}(\mathsf{M}|\mathsf{Y}^t) + \sum_{j=1}^{t} \mathcal{I}\left(\mathsf{M};\mathsf{Y}_j\left|\mathsf{Y}^{j-1}\right.\right). \tag{3.42}$$

Clearly $\mathbf{E}\left[V_{t+1}\middle|\,\mathsf{Y}^t\right] = V_t$ and $\mathbf{E}[|V_t|] < \infty$, thus V_t is a martingale. Furthermore, $\mathbf{E}\left[|V_{t+1}-V_t|\middle|\,\mathsf{Y}^t\right] < \infty$ and $\mathbf{E}[\tau_1] \leq \mathbf{E}[\tau_2] < \infty$, thus using Doob's optimal stopping theorem, [40, Theorem 2, p 487] we get $\mathbf{E}[V_{\tau_2}] = \mathbf{E}[V_{\tau_1}]$:

$$\mathbf{E}\left[\sum_{t=\tau_1+1}^{\tau_2} \mathcal{I}\left(\mathsf{M};\mathsf{Y}_t\left|\mathsf{Y}^{t-1}\right.\right)\right] = \mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_1}) - \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2})]. \tag{3.43}$$

Using equations (3.40), (3.41) and (3.43)

$$\mathbf{E}\left[\ln \frac{\mathbf{P}\left[\mathsf{Y}_{\tau_1+1}^{\tau}\middle|\mathsf{Y}^{\tau_1}\right]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}\left[\mathsf{Y}_{\tau_1+1}^{\tau}\middle|\mathsf{Y}^{\tau_1}\right]}\right] \leq \mathbf{E}[\tau_2-\tau_1]\,\mathcal{J}\left(\frac{\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_1})-\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2})]}{\mathbf{E}[\tau_2-\tau_1]}\right) + \mathbf{E}\left[\ln \frac{\mathbf{P}\left[\mathsf{Y}_{\tau_2+1}^{\tau}\middle|\mathsf{Y}^{\tau_2}\right]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}\left[\mathsf{Y}_{\tau_2+1}^{\tau}\middle|\mathsf{Y}^{\tau_2}\right]}\right] \tag{3.44}$$

Repeating the arguments (3.40) through (3.44) for $(\tau_j, \tau_{j+1}]$ for $j = 2, 3, \ldots, k$ we get,

$$\mathbf{E}\left[\ln \frac{\mathbf{P}\left[\mathsf{Y}_{\tau_1+1}^{\tau}\middle|\mathsf{Y}^{\tau_1}\right]}{\mathbf{P}_{\{\mathcal{A}_{\tau_1}\}}\left[\mathsf{Y}_{\tau_1+1}^{\tau}\middle|\mathsf{Y}^{\tau_1}\right]}\right] \leq \sum_{j=1}^{k} \mathbf{E}[\tau_{j+1}-\tau_j]\,\mathcal{J}\left(\frac{\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_j})-\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_{j+1}})]}{\mathbf{E}[\tau_{j+1}-\tau_j]}\right). \tag{3.45}$$

QED

3.3.2 Single Special Message

We first derive a lower bound on the minimum conditional error probability of a message in a variable-length block code, using Lemma 10. This derivation introduces some of the ideas used for the outer bound for the multi-layer *bit-wise UEP* problem.

Lemma 11 *For any variable-length block code with feedback with $|\mathcal{M}| = e^{\mathbf{E}[\tau]R}$ messages and average error probability $P_{\mathbf{e}} = e^{-\mathbf{E}[\tau]E}$, and for any $\delta \in (0, 0.5)$*

$$-\frac{\ln \mathbf{P}\left[\hat{\mathsf{M}}\neq m\middle|\mathsf{M}=m\right]}{\mathbf{E}[\tau]} \leq E + \left(1-\frac{E-\tilde{\epsilon}}{\mathcal{D}}\right)\mathcal{J}\left(\frac{R-\tilde{\epsilon}}{1-\frac{E-\tilde{\epsilon}}{\mathcal{D}}}\right) \qquad \forall m \in \mathcal{M} \tag{3.46}$$

where $\tilde{\epsilon} = \frac{\tilde{\epsilon}_1\mathcal{D}+\tilde{\epsilon}_2}{1-\tilde{\epsilon}_1}$, $\tilde{\epsilon}_1 = P_{\mathbf{e}} + \delta + \frac{P_{\mathbf{e}}}{\delta} + |\mathcal{M}|^{-1}$ *and* $\tilde{\epsilon}_2 = \frac{\ln 2 - \ln \lambda\delta}{\mathbf{E}[\tau]}$

Lemma 11 is a generalization of [3, Theorem 8], which is tight not only for small but all positive values of E. Furthermore unlike the proof for [3, Theorem 8] which needs previous results like [2, Lemma 1], our proof is self sufficient.

Proof:
Let $\tau_1 \triangleq 0$, $\mathcal{A}_{\tau_1} \triangleq \{m\}$ and τ_2 be the first time instance before τ such that one message has a posteriori probability $1-\delta$ or higher:

$$\tau_2 \triangleq \min\{t : \max_{\tilde{m}} \mathbf{P}\left[\mathsf{M} = \tilde{m} \middle| \mathsf{Y}^t\right] \geq 1-\delta \text{ or } t = \tau\}.$$

Let $\mathcal{A}_{\tau_2}$ be the set of all messages except the most likely one, if the most likely one has a probability $(1-\delta)$ or higher; $\mathcal{M}$ otherwise:

$$\mathcal{A}_{\tau_2} \triangleq \{\tilde{m} \in \mathcal{M} : \mathbf{P}[\mathsf{M} = \tilde{m} | \mathsf{Y}^{\tau_2}] < (1-\delta)\}.$$

Since $\min_{x\in\mathcal{X}, y\in\mathcal{Y}} W(y|x) = \lambda$ the posterior probability of a message can not decrease by a (multiplicative) factor less than λ in one time instance. Hence $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \mathsf{Y}^{\tau_2}] \geq \lambda\delta$ for all values of Y^{τ_2} and as a result of the definition of $\mathbf{P}_{\{\mathcal{A}_{\tilde{\tau}}\}}[m, y^\tau]$ given in equation (3.28) we have,

$$\mathbf{P}_{\{\mathcal{A}_{\tau_2}\}}[m, y^\tau] \leq \mathbf{P}[m, y^\tau] \tfrac{\mathbb{1}_{\{m \in \mathcal{A}_{\tau_2}\}}}{\lambda\delta}. \tag{3.47}$$

Evidently if $\mathsf{M} \in \mathcal{A}_{\tau_2}$ and $\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_2}$ then $\hat{\mathsf{M}} \neq \mathsf{M}$:

$$\mathbb{1}_{\{\mathsf{M}\in\mathcal{A}_{\tau_2}\}} \mathbb{1}_{\{\hat{\mathsf{M}}\notin\mathcal{A}_{\tau_2}\}} \leq \mathbb{1}_{\{\hat{\mathsf{M}}\neq\mathsf{M}\}}. \tag{3.48}$$

Using equations (3.47) and (3.48) we get

$$\mathbf{P}_{\{\mathcal{A}_{\tau_2}\}}[m, y^\tau] \mathbb{1}_{\{\hat{\mathsf{M}}\notin\mathcal{A}_{\tau_2}\}} \leq \mathbf{P}[m, y^\tau] \tfrac{\mathbb{1}_{\{\hat{\mathsf{M}}\neq m\}}}{\lambda\delta}.$$

If we sum over $\mathcal{M} \times \mathcal{Y}^\tau$ we get

$$\mathbf{P}_{\{\mathcal{A}_{\tau_2}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_2}\right] \leq \tfrac{P_{\mathbf{e}}}{\lambda\delta}. \tag{3.49}$$

Using Lemma 10, equation (3.49) and the fact that $\mathcal{J}(R) \leq \mathcal{D}$ we get

$$\tfrac{1-P_{\mathbf{e}}-|\mathcal{M}|^{-1}}{\mathbf{E}[\tau]} \ln \tfrac{1}{\mathbf{P}\left[\hat{\mathsf{M}}\neq m \middle| \mathsf{M}=m\right]} \leq \ln 2 + \eta \mathcal{J}\left(\tfrac{\ln|\mathcal{M}| - \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2})}{\eta\mathbf{E}[\tau]}\right) + (1-\eta)\mathcal{D} \tag{3.50a}$$

$$\tfrac{1-P_{\mathbf{e}}-\mathbf{P}\left[\mathsf{M}\in\mathcal{A}_{\tau_2}\right]}{\mathbf{E}[\tau]} \ln \tfrac{\lambda\delta}{P_{\mathbf{e}}} \leq \ln 2 + (1-\eta)\mathcal{D} \tag{3.50b}$$

where $\eta = \frac{\mathbf{E}[\tau_2]}{\mathbf{E}[\tau]}$.
Now we bound $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2}]$ and $\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2})$ from above. As a result of Bayes' rule

$$\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] = \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathcal{A}_{\tau_2} = \mathcal{M}\right] \mathbf{P}[\mathcal{A}_{\tau_2} = \mathcal{M}] + \mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathcal{A}_{\tau_2} \neq \mathcal{M}\right] \mathbf{P}[\mathcal{A}_{\tau_2} \neq \mathcal{M}].$$

Since $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M}\right] = P_{\mathbf{e}}$ and $\mathbf{P}\left[\hat{\mathsf{M}} \neq \mathsf{M} \middle| \mathcal{A}_{\tau_2} = \mathcal{M}\right] \geq \delta$ we have

$$P_{\mathbf{e}} \geq \delta \mathbf{P}[\mathcal{A}_{\tau_2} = \mathcal{M}]. \tag{3.51}$$

Using the fact that $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathcal{A}_{\tau_2} \neq \mathcal{M}] \leq \delta$ together with equation (3.51) and Bayes' rule we get

$$\begin{aligned}\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2}] &= \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathcal{A}_{\tau_2} \neq \mathcal{M}] \, \mathbf{P}[\mathcal{A}_{\tau_2} \neq \mathcal{M}] + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathcal{A}_{\tau_2} = \mathcal{M}] \, \mathbf{P}[\mathcal{A}_{\tau_2} = \mathcal{M}] \\ &\leq \delta + \tfrac{P_{\mathbf{e}}}{\delta}. \end{aligned} \tag{3.52}$$

Note that $\mathcal{A}_{\tau_2}$ has at most $|\mathcal{M}|$ elements and its complement, $\overline{\mathcal{A}_{\tau_2}}$, has at most one element thus,

$$\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2}, \mathcal{A}_{\tau_2}) \leq \ln |\mathcal{M}| \qquad\qquad \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2}, \overline{\mathcal{A}_{\tau_2}}) = 0 \tag{3.53}$$

Furthermore $h(x) = -x \ln x - (1-x) \ln(1-x) \leq \ln 2$ for all $x \in [0,1]$. Hence from equation (3.53) we get

$$\begin{aligned}\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2}) &= h(\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathsf{Y}^{\tau_2}]) + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathsf{Y}^{\tau_2}] \, \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2}, \mathcal{A}_{\tau_2}) + \mathbf{P}[\mathsf{M} \notin \mathcal{A}_{\tau_2} | \, \mathsf{Y}^{\tau_2}] \, \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2}, \overline{\mathcal{A}_{\tau_2}}) \\ &\leq \ln 2 + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_2} | \, \mathsf{Y}^{\tau_2}] \ln |\mathcal{M}|. \end{aligned} \tag{3.54}$$

Using equations (3.52) and (3.54) we get

$$\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_2})] \leq \ln 2 + (\delta + \tfrac{P_{\mathbf{e}}}{\delta}) \ln |\mathcal{M}|. \tag{3.55}$$

Using equations (3.50), (3.52) and (3.55) we get

$$\tfrac{1-\tilde{\epsilon}_1}{\mathbf{E}[\tau]} \ln \tfrac{1}{\mathbf{P}\left[\hat{\mathsf{M}} \neq m \middle| \mathsf{M} = m\right]} \leq \eta \mathcal{J}\left(\tfrac{(1-\tilde{\epsilon}_1)R - \tilde{\epsilon}_2}{\eta}\right) + (1-\eta)\mathcal{D} + \tilde{\epsilon}_2 \tag{3.56a}$$

$$\tfrac{1-\tilde{\epsilon}_1}{\mathbf{E}[\tau]} \ln \tfrac{1}{P_{\mathbf{e}}} \leq (1-\eta)\mathcal{D} + \tilde{\epsilon}_2 \tag{3.56b}$$

where $\eta = \frac{\mathbf{E}[\tau_2]}{\mathbf{E}[\tau]}$.
Since $\mathcal{J}(R)$ is a concave function, it lies below its tangents, i.e. $\forall R \in [0, C]$ and $\eta \leq \frac{R}{C}$

$$\begin{aligned}\frac{d}{d\eta}\left(\eta \mathcal{J}\left(\tfrac{R}{\eta}\right) + (1-\eta)\mathcal{D}\right) &= \mathcal{J}\left(\tfrac{R}{\eta}\right) - \tfrac{R}{\eta}\mathcal{J}'\left(\tfrac{R}{\eta}\right) - \mathcal{D} \\ &\geq 0. \end{aligned} \tag{3.57}$$

Thus the bound in equation (3.56a) has its maximum value for the maximum value of η. Furthermore equation (3.56b) gives an upper bound on η. These two observations leads to Lemma 11.

QED

3.3.3 Special Bits

In this section we prove an outer bound for achievable rate and error exponent vectors in a k-level *bit-wise UEP* code with finite $\mathbf{E}[\tau]$, i.e. we derive a necessary condition that is satisfied by all achievable $(\vec{R}, \vec{E}, \mathbf{E}[\tau])$ triples. This leads to an outer bound to the achievable points on rate error exponent vectors, which matches the inner bound described in Section 3.2.4 as $\mathbf{E}[\tau]$ increases for all reliable sequences.

Lemma 12 *For any* $\mathbf{E}[\tau]$, k, $\delta \in (0, 0.5)$, *all achievable rate vector error exponent vector pairs,* $(\vec{R}, \vec{E})$ *satisfy*

$$(1-\tilde{\epsilon}_3)E_i - \tilde{\epsilon}_4 \leq (1-\sum\nolimits_{j=1}^{k}\eta_j)\mathcal{D} + \sum\nolimits_{j=i+1}^{k}\eta_j \mathcal{J}\left(\tfrac{(1-\tilde{\epsilon}_3)R_j}{\eta_j}\right) \qquad i = 1, 2, \ldots, k \tag{3.58a}$$

$$\tfrac{(1-\tilde{\epsilon}_3)R_j - \tilde{\epsilon}_4 \mathbb{1}_{\{j=1\}}}{\eta_j} \leq \mathcal{C} \qquad i = 1, 2, \ldots, k \tag{3.58b}$$

$$\sum\nolimits_{j=1}^{k}\eta_j \leq 1 \tag{3.58c}$$

for some $\vec{\eta}$ *where* $\tilde{\epsilon}_3 = \delta + \frac{P_{\mathbf{e}}}{\delta}$ *and* $\tilde{\epsilon}_4 = \frac{\ln 2k - \ln \lambda\delta}{\mathbf{E}[\tau]}$.

Proof:
Let $\mathcal{M}^i$ be $\mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_i$. If a member of $\mathcal{M}^i$ gains a posterior probability larger than or equal to $1-\delta$ at or before τ, τ_i is the first times instance it happens and $\mathcal{A}_{\tau_i}$ is the set of messages of the form $(m^i, m_{i+1}, \ldots, m_k)$ where $m^i \neq \underset{\tilde{m}^i}{\operatorname{argmax}} \mathbf{P}\left[\tilde{m}^i \middle| \mathsf{Y}^{\tau_i}\right]$; else $\tau_i = \tau$ and $\mathcal{A}_{\tau_i} = \mathcal{M}$:

$$\tau_i \triangleq \min\{t : \max_{m^i} \mathbf{P}\left[\mathsf{M}^i = m^i \middle| \mathsf{Y}^t\right] \geq 1-\delta \text{ or } t = \tau\} \tag{3.59}$$

$$\mathcal{A}_{\tau_i} \triangleq \{(m^i, m_{i+1}, \ldots, m_k) \in \mathcal{M} : m^i \in \mathcal{M}^i \text{ and } \mathbf{P}\left[\mathsf{M}^i = m^i \middle| \mathsf{Y}^{\tau_i}\right] < 1-\delta\}. \tag{3.60}$$

Recall that $\min_{x\in\mathcal{X}, y\in\mathcal{Y}} W(y|x) = \lambda$, thus the posterior probability of a $m^i \in \mathcal{M}^i$ can not decrease by a (multiplicative) factor less than λ in one time instance. Hence as a result of equations (3.59) and (3.60), $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \mathsf{Y}^{\tau_i}] \geq \lambda\delta$ for all values of Y^{τ_i}. Then as a result of the definition of $\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[m, y^\tau]$ given in equation (3.28) we have,

$$\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}[m, y^\tau] \leq \mathbf{P}[m, y^\tau] \, \tfrac{\mathbb{1}_{\{m\in\mathcal{A}_{\tau_i}\}}}{\lambda\delta}. \tag{3.61}$$

As a result of the definition of $\mathcal{A}_{\tau_i}$ given in equation (3.60) we have,

$$\mathbb{1}_{\left\{\mathsf{M}\in\mathcal{A}_{\tau_i}\right\}} \mathbb{1}_{\left\{\hat{\mathsf{M}}\notin\mathcal{A}_{\tau_i}\right\}} \leq \mathbb{1}_{\left\{\hat{\mathsf{M}}^i\neq\mathsf{M}^i\right\}} \tag{3.62}$$

From equations (3.61) and (3.62) we get

$$\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right] \leq \tfrac{\mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i\right]}{\lambda\delta}. \tag{3.63}$$

Recall that we have assumed $P_{\mathbf{e},1} \leq P_{\mathbf{e},2} \leq \ldots \leq P_{\mathbf{e},k}$. Thus

$$\mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i\right] \leq \sum\nolimits_{j\leq i} \mathbf{P}\left[\hat{\mathsf{M}}_j \neq \mathsf{M}_j\right] \leq i P_{\mathbf{e},i}. \tag{3.64}$$

As a result of equations (3.63), (3.64) and the fact that $i \leq k$ we have

$$\mathbf{P}_{\{\mathcal{A}_{\tau_i}\}}\left[\hat{\mathsf{M}} \notin \mathcal{A}_{\tau_i}\right] \leq \tfrac{k}{\lambda\delta} P_{\mathbf{e},i}. \tag{3.65}$$

Using Bayes' rule we get

$$\mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i\right] = \mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i \middle| \mathcal{A}_{\tau_i} = \mathcal{M}\right] \mathbf{P}[\mathcal{A}_{\tau_i} = \mathcal{M}] + \mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i \middle| \mathcal{A}_{\tau_i} \neq \mathcal{M}\right] \mathbf{P}[\mathcal{A}_{\tau_i} \neq \mathcal{M}].$$

Plugging in the inequalities $P_{\mathbf{e}} \geq \mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i\right]$ and $\mathbf{P}\left[\hat{\mathsf{M}}^i \neq \mathsf{M}^i \middle| \mathcal{A}_{\tau_i} = \mathcal{M}\right] \geq \delta$ we get

$$P_{\mathbf{e}} \geq \delta \mathbf{P}[\mathcal{A}_{\tau_i} = \mathcal{M}]\,. \tag{3.66}$$

In addition using Bayes' rule together with equation (3.66) and $\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathcal{A}_{\tau_i} \neq \mathcal{M}] \leq \delta$ we get

$$\begin{aligned} \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}] &= \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathcal{A}_{\tau_i} = \mathcal{M}] \, \mathbf{P}[\mathcal{A}_{\tau_i} = \mathcal{M}] + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathcal{A}_{\tau_i} \neq \mathcal{M}] \, \mathbf{P}[\mathcal{A}_{\tau_i} \neq \mathcal{M}] \\ &\leq \tfrac{P_{\mathbf{e}}}{\delta} + \delta. \end{aligned} \tag{3.67}$$

Using Lemma 10 together with equations (3.65) and (3.67) we get

$$(1 - \tilde{\epsilon}_3) E_i \leq \tilde{\epsilon}_4 + \sum\nolimits_{j=i+1}^{k+1} \beta_j \mathcal{J}\left(\tfrac{f_j}{\beta_j}\right) \qquad i = 1, 2, \ldots, k \tag{3.68}$$

where β_j and f_j are defined for $j \in \{1, 2, \ldots, k+1\}$ as follows[7]

$$\beta_j \triangleq \tfrac{\mathbf{E}[\tau_j] - \mathbf{E}[\tau_{j-1}]}{\mathbf{E}[\tau]} \qquad f_j \triangleq \tfrac{\mathbf{E}\left[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_{j-1}}) - \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_j})\right]}{\mathbf{E}[\tau]}. \tag{3.69}$$

Note that $\mathcal{A}_{\tau_i}$ has at most $|\mathcal{M}|$ elements and its complement, $\overline{\mathcal{A}_{\tau_i}}$ has at most $\frac{|\mathcal{M}|}{|\mathcal{M}^i|}$ elements. Thus,

$$\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i}, \mathcal{A}_{\tau_i}) \leq \ln |\mathcal{M}| \qquad \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i}, \overline{\mathcal{A}_{\tau_i}}) = \ln \tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \tag{3.70}$$

Using the fact that $h(x) = -x \ln x - (1-x) \ln(1-x) \leq \ln 2$ for all $x \in [0,1]$ together with equations (3.70) we get

$$\begin{aligned} \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i}) &= h(\mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}]) + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}] \, \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i}, \mathcal{A}_{\tau_i}) + \mathbf{P}[\mathsf{M} \notin \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}] \, \mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i}, \overline{\mathcal{A}_{\tau_i}}) \\ &\leq \ln 2 + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}] \ln |\mathcal{M}| + \mathbf{P}[\mathsf{M} \notin \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}] \ln \tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} \\ &= \ln 2 + \ln \tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i} | \, \mathsf{Y}^{\tau_i}] \ln |\mathcal{M}^i|. \end{aligned} \tag{3.71}$$

By calculating the expected value of both sides of the inequality (3.71) we get,

$$\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i})] = \ln 2 + \ln \tfrac{|\mathcal{M}|}{|\mathcal{M}^i|} + \mathbf{P}[\mathsf{M} \in \mathcal{A}_{\tau_i}] \ln |\mathcal{M}^i|. \tag{3.72}$$

Using equations (3.67) and (3.72) together with $\mathbf{E}[\mathcal{H}(\mathsf{M}|\mathsf{Y}^{\tau_i})] = R - \sum_{j=1}^{i} f_j$ we get,

$$R - \sum\nolimits_{j=1}^{i} f_j \leq \tilde{\epsilon}_4 + R + (\tilde{\epsilon}_3 - 1) \sum\nolimits_{j=1}^{i} R_j$$

Hence

$$\sum\nolimits_{j=1}^{i} f_j \geq (1 - \tilde{\epsilon}_3) \sum\nolimits_{j=1}^{i} R_j - \tilde{\epsilon}_4 \qquad i = 1, 2, \ldots, (k+1) \tag{3.73}$$

As a result of equation (3.43) and the fact that $\mathcal{I}\left(\mathsf{Y}_{t+1}; \mathsf{M} \middle| \mathsf{Y}^t\right) \leq \mathcal{C}$ we have

$$f_i \leq \mathcal{C} \beta_i \qquad i \in \{1, 2, \ldots, (k+1)\}. \tag{3.74}$$

[7] We use the convention $\tau_0 = 0$ and $\tau_{k+1} = \tau$.

Thus equations (3.68), (3.73), (3.74) imply that the following set of necessary conditions, that are satisfied by all achievable $(\vec{R},\vec{E})$ pairs for some $(\beta, f)=(\beta_1,\ldots,\beta_{k+1},f_1,\ldots,f_{k+1})$:

$$(1-\tilde{\epsilon}_3)E_i-\tilde{\epsilon}_4 \leq \sum\nolimits_{j=i+1}^{k+1}\beta_j \mathcal{J}\left(\tfrac{f_j}{\beta_j}\right) \qquad i=1,2,\ldots,k \tag{3.75a}$$

$$\sum\nolimits_{j=1}^{i} f_j \geq (1-\tilde{\epsilon}_3)\sum\nolimits_{j=1}^{i} R_j-\tilde{\epsilon}_4 \qquad i=1,2,\ldots,(k+1) \tag{3.75b}$$

$$f_i \leq \mathcal{C}\beta_i \qquad i=1,2,\ldots,(k+1). \tag{3.75c}$$

We show below that the necessary conditions given in equation (3.75) implies the ones in equation (3.58). First we show that the inequality constraint in equation (3.75b) can be replaced by an equality constraint for all $i\in\{1,2,\ldots,k\}$ without changing the set of rate vector, error exponent vector pairs satisfying the constraints. Since we are imposing a more stringent condition clearly we are not expanding the set of $(\vec{R},\vec{E})$ pairs satisfying the constraints. In order to see why we are not curtailing the set of $(\vec{R},\vec{E})$ pairs satisfying the constraints, let ℓ be the first integer for which equation (3.75b) is a strict inequality. Let ζ be

$$\zeta \triangleq f_\ell-(1-\tilde{\epsilon}_3)R_\ell+\tilde{\epsilon}_4 \mathbb{1}_{\{\ell=1\}}. \tag{3.76}$$

Let us define $(\tilde{f},\tilde{\beta})$ in terms of (f,β) and ζ as

$$\tilde{f}_i=f_i+\zeta(\mathbb{1}_{\{i=\ell+1\}}-\mathbb{1}_{\{i=\ell\}}) \qquad \tilde{\beta}_i=\beta_i+\tfrac{\beta_\ell}{f_\ell}\zeta(\mathbb{1}_{\{i=\ell+1\}}-\mathbb{1}_{\{i=\ell\}}). \tag{3.77a}$$

Then we have

$$\tfrac{\tilde{f}_i}{\tilde{\beta}_i}=\tfrac{f_i}{\beta_i}+\frac{\beta_\ell\frac{\zeta}{f_\ell}}{\beta_{\ell+1}+\beta_\ell\frac{\zeta}{f_\ell}}\left(\tfrac{f_\ell}{\beta_\ell}-\tfrac{f_{\ell+1}}{\beta_{\ell+1}}\right)\mathbb{1}_{\{i=\ell+1\}}. \tag{3.78}$$

Since (f,β) satisfies (3.75c) equation (3.78) implies that $(\tilde{f},\tilde{\beta})$ satisfies (3.75c).
Definition of ζ given in equation (3.76), definition of $(\tilde{f},\tilde{\beta})$ given in (3.77) and the fact that (f,β) satisfies (3.75b), implies that $(\tilde{f},\tilde{\beta})$ satisfies (3.75b). Furthermore (3.75b) is satisfied with equality for all $i\leq\ell$ for $(\tilde{f},\tilde{\beta})$ because (3.75b) is satisfied with equality for all $i<\ell$ for (f,β).
Finally note that as result of concavity of the $\mathcal{J}(\cdot)$ function we have,

$$\beta_{\ell+1}\mathcal{J}\left(\tfrac{f_{\ell+1}}{\beta_{\ell+1}}\right)+\beta_\ell\tfrac{\zeta}{f_\ell}\mathcal{J}\left(\tfrac{f_\ell}{\beta_\ell}\right)\leq\tilde{\beta}_{\ell+1}\mathcal{J}\left(\tfrac{\tilde{f}_{\ell+1}}{\tilde{\beta}_{\ell+1}}\right). \tag{3.79}$$

Consequently

$$\sum\nolimits_{j=i+1}^{k+1}\beta_j\mathcal{J}\left(\tfrac{f_j}{\beta_j}\right)\leq\sum\nolimits_{j=i+1}^{k+1}\tilde{\beta}_j\mathcal{J}\left(\tfrac{\tilde{f}_j}{\tilde{\beta}_j}\right) \qquad i=1,2,\ldots,k. \tag{3.80}$$

Since (f,β) satisfies (3.75a), equation (3.80) ensures that $(\tilde{f},\tilde{\beta})$ satisfies (3.75a).
We can repeat this procedure successively and replace the inequality constraints in (3.75b) with equality constraint in all $i\in\{1,2,\ldots,k\}$. Then the lemma follows from the modified (3.75) and fact that $\mathcal{J}(\cdot)\leq\mathcal{D}$.

QED

Note that the necessary condition for the achievability of $(\vec{R},\vec{E})$ pair is same as the sufficient condition apart from the error terms that vanish as $\mathbf{E}[\tau]$ diverges.

3.4 Conclusions

We have considered the single message *message-wise* and the multi-layer *bit-wise UEP* problems and characterized the achievable rate, error exponent region completely for both of the problems.

We have shown that, like the conventional variable-length block coding schemes without *UEP*, it is possible to decouple communication and bulk of the error correction both at the transmitter and at the receiver in *bit-wise UEP* schemes. Unlike the conventional case, however, in *bit-wise UEP* schemes there is hierarchy of bits and thus:

(a) The communication phase of each layer of bits needs to be merged with the error correction phases of the more important layers.

(b) The error correction phase of each layer of bits needs to be merged with the error correction phases of the more important layers.

This is done using the implicit confirmation explicit rejection protocols. They were first suggested by Kudryashov [22] for non-block encoding schemes, it turns out that they also play a key role in *UEP* schemes for block codes.

We have also suggested a new technique for establishing outer bounds to the performance of the variable-length block codes. Lemma 10 relating the missed detection probability of a hypothesis chosen at a stopping time to the decay rate of the entropy of the messages, is at the core of this new technique.

Chapter 4

Feedback Encoding Schemes for Fixed-Length Block Codes[1]

In this chapter we derive upper bounds to the error probability of fixed-length block codes with ideal feedback by modifying the analysis technique of Gallager [16]. Using encoding schemes suggested by Zigangirov [44], D'yachkov [14] and Burnashev [7] we recover previously known best results on binary symmetric channels and improve on the previously known best results on k-ary symmetric channels and binary input channels. Let us start with a brief overview of the previous studies on the problem.

Berlekamp [1] analyzed the decay rate of error probability with block length in fixed-length block codes and gave a closed form expression for the error exponent at zero rate on binary symmetric channels (BSCs). Later Zigangirov [45] presented a rigorous proof of the converse part of Berlekamp's claim. In another paper Zigangirov [44] proposed an encoding scheme, for BSCs which is optimal for all rates larger than a critical rate R_{Zcrit}[2] and at zero rate, i.e. his encoding scheme reaches sphere packing exponent for all rates larger than R_{Zcrit} and reaches the optimal error exponent derived by Berlekamp [1] at zero rate. After that Burnashev [7] improved Zigangirov's inner bound for error exponent at all positive rates below R_{Zcrit} by modifying his encoding scheme.

D'yachkov [14] on the other hand proposed a generalization of the encoding scheme of Zigangirov and obtained a coding theorem for general DMCs. However the optimization problem resulting from his coding theorem, is quite involved and does not allow for simplifications that will lead to conclusions about the error exponents of general DMCs. In [14] after pointing out this fact, D'yachkov focuses on binary input channels and k-ary symmetric channels and derives the error exponent expressions for these channels.

Recently Burnashev and Yamamoto investigated the problem for a noisy feedback link. They considered the binary symmetric channel both in the forward and in the feedback channels and derived a lower bound to the error exponent first at zero rate [8] and then at positive rates [9].

We first bound the error probability of a maximum likelihood decoder for a feedback encoder by modifying Gallager's technique [16]. After that we show how to use the encoding scheme of Zigangirov [44] and D'yackov [14] within this framework. Then in Section 4.2 we discuss the weaknesses of the analysis presented in Section 4.1 and propose a way to

[1]Results presented in this chapter have been reported previously in various conferences [29], [30], [31].

[2]The rate above which random coding exponent is equal to the sphere packing exponent is called the critical rate R_{crit} and [44] that $R_{Zcrit} < R_{crit}$.

improve it by using a weighted maximum likelihood decoder. Then using the encoding scheme suggested by Burnashev [7] we obtain an upper bound to the error exponent and compare it with the results of the previous studies. In the binary symmetric channels we simply recover results of Burnashev in [7], in all other channel this techniques results in an improvement.

4.1 Basic Bound on Error Probability with Feedback

In this section we introduce the basic idea behind our bounds and derive a preliminary bound on the error probability. We start with deriving an upper bound on the error probability, in terms of the expected value of a function of the likelihoods of the messages, ξ_{n}. Function ξ_{n} has a natural extension of to the time instances before n, in terms of the likelihoods of the messages at those time instances. By deriving bounds on the expected rate of decrease in the value of ξ_t we bound the error probability from above. For doing that we use random coding argument and the matching scheme of Zigangirov and D'yachkov.

4.1.1 Error Analysis

We have a discrete memoryless channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and a noiseless, delay-free feedback link from the receiver to the transmitter as before. In this section we assume the encoding scheme Φ is deterministic, i.e. the encoding function at time t is of the form,

$$\Phi_t(\cdot) : \mathcal{M} \times \mathcal{Y}^{t-1} \to \mathcal{X} \qquad \forall t \in \{1, 2, \ldots, \mathsf{n}\}. \tag{4.1}$$

We will specify $\Phi_t(\cdot)$ later. For any encoding scheme the maximum likelihood decoder is given by,[3][4]

$$\hat{\mathsf{M}} = \operatorname*{argmax}_m \mathbf{P}[\mathsf{Y}^{\mathsf{n}} |\, m] \,. \tag{4.2}$$

A maximum likelihood decoder, decodes erroneously only when the likelihood of the actual message is less than or equal to the likelihood of another message. Thus the indicator function of the error event is upper bounded as,

$$\mathbb{1}_{\{\hat{\mathsf{M}} \neq \mathsf{M}\}} \leq \left(\frac{\sum_{m \neq \mathsf{M}} \mathbf{P}[\mathsf{Y}^{\mathsf{n}} |\, m]^{\lambda}}{\mathbf{P}[\mathsf{Y}^{\mathsf{n}} |\, \mathsf{M}]^{\lambda}} \right)^{\rho} \qquad \forall \lambda \geq 0,\ \rho \geq 0. \tag{4.3}$$

Since $P_{\mathbf{e}} = \mathbf{E}\left[\mathbb{1}_{\{\hat{\mathsf{M}} \neq \mathsf{M}\}}\right]$, equation (4.3) implies that,

$$P_{\mathbf{e}} \leq \mathbf{E}[\xi_{\mathsf{n}}] \tag{4.4}$$

where

$$\xi_t \triangleq \mathbf{E}\left[\left(\frac{\sum_{m \neq \mathsf{M}} \mathbf{P}\left[\mathsf{Y}^t |\, m\right]^{\lambda}}{\mathbf{P}[\mathsf{Y}^t |\, \mathsf{M}]^{\lambda}} \right)^{\rho} \middle| \mathsf{Y}^t \right]. \tag{4.5}$$

Note that ξ_t is only a function of Y^t and the encoding scheme up to t, it does not depend on the transmitted message M.

[3] If there are multiple messages with same likelihood the one with the smaller index is chosen.

[4] We make a slight abuse of notation and use $\mathbf{P}[\mathsf{Y}^{\mathsf{n}} |\, m]$ instead of $\mathbf{P}[\mathsf{Y}^{\mathsf{n}} |\, \mathsf{M} = m]$.

Let us assume for the moment that there exists an encoding scheme that satisfies following inequality for all realizations of Y^t and for all t in $\{1, 2, \ldots, \mathsf{n}\}$

$$\mathbf{E}\left[\xi_{t+1} \middle| \mathsf{Y}^t\right] \leq e^{-\mathsf{G}(\rho,\lambda)} \xi_t \tag{4.6}$$

where $\mathsf{G}(\rho, \lambda)$ is a real valued function of ρ and λ. For such an encoding scheme we have

$$\begin{aligned}\mathbf{E}[\xi_{\mathsf{n}}] &\leq e^{-\mathsf{G}(\rho,\lambda)} \mathbf{E}[\xi_{\mathsf{n}-1}] \\ &\leq e^{-\mathsf{n}\mathsf{G}(\rho,\lambda)} \mathbf{E}[\xi_0] \,.\end{aligned} \tag{4.7}$$

Thus using equations (4.4) and (4.7) together with the fact that $\xi_0 = |\mathcal{M} - 1|^\rho < e^{\mathsf{n}R}$ we can upper bound error probability as follows

$$P_{\mathbf{e}} < e^{-\mathsf{n}(\mathsf{G}(\rho,\lambda) - \rho R)}. \tag{4.8}$$

Any lower bound on the achievable values of $\mathsf{G}(\rho, \lambda)$ in equation (4.6) gives us an upper bound on the error probability via equation (4.8). In the following two subsections we establish lower bounds to achievable values of $\mathsf{G}(\rho, \lambda)$ using a random coding argument and using the encoding scheme of Zigangirov and D'yachkov. Before starting that discussion let us point out a fact that will become helpful later on.

For any Y^t such that $\xi_t = 0$ we have $\xi_{t+1} = \ldots = \xi_{\mathsf{n}} = 0$. Hence, $\mathbf{E}\left[\xi_{t+1} \middle| \mathsf{Y}^t\right] \leq e^{-\mathsf{G}(\rho,\lambda)} \xi_t$ holds trivially for all real $\mathsf{G}(\rho, \lambda)$'s. Because of that we assume from now on without loss of generality that $\xi_t > 0$. For Y^t such that $\xi_t > 0$, $\mathbf{E}\left[\xi_{t+1} \middle| \mathsf{Y}^t\right] / \xi_t$ can be written only as a function of the likelihoods of the messages at time t and the encoding scheme at time $t+1$ as follows,

$$\frac{\mathbf{E}\left[\xi_{t+1} | \mathsf{Y}^t\right]}{\xi_t} = \frac{\sum_{m, \mathsf{Y}_{t+1}} \mathbf{P}\left[\mathsf{Y}^t \middle| m\right]^{1-\lambda\rho} W(\mathsf{Y}_{t+1} | \Phi_{t+1}(m, \mathsf{Y}^t))^{1-\lambda\rho} \left(\sum_{\tilde{m} \neq m} \mathbf{P}\left[\mathsf{Y}^t \middle| \tilde{m}\right]^{\lambda} W(\mathsf{Y}_{t+1} | \Phi_{t+1}(\tilde{m}, \mathsf{Y}^t))^{\lambda} \right)^{\rho}}{\sum_m \mathbf{P}[\mathsf{Y}^t | m]^{1-\lambda\rho} \left(\sum_{\tilde{m} \neq m} \mathbf{P}[\mathsf{Y}^t | m]^{\lambda} \right)^{\rho}}. \tag{4.9}$$

4.1.2 Encoding for The Basic Bound on Error Probability

We use different techniques for establishing lower bounds to the achievable $\mathsf{G}(\rho, \lambda)$'s depending on the value of ρ. For $\rho \in [0, 1]$ case we use a random coding argument, to show that there exists a mapping with the given $\mathsf{G}(\rho, \lambda)$. For $\rho \geq 1$ case on the other hand we use a slightly modified version of the encoding scheme suggested by Zigangirov [44] and D'yachkov [14] to obtain an achievable value of $\mathsf{G}(\rho, \lambda)$.

Random Coding

Consider the ensemble of assignments of messages $m \in \mathcal{M}$ to $x \in \mathcal{X}$ at time $t+1$ in which each message m is assigned to the input letter x with probability $P(x)$, independently of the encoding in previous time instances Φ_1^t, previous channel outputs Y^t and assignments of the other messages at time $t+1$, i.e. for all realizations of previous channel outputs Y^t, for all previous assignments Φ_1^t and for all $(x_1, x_2, \ldots, x_{|\mathcal{M}|})$ in $\mathcal{X}^{|\mathcal{M}|}$ we have

$$\mathbb{E}\left[\prod\nolimits_m \mathbb{1}_{\{\Phi_{t+1}(m, \mathsf{Y}^t) = x_m\}} \middle| \mathsf{Y}^t, \Phi_1^t\right] = \prod\nolimits_m P(x_m) \tag{4.10}$$

where $\mathbb{E}[\cdot]$ stands for the ensemble average.

As a result of equation (4.9) and the independence of assignments of different messages i.e. equation (4.10) we have,

$$\mathbb{E}\left[\frac{\mathbf{E}[\xi_{t+1}|\mathsf{Y}^t]}{\xi_t}\middle|\mathsf{Y}^t, \Phi_1^t\right] = \frac{\sum_{m,\mathsf{Y}_{t+1}} \mathbf{P}[\mathsf{Y}^t|m]^{1-\lambda\rho}\mathbb{E}[W(\mathsf{Y}_{t+1}|\Phi_{t+1}(m,\mathsf{Y}^t))^{1-\lambda\rho}|\mathsf{Y}^t,\Phi_1^t]}{\sum_m \mathbf{P}[\mathsf{Y}^t|m]^{1-\lambda\rho}\left(\sum_{\tilde{m}\neq m}\mathbf{P}[\mathsf{Y}^t|m]^{\lambda}\right)^{\rho}}$$

$$\cdot\,\mathbb{E}\left[\left(\sum\nolimits_{\tilde{m}\neq m}\mathbf{P}\left[\mathsf{Y}^t\middle|\,\tilde{m}\right]^{\lambda}W(\mathsf{Y}_{t+1}|\Phi_{t+1}(\tilde{m},\mathsf{Y}^t))^{\lambda}\right)^{\rho}\middle|\,\mathsf{Y}^t,\Phi_1^t\right].$$

Using the concavity of z^ρ function for $\rho\in[0,1]$ together with Jensen's inequality we get,

$$\mathbb{E}\left[\frac{\mathbf{E}[\xi_{t+1}|\mathsf{Y}^t]}{\xi_t}\middle|\mathsf{Y}^t, \Phi_1^t\right] \leq \frac{\sum_{m,\mathsf{Y}_{t+1}} \mathbf{P}[\mathsf{Y}^t|m]^{1-\lambda\rho}\mathbb{E}[W(\mathsf{Y}_{t+1}|\Phi_{t+1}(m,\mathsf{Y}^t))^{1-\lambda\rho}|\mathsf{Y}^t,\Phi_1^t]}{\sum_m \mathbf{P}[\mathsf{Y}^t|m]^{1-\lambda\rho}\left(\sum_{\tilde{m}\neq m}\mathbf{P}[\mathsf{Y}^t|m]^{\lambda}\right)^{\rho}}$$

$$\left(\sum\nolimits_{\tilde{m}\neq m}\mathbf{P}\left[\mathsf{Y}^t\middle|\,\tilde{m}\right]^{\lambda}\mathbb{E}\left[W(\mathsf{Y}_{t+1}|\Phi_{t+1}(\tilde{m},\mathsf{Y}^t))^{\lambda}\middle|\,\mathsf{Y}^t,\Phi_1^t\right]\right)^{\rho}$$

$$=\sum\nolimits_x P(x)\mu_x(P,\rho,\lambda) \tag{4.11}$$

where

$$\mu_x(P,\rho,\lambda)=\sum\nolimits_y W(y|x)^{(1-\rho\lambda)}\left(\sum\nolimits_{\tilde{x}}P(\tilde{x})W(y|\tilde{x})^{\lambda}\right)^{\rho}. \tag{4.12}$$

The inequality (4.11) holds universally for all realizations of Y^t and Φ_1^t. If the ensemble average over the mappings at time $t{+}1$ satisfies inequality (4.11) then there exists at least one mapping of $\mathcal{M}$ to $\mathcal{X}$ that satisfies (4.11). Thus for all realizations of Y^t and Φ_1^t there exists a mapping at time $t{+}1$ such that

$$\frac{\mathbf{E}[\xi_{t+1}|\mathsf{Y}^t]}{\xi_t}\leq\sum\nolimits_x P(x)\mu_x(P,\rho,\lambda) \qquad \forall\rho\in[0,1],\ \lambda\geq 0. \tag{4.13}$$

where $\mu_x(P,\rho,\lambda)$ is defined in equation (4.12).

Zigangirov-D'yachkov Encoding Scheme

$Z-D$ encoding scheme is given in terms of a probability distribution P on the input alphabet $\mathcal{X}$ and a likelihood vector on $\mathcal{M}$. The assignment of messages to the input letters at time $t{+}1$ depends on previous channel outputs Y^t and the assignments $\Phi_1,\Phi_2,\ldots,\Phi_t$ only through the likelihoods of the messages $\mathbf{P}\left[\mathsf{Y}^t\middle|\,m\right]$ for $m\in\mathcal{M}$. First, messages are reordered according to their likelihoods in a decreasing fashion and tilted mass γ of all input letters are set to zero. Then starting from the most likely message, messages are assigned to the input letters with the smallest $\gamma(x)/P(x)$ ratio, one by one. After the assignment of each message the tilted mass $\gamma(x)$ of the corresponding letter is increased by $\mathbf{P}\left[\mathsf{Y}^t\middle|\,m\right]^{\lambda}$.

We assume without loss of generality[5] $\forall m,\tilde{m}\in\mathcal{M}$ if $m\leq\tilde{m}$ then $\mathbf{P}\left[\mathsf{Y}^t\middle|\,m\right]\geq\mathbf{P}\left[\mathsf{Y}^t\middle|\,\tilde{m}\right]$. With that assumption the encoding scheme at time $(t+1)$ for any $P(\cdot)$ is given by:

$$\begin{aligned}\gamma_0(x)&=0 && \forall x\in\mathcal{X}\\ \Phi_{t+1}(m,\mathsf{Y}^t)&=\operatorname*{argmin}_{x\in\mathsf{supp}P}\tfrac{\gamma_{m-1}(x)}{P(x)}\\ \gamma_m(x)&=\sum\nolimits_{1\leq\tilde{m}\leq m:\,\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=x}\mathbf{P}\left[\mathsf{Y}^t\middle|\,\tilde{m}\right]^{\lambda}.\end{aligned}$$

[5] If this is not the case we can rearrange the messages $m\in\mathcal{M}$, according to their likelihoods in decreasing order. If two or more messages have same mass we order them according to their indices.

In [44] and [14] instead of tilted likelihoods of the messages, posterior probabilities of the messages calculated according to a “noisier” DMC V from $\mathcal{X}$ to $\mathcal{Y}$ is used for calculating γ's. Apart from this above encoding scheme is identical to the ones in [44] and [14].

Z-D encoding scheme distributes the tilted likelihoods over the input letters in a particular way: if we consider all the tilted mass except that of $m \in \mathcal{M}$ and normalize it to sum up to one, it is a convex combination of $P(x)$ and $\delta_{x,\tilde{x}}$'s for $\tilde{x} \neq \Phi_{t+1}(m, \mathsf{Y}^t)$, i.e.

$$\frac{\sum_{\tilde{m}\neq m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=x\}} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda}{\sum_{\tilde{m}\neq m} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda} = p_m(0)P(x) + p_m(x) \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M} \tag{4.14a}$$

$$p_m(0) + \sum\nolimits_{x\in\mathcal{X}} p_m(x) = 1 \qquad \forall m \in \mathcal{M} \tag{4.14b}$$

where $p_m \geq 0$ and $p_m(\Phi_{t+1}(m, \mathsf{Y}^t)) = 0$.

In order to see this first assume that m is the last message assigned to the the input letter $\Phi_{t+1}(m, \mathsf{Y}^t)$. Then as a result of the construction we know that

$$\frac{\sum_{\tilde{m}< m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=\Phi_{t+1}(m,\mathsf{Y}^t)\}} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda}{P(\Phi_{t+1}(m,\mathsf{Y}^t))} \leq \frac{\sum_{\tilde{m}< m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=x\}} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda}{P(x)} \qquad \forall x \in \mathcal{X}$$

Since no other message is assigned to $\Phi_{t+1}(m, \mathsf{Y}^t)$ after m we also have

$$\frac{\sum_{\tilde{m}\neq m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=\Phi_{t+1}(m,\mathsf{Y}^t)\}} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda}{P(\Phi_{t+1}(m,\mathsf{Y}^t))} \leq \frac{\sum_{\tilde{m}\neq m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=x\}} \mathbf{P}\left[\mathsf{Y}^t\middle|\tilde{m}\right]^\lambda}{P(x)} \qquad \forall x \in \mathcal{X} \tag{4.15}$$

The likelihood of the messages that are assigned prior to m can not be less than $\mathbf{P}\left[\mathsf{Y}^t\middle| m\right]$, hence equation (4.15) holds for all the messages, not just the last message assigned to each input letter. Consequently we can conclude that there exists a $\gamma_m(\cdot) \geq 0$ such that

$$\sum\nolimits_{\tilde{m}\neq m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=x\}} \mathbf{P}\left[\mathsf{Y}^t\middle| \tilde{m}\right]^\lambda = \gamma_m(0)P(x) + \gamma_m(x) \qquad \forall x \in \mathcal{X}, \quad \forall m \in \mathcal{M}$$

$$\sum\nolimits_{\tilde{m}\neq m} \mathbb{1}_{\{\Phi_{t+1}(\tilde{m},\mathsf{Y}^t)=\Phi_{t+1}(m,\mathsf{Y}^t)\}} \mathbf{P}\left[\mathsf{Y}^t\middle| \tilde{m}\right]^\lambda = \gamma_m(0)P(\Phi_{t+1}(m, \mathsf{Y}^t)) \quad \forall m \in \mathcal{M}$$

If we normalize $\gamma_m(x)$'s over x for each m to sum up to one, we obtain equation (4.14).

Using equation (4.14b), the convexity of z^ρ for $\rho \geq 1$ and Jensen's inequality we get,

$$\begin{aligned}
&\left[\sum\nolimits_x \left(p_m(0)P(x) + p_m(x)\right) W(\mathsf{Y}_{t+1}|x)^\lambda\right]^\rho \\
&\qquad = \left[p_m(0) \left(\sum\nolimits_x P(x)W(\mathsf{Y}_{t+1}|x)^\lambda\right) + \sum\nolimits_x p_m(x) \left(W(\mathsf{Y}_{t+1}|x)^\lambda\right)\right]^\rho \\
&\qquad \leq p_m(0) \left(\sum\nolimits_x P(x)W(\mathsf{Y}_{t+1}|x)^\lambda\right)^\rho + \sum\nolimits_x p_m(x)W(\mathsf{Y}_{t+1}|x)^{\lambda\rho}
\end{aligned} \tag{4.16}$$

As a result of equations (4.9), (4.14a) and (4.16) for all $\rho \geq 1$, $\lambda \geq 0$ input distributions P we have

$$\frac{\mathbf{E}\left[\xi_{t+1}\middle|\mathsf{Y}^t\right]}{\xi_t} \leq \max_{x\in\mathsf{supp}P} \max\{\mu_x(P, \rho, \lambda), \nu_x(\rho\lambda)\} \tag{4.17}$$

where $\mu_x(P, \rho, \lambda)$ is given in equation (4.12) and $\nu_x(\rho\lambda)$ is give by

$$\nu_x(\rho\lambda) = \max\nolimits_{\tilde{x}\neq x, \tilde{x}\in\mathsf{supp}P} \sum\nolimits_y W(y|x)^{(1-\rho\lambda)} W(y|\tilde{x})^{\rho\lambda}. \tag{4.18}$$

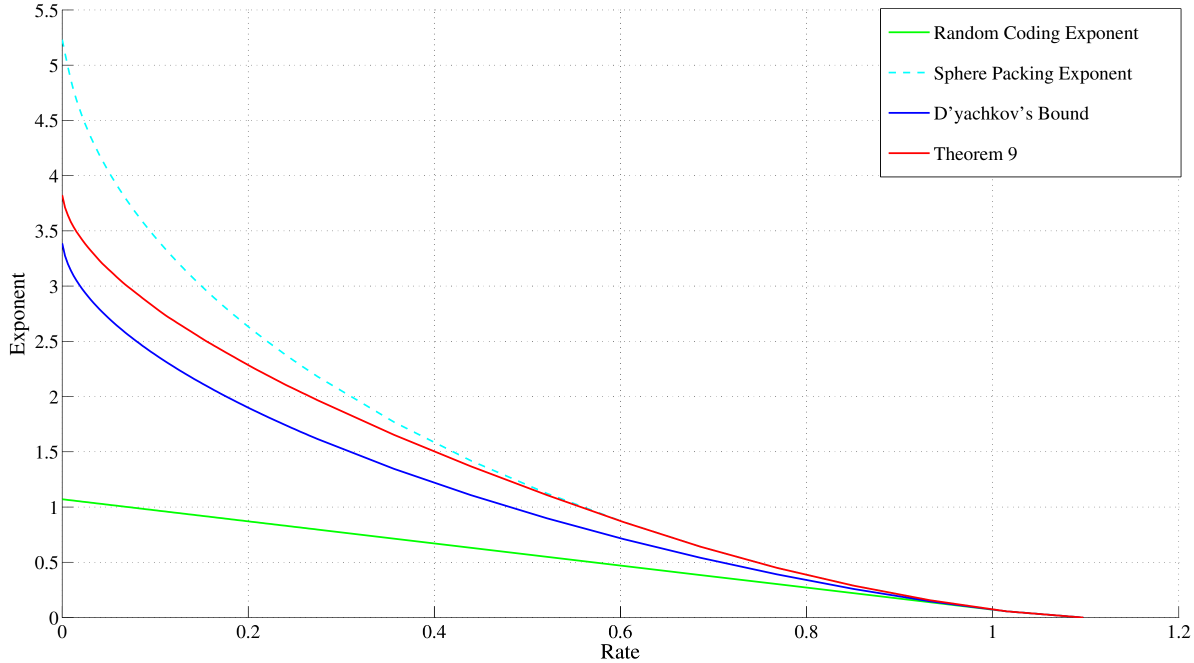

Figure 4-1: Lower bound on the error exponent given in Theorem 9 and previously known best, Dyachkov's lower bound given in [14], are plotted for a ternary symmetric channel with $\delta = 10^{-4}$, together with random coding exponent and sphere packing exponent. Recall that sphere packing exponent is an upper bound on the error exponent with feedback for symmetric channels even.

4.1.3 Basic Bound Error Probability

We summarize the results of our analyzes given in (4.8), (4.13) and (4.17) as follows.

Theorem 9 *For any block length n, rate $R > 0$, input distribution P and tilting factors $\rho \geq 0$, $\lambda \geq 0$ there exists a feedback encoder with $|\mathcal{M}| = \lfloor e^{\mathsf{n}R} \rfloor$ messages whose error probability with maximum likelihood decoder is upper bounded as,*

$$P_{\mathbf{e}} < e^{-\mathsf{n}(\mathsf{G}(P,\rho,\lambda)-\rho R)}$$

where

$$\mathsf{G}(P,\rho,\lambda) = \begin{Bmatrix} -\ln \max_{x \in \mathsf{supp}P} \max\{\mu_x(P,\rho,\lambda), \nu_x(\rho\lambda)\} & \rho \geq 1 \\ -\ln \sum_x P(x)\mu_x(P,\rho,\lambda) & \rho \in [0,1] \end{Bmatrix} \tag{4.19}$$

and $\mu_x(P,\rho,\lambda)$ and $\nu_x(\rho\lambda)$ are given in equations (4.12) and (4.18) respectively.

When optimized over P, ρ and λ, Theorem 9 recovers the results of Zigangirov [44] and D'yachkov [14] for binary input channels. The improvements due to the analysis presented in this section become salient only in channels with more than two input letters. The channels whose transition probabilities are of the form

$$W(y|x) = \begin{cases} 1-\delta & x = y \\ \frac{\delta}{|\mathcal{X}|-1} & x \neq y \end{cases} \tag{4.20}$$

are called k-ary symmetric channels. On k-ary symmetric channels previously known best performance is that of [14]. In Figure 4-1 we have plotted the exponent curves resulting from [14] and Theorem 9 for a ternary symmetric channel with $\delta = 10^{-4}$ for comparison.

4.2 Improved Error Analysis with Weighted Likelihoods and Stopping Time

In the previous section we have required the encoding scheme to satisfy

$$\mathbf{E}\left[\xi_{t+1}\middle|\,\mathsf{Y}^t\right] \leq e^{-\mathsf{G}(\rho,\lambda)}\xi_t$$

for all values of Y^t. In order to ensure that when finding the feasible $\mathsf{G}(\rho, \lambda)$'s we established bounds that hold for all possible $|\mathcal{M}|$ dimensional likelihood vectors. However we do know that likelihoods of the messages will have comparable values at least for the initial part of the block. In other words, if we calculate a posteriori probability distribution by normalizing the likelihoods; for some part of the block, even the message with the maximum posterior probability will have a posterior probability smaller than ϵ for an $\epsilon \ll 1$. In order make use of this fact we will keep track of the likelihoods using a stopping time and apply different encoding schemes before and after this stopping time.

In addition in previous section we have used the same tilting factor in encoding throughout the block. But depending on the tilted probability distribution we might want to allow changes in tilting factor of the encoding. For doing that we need to replace the maximum likelihood decoder with a weighted maximum likelihood decoder. Clearly maximum likelihood decoder is the best decoder for any encoder in terms of the average error probability. However, it does not necessarily lead to the best bounds on the error probability when used in conjunction with our analysis. As we will see later in this section for certain values of the rate, maximum weighted likelihood decoders result in better bounds than the maximum likelihood decoder.

Burnashev [7] has already used similar techniques within the framework of [44] for binary symmetric channels and improved the results of [44]. For binary symmetric channels analysis presented in this section is simply an alternative derivation of Burnashev's results in [7]. But our analysis technique allow us to extend the gains of these observations to a broader class of memoryless channels.

In the rest of this section we do the error analysis again from scratch in order to account for the above mentioned modifications. We first derive an upper bound on the error probability, in terms of a general encoding scheme and a decoder. Then we specify the encoding scheme and the decoder and obtain parametric bounds on the error probability.

4.2.1 Error Analysis Part I

Incorporating the two modifications discussed above require us to not only do the error analysis again from scratch but also change our model of the encoder slightly. We will assume, that at each time t receiver sends an additional random variable of its choice U_t together with the channel output Y_t to the transmitter. The transmitter receives the feedback link symbol for time t, $\mathsf{Z}_t = (\mathsf{Y}_t, \mathsf{U}_t)$ before the transmission of X_{t+1} and use it in the encoding function Φ_{t+1} at time $t+1$. Thus the feedback encoder Φ is a sequence $(\Phi_1, \Phi_2, \dots, \Phi_\mathsf{n})$ of mappings of the form

$$\Phi_t(\cdot) : \mathcal{M} \times \mathcal{Z}^{t-1} \to \mathcal{X} \qquad \forall t \in \{1, 2, \dots, \mathsf{n}\}. \tag{4.21}$$

Any performance achievable using a feedback encoder of the form given in equation (4.21) is also achievable by deterministic encoders of the from given in equation (4.1). We use the encoder of the form (4.21) in order to simplify the analysis.

In this section we use weighted likelihoods, $\phi(\mathsf{Z}^{\mathsf{n}}|m)$'s, instead of likelihoods, $\mathbf{P}[\mathsf{Z}^{\mathsf{n}}|\, m]$'s, in our decoder.

$$\hat{\mathsf{M}} = \underset{m}{\operatorname{argmax}} \phi(\mathsf{Z}^{\mathsf{n}}|m) \tag{4.22}$$

where $\phi(\mathsf{Z}^{\mathsf{n}}|m)$ is non-negative function to be specified later.

A decoder of the form given in (4.22) decodes erroneously only when $\phi(\mathsf{Z}^{\mathsf{n}}|\mathsf{M})$ is less than or equal to $\phi(\mathsf{Z}^{\mathsf{n}}|m)$ for some $m \neq \mathsf{M}$. Hence the indicator function of the error event is upper bounded as,

$$\mathbb{1}_{\{\hat{\mathsf{M}}\neq \mathsf{M}\}} \leq \left(\frac{\sum_{m\neq \mathsf{M}} \phi(\mathsf{Z}^{\mathsf{n}}|m)}{\phi(\mathsf{Z}^{\mathsf{n}}|\mathsf{M})} \right)^{\rho} \qquad \forall \rho \geq 0. \tag{4.23}$$

Unlike equation (4.3) we have not included λ in the bound, simply because we include λ within the definition of $\phi(\cdot|\cdot)$. We define ξ_t and use it to bound the error probability as we did in the last section.

$$\xi_t \triangleq \mathbf{E}\left[\left(\frac{\sum_{m\neq \mathsf{M}} \phi(\mathsf{Z}^t|m)}{\phi(\mathsf{Z}^t|\mathsf{M})} \right)^{\rho} \middle| \mathsf{Z}^t \right]. \tag{4.24}$$

As it was the case in the last section, ξ_t is only a function of Z^t and the encoding scheme up to t; it does not depend on the transmitted message M. Furthermore as a result of equations (4.23) and (4.24) the conditional error probability given Z^{n} is bounded as

$$\mathbf{E}\left[\mathbb{1}_{\{\hat{\mathsf{M}}\neq \mathsf{M}\}} \middle| \mathsf{Z}^{\mathsf{n}} \right] \leq \xi_{\mathsf{n}} \tag{4.25}$$

In order keep track of the weighted likelihoods of the messages we use a stopping time measurable with respect to receivers observation. Let τ be a stopping time with respect to the stochastic sequence $\mathsf{Z}_1, \mathsf{Z}_2, \mathsf{Z}_3, \ldots$, i.e. with respect to the receivers observation. For each t let ζ_t be a high probability subset set of $\mathcal{Z}^t$ to be determined later.[6] Let ζ^τ be the set of Z^τ's such that all subsequences are in the corresponding high probability subset and $\overline{\zeta^\tau}$ be its complement, i.e.

$$\zeta^\tau \triangleq \{ z^\tau : \forall t \leq \tau, z^t \in \zeta_t \} \tag{4.26a}$$

$$\overline{\zeta^\tau} \triangleq \{ z^\tau : z^\tau \notin \zeta^\tau \}. \tag{4.26b}$$

Then for all t, $\mathbb{1}_{\{\zeta^\tau\}} \mathbb{1}_{\{\tau=t\}}$ is measurable in Z^t, i.e.

$$\mathbf{E}\left[\mathbb{1}_{\{\zeta^\tau\}} \mathbb{1}_{\{\tau=t\}} \middle| \mathsf{Z}^t \right] = \mathbb{1}_{\{\zeta^\tau\}} \mathbb{1}_{\{\tau=t\}}. \tag{4.27}$$

Note that $\mathbb{1}_{\{\overline{\zeta^\tau}\}} + \mathbb{1}_{\{\zeta^\tau\}} = 1$ and $\mathbb{1}_{\{\tau>\mathsf{n}\}} + \sum_{t=1}^{\mathsf{n}} \mathbb{1}_{\{\tau=t\}} = 1$ thus we have

$$\mathbb{1}_{\{\hat{\mathsf{M}}\neq \mathsf{M}\}} \leq \mathbb{1}_{\{\overline{\zeta^\tau}\}} + \mathbb{1}_{\{\zeta^\tau\}} \mathbb{1}_{\{\tau>\mathsf{n}\}} + \mathbb{1}_{\{\zeta^\tau\}} \sum_{t=1}^{\mathsf{n}} \mathbb{1}_{\{\tau=t\}} \mathbb{1}_{\{\hat{\mathsf{M}}\neq \mathsf{M}\}}. \tag{4.28}$$

Taking the expectation of both sides of (4.28) over M and using (4.25) and (4.27) we get

$$\mathbf{E}\left[\mathbb{1}_{\{\hat{\mathsf{M}}\neq \mathsf{M}\}} \middle| \mathsf{Z}^{\mathsf{n}} \right] \leq \mathbb{1}_{\{\overline{\zeta^\tau}\}} + \mathbb{1}_{\{\zeta^\tau\}} \mathbb{1}_{\{\tau>\mathsf{n}\}} + \mathbb{1}_{\{\zeta^\tau\}} \sum_{t=1}^{\mathsf{n}} \mathbb{1}_{\{\tau=t\}} \xi_{\mathsf{n}}. \tag{4.29}$$

[6] Say with probability $\mathbf{P}[\zeta_t] = 1 - e^{-\mathsf{n}^2}$.

If we take the expectation over Z^{n} we get,

$$P_{\mathbf{e}} \leq P_{\mathbf{e}}{}^{*} + P_{\mathbf{e}_{\mathsf{n}}}^{*} + \sum\nolimits_{t=1}^{\mathsf{n}} P_{\mathbf{e}t} \tag{4.30}$$

where

$$P_{\mathbf{e}}{}^{*} \triangleq \mathbf{E}\left[\mathbb{1}_{\{\overline{\zeta^{\tau}}\}}\right] \tag{4.31a}$$

$$P_{\mathbf{e}_{\mathsf{n}}}^{*} \triangleq \mathbf{E}\left[\mathbb{1}_{\{\zeta^{\tau}\}}\mathbb{1}_{\{\tau>\mathsf{n}\}}\right] \tag{4.31b}$$

$$P_{\mathbf{e}t} \triangleq \mathbf{E}\left[\mathbb{1}_{\{\zeta^{\tau}\}}\mathbb{1}_{\{\tau=t\}}\xi_{\mathsf{n}}\right]. \tag{4.31c}$$

In order to bound the error probability further we need to specify the decoder, high probability sets ζ_t, the stopping time τ and the encoding scheme.

4.2.2 Weighted Likelihood Decoder

We use the following weighted likelihood function in the decoder,

$$\phi(\mathsf{Z}^t|m) \triangleq \left\{ \begin{array}{rl} \mathbf{P}\left[\mathsf{Z}^t\middle|\, m\right]^{\lambda_a} & \text{if } t \leq \tau \\ \mathbf{P}\left[\mathsf{Z}_{\tau+1}^t\middle|\, m, \mathsf{Z}^{\tau}\right]^{\lambda_b} \mathbf{P}\left[\mathsf{Z}^{\tau}\middle|\, m\right]^{\lambda_a} & \text{if } t > \tau \end{array} \right\} \tag{4.32}$$

where τ is the first time instance at which a message reaches a tilted posterior probability higher than ϵ.

$$\tau \triangleq \min\left\{ t : \max_{m\in\mathcal{M}} \frac{\mathbf{P}\left[\mathsf{Z}^t|m\right]^{\lambda_a}}{\sum_{\tilde{m}} \mathbf{P}\left[\mathsf{Z}^t|\tilde{m}\right]^{\lambda_a}} \geq \epsilon \right\} \tag{4.33}$$

For some $\mathsf{Z}^{\mathsf{n}} \in \mathcal{Z}^{\mathsf{n}}$ the largest tilted posterior probability among the messages can be smaller than ϵ in first n times; in such cases $\tau > \mathsf{n}$.

Note that τ given in equation (4.33) is a stopping time with respect to the stochastic sequence $\mathsf{Z}_1, \mathsf{Z}_2, \ldots$ as we have assumed in Section 4.2.1. In fact τ is also a stopping time with respect to the stochastic sequence $\mathsf{Y}_1, (\mathsf{U}_1, \mathsf{Y}_2), (\mathsf{U}_2, \mathsf{Y}_3), \ldots$. In order to see this first note that for any t receiver chooses U_t after observing $(\mathsf{Z}^{t-1}, \mathsf{Y}_t)$, without knowing the actual message M. Thus given $(\mathsf{Z}^{t-1}, \mathsf{Y}_t)$, U_t is independent of M:

$$\mathbf{P}\left[\mathsf{U}_t\middle|\, \mathsf{Z}^{t-1}, \mathsf{Y}_t\right] = \mathbf{P}\left[\mathsf{U}_t\middle|\, \mathsf{Z}^{t-1}, \mathsf{Y}_t, \mathsf{M}\right].$$

Then the likelihoods of the messages depend on U_t as follows

$$\mathbf{P}\left[\mathsf{Z}^t\middle|\, m\right] = \mathbf{P}\left[\mathsf{Z}^{t-1}, \mathsf{Y}_t\middle|\, m\right] \cdot \mathbf{P}\left[\mathsf{U}_t\middle|\, \mathsf{Z}^{t-1}, \mathsf{Y}_t\right]. \tag{4.34}$$

Hence the receiver can determine whether $\tau > t$ or not without the knowledge of U_t, i.e. τ is also a stopping time with respect to the stochastic sequence $\mathsf{Y}_1, (\mathsf{U}_1, \mathsf{Y}_2), (\mathsf{U}_2, \mathsf{Y}_3), \ldots$.

4.2.3 Encoding Scheme

The encoding scheme is composed of two phases and the transition between them happens at τ. In the first phase, $[1, \tau]$, each message is assigned to input letters independently of the others according to some probability distribution $\mathsf{P}_a(\cdot)$ on $\mathcal{X}$. Since all the messages have small tilted posteriors we will be able to bound the expected change in ξ_t pretty accurately. In the second phase, $[\tau + 1, \mathsf{n}]$, we use an encoding scheme like the one in Section 4.1 that ensures at each time ξ_t decreases in expectation by some fixed multiplicative factor.

Encoding in $[1,\tau]$: Random Coding

At each time t, encoding function Φ_t assigns each message to an input letter. For all $(\mathsf{Z}^{t-2},\mathsf{Y}_{t-1})$ such that $t \leq \tau$ we use U_{t-1} to choose these assignments randomly using a probability distribution $\mathsf{P}_a(\cdot)$ on $\mathcal{X}$. For all $(\mathsf{Z}^{t-2},\mathsf{Y}_{t-1})$ such that $t \leq \tau$ assignments of the messages are independent of one another and $(\mathsf{Z}^{t-2},\mathsf{Y}_{t-1})$:

$$\mathbf{E}\left[\prod_m \mathbb{1}_{\{\Phi_t(m,\mathsf{Z}^{t-1})=x_m\}}\middle|\, \mathsf{Z}^{t-2},\mathsf{Y}_{t-1}\right] = \prod_m \mathbf{E}\left[\mathbb{1}_{\{\Phi_t(m,\mathsf{Z}^{t-1})=x_m\}}\middle|\, \mathsf{Z}^{t-2},\mathsf{Y}_{t-1}\right]$$
$$= \prod_m \mathsf{P}_a(x_m). \tag{4.35}$$

For such an encoding scheme following approximate equality will hold with high probability whenever $\tau \geq t$

$$\sum_{m:\Phi_t(m,\mathsf{Z}^{t-1})=x} \phi(\mathsf{Z}^{t-1}|m) \approx P(x) \sum_m \phi(\mathsf{Z}^{t-1}|m).$$

The reason is that if $t \leq \tau$ then tilted posterior probability of each message is small, i.e. less than ϵ, and there are many of them, i.e. $|\mathcal{M}|$. Thus if we assign each one of them to the input letter x with probability $\mathsf{P}_a(x)$ independently, the total tilted posterior probability of the messages that are assigned to input letter x is very close to $\mathsf{P}_a(x)$ most of the time for all x in $\mathcal{X}$. Following lemma states this fact more precisely,

Lemma 13 *Let ζ_t be*

$$\zeta_t = \left\{\mathsf{Z}^t : (t \geq \tau) \text{ or } (t < \tau \text{ and } \left|\frac{\sum_{m:\Phi_{t+1}(m,\mathsf{Z}^t)=x}\phi(\mathsf{Z}^t|m)}{\sum_m \phi(\mathsf{Z}^t|m)} - \mathsf{P}_a(x)\right| \leq \epsilon_1 \mathsf{P}_a(x),\ \forall x \in \mathcal{X})\right\} \tag{4.36}$$

then

$$\mathbf{P}\left[Z^t \in \zeta_t \middle|\, \mathsf{Z}^{t-1},\mathsf{Y}_t\right] = 1 \qquad t \geq \tau \tag{4.37a}$$
$$\mathbf{P}\left[Z^t \in \zeta_t \middle|\, \mathsf{Z}^{t-1},\mathsf{Y}_t\right] \leq 1 - 2|\mathcal{X}|e^{-\frac{\epsilon_1^2}{2\epsilon}\min_{x:\mathsf{P}_a(x)>0}\mathsf{P}_a(x)} \qquad t < \tau \tag{4.37b}$$

Proof:
Note that (4.37a) follows from the definition of ζ_{t-1} trivially, so we focus on (4.37b). Let a_m, $\overline{a_m}(x)$ and $\sigma(a_m(x))$ be

$$a_m(x) \triangleq \mathbb{1}_{\{\Phi_t(m,\mathsf{Z}^t)=x\}}\phi(\mathsf{Z}^t|m) \tag{4.38a}$$
$$\overline{a_m}(x) \triangleq \mathbf{E}\left[a_m(x)\middle|\, \mathsf{Z}^{t-1},\mathsf{Y}_t\right] \tag{4.38b}$$
$$\sigma(a_m(x))^2 \triangleq \mathbf{E}\left[(a_m(x)-\overline{a_m}(x))^2\middle|\, \mathsf{Z}^{t-1},\mathsf{Y}_t\right] \tag{4.38c}$$

Then

$$\overline{a_m}(x) = \phi(\mathsf{Z}^t|m)\mathsf{P}_a(x) \tag{4.39a}$$
$$\sigma(a_m(x))^2 = \phi(\mathsf{Z}^t|m)^2\mathsf{P}_a(x)(1-\mathsf{P}_a(x))$$
$$\leq \phi(\mathsf{Z}^t|m)\epsilon \sum_{\tilde{m}} \phi(\mathsf{Z}^t|\tilde{m})\mathsf{P}_a(x)(1-\mathsf{P}_a(x)) \tag{4.39b}$$

where the inequality follows from the fact that $\max_m \phi(\mathsf{Z}^t|m) < \epsilon \sum_m \phi(\mathsf{Z}^t|m)$ when $\tau > t$.

As result of [10, Theorem 5.3] we have,

$$\mathbf{P}\left[|\sum_{m}\tilde{a}_m(x)| \geq \beta \middle| \mathsf{Z}^{t-1},\mathsf{Y}_t\right] \leq 2e^{-\frac{\beta^2}{2\sum_m \sigma(a_m(x))^2}} \tag{4.40}$$

If we choose $\beta = \epsilon_1 \mathsf{P}_a(x) \sum_m \phi(\mathsf{Z}^t|m)$ and apply union bound over $x \in \mathcal{X}$, Lemma 13 follows from equations (4.38), (4.39) and (4.40).

QED

The closeness of tilted posterior distribution of input letters to $\mathsf{P}_a(\cdot)$ can be used to bound ξ_τ from above. Note that for ζ_t given in equation (4.36), $\forall \mathsf{Z}^t \in \zeta_t$ and $\forall \mathsf{Y}_{t+1} \in \mathcal{Y}$ we have

$$\left|\frac{\sum_m \mathbf{P}\left[\mathsf{Y}_{t+1}|m,\mathsf{Z}^t\right]^{\lambda_a}\phi(\mathsf{Z}^t|m)}{\sum_m \phi(\mathsf{Z}^t|m)} - \sum_x W(\mathsf{Y}_{t+1}|x)^{\lambda_a}\mathsf{P}_a(x)\right| \leq \epsilon_1 \sum_x W(\mathsf{Y}_t|x)^{\lambda_a}\mathsf{P}_a(x) \tag{4.41}$$

Thus for all $\mathsf{Z}^t \in \zeta^\tau$ we have

$$\begin{aligned}
\frac{\sum_{\tilde{m}}\phi(\mathsf{Z}^t|\tilde{m})}{\phi(\mathsf{Z}^t|m)} &= \frac{\sum_{\tilde{m}}\mathbf{P}\left[\mathsf{Y}_t|\tilde{m},\mathsf{Z}^{t-1}\right]^{\lambda_a}\phi(\mathsf{Z}^{t-1}|\tilde{m})}{\mathbf{P}\left[\mathsf{Y}_t|m,\mathsf{Z}^{t-1}\right]^{\lambda_a}\phi(\mathsf{Z}^{t-1}|m)} \\
&\leq (1+\epsilon_1)\,\frac{\sum_x W(\mathsf{Y}_t|x)^{\lambda_a}\mathsf{P}_a(x)}{\mathbf{P}\left[\mathsf{Y}_t|m,\mathsf{Z}^{t-1}\right]^{\lambda_a}}\,\frac{\sum_{\tilde{m}}\phi(\mathsf{Z}^{t-1}|\tilde{m})}{\phi(\mathsf{Z}^{t-1}|m)} \\
&\leq e^{\mathsf{n}R}\,(1+\epsilon_1)^t \prod_{\ell=1}^{t} \frac{\sum_x W(\mathsf{Y}_\ell|x)^{\lambda_a}\mathsf{P}_a(x)}{\mathbf{P}\left[\mathsf{Y}_\ell|m,\mathsf{Z}^{\ell-1}\right]^{\lambda_a}} \\
&= \frac{e^{\mathsf{n}R}(1+\epsilon_1)^t}{\Gamma_t(m)}
\end{aligned} \tag{4.42}$$

where $\Gamma_t(m)$ is defined as

$$\Gamma_t(m) \triangleq \prod_{\ell=1}^{t} \frac{\mathbf{P}\left[\mathsf{Y}_\ell|m,\mathsf{Z}^{\ell-1}\right]^{\lambda_a}}{\sum_x W(\mathsf{Y}_\ell|x)^{\lambda_a}\mathsf{P}_a(x)} \tag{4.43}$$

Similarly one can also prove that for all $\mathsf{Z}^t \in \zeta^\tau$ we have

$$\frac{\sum_{\tilde{m}}\phi(\mathsf{Z}^t|\tilde{m})}{\phi(\mathsf{Z}^t|m)} \geq \frac{e^{\mathsf{n}R}(1-\epsilon_1)^t}{\Gamma_t(m)}. \tag{4.44}$$

Note that for any Z^t and m,

$$\sum_{\tilde{m}\neq m}\phi(\mathsf{Z}^t|\tilde{m}) \leq \sum_{\tilde{m}}\phi(\mathsf{Z}^t|\tilde{m})$$

Thus as a result of equation (4.42)

$$\begin{aligned}
\xi_t &\leq \mathbf{E}\left[\left(\frac{\sum_m \phi(\mathsf{Z}^t|m)}{\phi(\mathsf{Z}^t|\mathsf{M})}\right)^\rho \middle| \mathsf{Z}^t\right] \\
&= (1+\epsilon_1)^{\rho t} e^{\rho \mathsf{n}R}\mathbf{E}\left[\Gamma_t(\mathsf{M})^{-\rho}\middle| \mathsf{Z}^t\right] \qquad \forall \mathsf{Z}^t \in \zeta^t
\end{aligned} \tag{4.45}$$

Encoding in $[\tau+1,\mathsf{n}]$

In the interval $[\tau+1,\mathsf{n}]$ we use the deterministic encoding schemes used in Section 4.1 to ensure a fixed multiplicative decrease in each step as before. For ξ_t defined in equation

(4.24), $\phi(\mathsf{Z}^t|m)$ defined in equation (4.32) and Z^t such that $t \geq \tau$ we have,

$$\frac{\mathbf{E}[\xi_{t+1}|\mathsf{Z}^t]}{\xi_t} = \frac{\sum_{m,\mathsf{Y}_{t+1}} \mathbf{P}[\mathsf{Z}^t|m] W(\mathsf{Y}_{t+1}|\Phi_{t+1}(m,\mathsf{Z}^t))^{1-\lambda_b\rho}\left(\sum_{\tilde{m}\neq m} \frac{\phi(\mathsf{Z}^t|\tilde{m})}{\phi(\mathsf{Z}^t|m)} W(\mathsf{Y}_{t+1}|\Phi_{t+1}(\tilde{m},\mathsf{Y}^t))^{\lambda_b}\right)^{\rho}}{\sum_m \mathbf{P}[\mathsf{Z}^t|m]\left(\sum_{\tilde{m}\neq m}\frac{\phi(\mathsf{Z}^t|\tilde{m})}{\phi(\mathsf{Z}^t|m)}\right)^{\rho}} \tag{4.46}$$

Repeating the analysis we have done in Section (4.1) for $\phi(\mathsf{Z}^t|m)$ instead of $\mathbf{P}\left[\mathsf{Y}^t\middle| m\right]^{\lambda}$ we can conclude that there exists an encoding scheme for which,

$$\mathbf{E}\left[\xi_{t+1}\middle| \mathsf{Z}^t\right] = e^{-\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}\xi_t \qquad\qquad \forall \mathsf{Z}^t : t \geq \tau \tag{4.47}$$

where $\mathsf{G}(\mathsf{P}_b, \rho, \lambda_b)$ is given in equation (4.19).

4.2.4 Error Analysis Part II

In this subsection we continue the error analysis we have started in Subsection 4.2.1 for the decoder and encoder specified in Subsection 4.2.2 and Subsection 4.2.3 respectively.

Bounding $P_{\mathbf{e}}{}^*$

For $\overline{\zeta^\tau}$ defined in equation (4.26) as a result of the union bound we have

$$\mathbf{E}\left[\mathbb{1}_{\{\overline{\zeta^\tau}\}}\right] \leq \sum\nolimits_{t=1}^{\mathsf{n}} \mathbf{E}\left[\mathbb{1}_{\{\bar{\zeta}_t\}}\right]. \tag{4.48}$$

Using Lemma 13 and the definition of $P_{\mathbf{e}}{}^*$ given in equation (4.31a) we get

$$P_{\mathbf{e}}{}^* \leq 2|\mathcal{X}|\mathsf{n}e^{-\frac{\epsilon_1^2}{2\epsilon}\min_{x:\mathsf{P}_a(x)>0}\mathsf{P}_a(x)}. \tag{4.49}$$

Bounding $P_{\mathbf{e}_\mathsf{n}}^*$

If $\tau > \mathsf{n}$ then $\max_{\tilde{m}} \phi(\mathsf{Z}^\mathsf{n}|\tilde{m}) \leq \epsilon \sum_m \phi(\mathsf{Z}^\mathsf{n}|m)$. Thus using equation (4.42) we get,

$$\begin{aligned}\mathbb{1}_{\{\tau>\mathsf{n}\}}\mathbb{1}_{\{\mathsf{Z}^\mathsf{n}\in\zeta^\mathsf{n}\}} &\leq \mathbb{1}_{\left\{\max_m \frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1+\epsilon_1)^\mathsf{n}}\leq\epsilon\right\}}\\ &\leq \mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1+\epsilon_1)^\mathsf{n}}\leq\epsilon\right\}}\end{aligned} \tag{4.50}$$

Thus using definition of $P_{\mathbf{e}_\mathsf{n}}^*$ given in equation (4.31b) we get

$$\begin{aligned}P_{\mathbf{e}_\mathsf{n}}^* &\leq \mathbf{E}\left[\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1+\epsilon_1)^\mathsf{n}}\leq\epsilon\right\}}\right]\\ &\leq (\epsilon e^{\mathsf{n}R}(1+\epsilon_1)^\mathsf{n})^{\rho_0}\mathbf{E}\left[\Gamma_\mathsf{n}(\mathsf{M})^{-\rho_0}\right]\\ &\leq (\epsilon e^{\mathsf{n}R}(1+\epsilon_1)^\mathsf{n})^{\rho_0}\left(\sum\nolimits_x \mathsf{P}_a(x)\mu_x(\mathsf{P}_a,\rho_0,\lambda_a)\right)^\mathsf{n}\end{aligned} \tag{4.51}$$

where $\mu_x(P, \rho, \lambda)$ is given in equation (4.12).

Indeed there is a slight abuse of notation in the above array of inequalities. The expectations that leads to inequality (4.51) are not the expectations resulting from the encoding scheme described in Subsection 4.2.3 in which after τ encoder stops using random coding and switches to a deterministic encoding scheme. In the expressions leading to inequality

(4.51) it is assumed that random coding continues after τ. Since the equations (4.31b), (4.45) and (4.50) hold for expectations calculated in either way; expectations calculated in either way gives us an upper bound on $P_{\mathbf{e}\mathsf{n}}^*$.

Bounding $P_{\mathbf{e}t}$

As a result of definition of τ given in equation (4.33), if $\tau = t$ then

$$\max_m \phi(\mathsf{Z}^t|m) \geq \epsilon \sum_{\tilde{m}} \phi(\mathsf{Z}^t|\tilde{m}).$$

Thus using equation (4.44) we get

$$\mathbb{1}_{\{\tau=t\}} \mathbb{1}_{\left\{\mathsf{Z}^t \in \zeta^t\right\}} \leq \mathbb{1}_{\left\{\max_m \frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon\right\}}. \tag{4.52}$$

Equation (4.52) is an implicit lower bound on τ. To see this first note that if $\mathsf{Z}^t \in \zeta^t$ and $t \leq \tau$ then

$$\Gamma_t(m) \leq \left(\max_{x,y} \frac{W(y|x)^{\lambda_a}}{\sum_{\tilde{x}} W(y|\tilde{x})^{\lambda_a} \mathsf{P}_a(\tilde{x})}\right)^t. \tag{4.53}$$

If in addition $t = \tau$ then $t \geq t_0$ where,

$$t_0 \triangleq \frac{\mathsf{n}(R + \ln(1-\epsilon_1)) + \ln \epsilon}{\max_{x,y} \frac{W(y|x)^{\lambda_a}}{\sum_{\tilde{x}} W(y|\tilde{x})^{\lambda_a} \mathsf{P}_a(\tilde{x})}}. \tag{4.54}$$

Hence considering $P_{\mathbf{e}t}$'s definition give in (4.31c) we can conclude that $P_{\mathbf{e}t} = 0$ for $t < t_0$.

In order to bound $P_{\mathbf{e}t}$ for $t \geq t_0$ first we use equation (4.47).

$$\begin{aligned}
P_{\mathbf{e}t} &\leq \mathbf{E}\left[\mathbb{1}_{\{\tau=t\}} \mathbb{1}_{\left\{\mathsf{Z}^t \in \zeta^t\right\}} \xi_{\mathsf{n}}\right] \\
&\leq \mathbf{E}\left[\mathbb{1}_{\{\tau=t\}} \mathbb{1}_{\left\{\mathsf{Z}^t \in \zeta^t\right\}} \xi_t\right] e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}.
\end{aligned}$$

Then using equations (4.45) and (4.52) we get

$$P_{\mathbf{e}t} \leq \mathbf{E}\left[\mathbb{1}_{\left\{\max_m \frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon\right\}} \Gamma_t(\mathsf{M})^\rho\right] (1+\epsilon_1)^{\rho t} e^{\mathsf{n}\rho R - (\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}$$

The message that gains the tilted posterior probability of ϵ can be the actual message or some other message. We analyze two cases separately:

$$\begin{aligned}
&P_{\mathbf{e}t} \leq P_{\mathbf{e}ta} + P_{\mathbf{e}tb} &(4.55)\\
&P_{\mathbf{e}ta} \triangleq \mathbf{E}\left[\mathbb{1}_{\left\{\max_m \frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon; \frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon\right\}} \Gamma_t(\mathsf{M})^{-\rho}\right] (1+\epsilon_1)^{\rho t} e^{\mathsf{n}\rho R} e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)} \\
&P_{\mathbf{e}tb} \triangleq \mathbf{E}\left[\mathbb{1}_{\left\{\max_m \frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon; \frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t} < \epsilon\right\}} \Gamma_t(\mathsf{M})^{-\rho}\right] (1+\epsilon_1)^{\rho t} e^{\mathsf{n}\rho R} e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}
\end{aligned}$$

Note that equation (4.55) and the definitions of $P_{\mathbf{e}ta}$ and $P_{\mathbf{e}tb}$ have the slight abuse of notation, like the one in equation (4.51): encoding scheme is assumed to continue to employ the random coding after τ while calculating these expectations.

Let us start with bounding $P_{\mathbf{e}ta}$. Note that if $\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t} \geq \epsilon$ then the first condition of the indicator function is always satisfied. Thus for all $\rho_a \geq 0$ we have

$$\begin{aligned}
P_{\mathbf{e}ta} &= \mathbf{E}\left[\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\right](1+\epsilon_1)^{\rho t}e^{\mathsf{n}\rho R}e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}\\
&\leq \mathbf{E}\left[(\tfrac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t\epsilon})^{\rho_a}\Gamma_t(\mathsf{M})^{-\rho}\right](1+\epsilon_1)^{\rho t}e^{\mathsf{n}\rho R}e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}\\
&= \tfrac{(1+\epsilon_1)^{\rho t}}{\epsilon^{\rho_a}(1-\epsilon_1)^{\rho_a t}}e^{\mathsf{n}(\rho-\rho_a)R}\left(\sum\nolimits_x \mathsf{P}_a(x)\mu_x(\mathsf{P}_a,\rho-\rho_a,\lambda_a)\right)^t e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}
\end{aligned} \tag{4.56}$$

For bounding $P_{\mathbf{e}tb}$ first note that,

$$\begin{aligned}
&\mathbf{E}\left[\mathbb{1}_{\left\{\max_{m\neq\mathsf{M}}\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon;\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\right]\\
&\qquad= \mathbf{E}\left[\mathbb{1}_{\left\{\max_{m\neq\mathsf{M}}\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\right]\\
&\qquad= \mathbf{E}\left[\mathbf{E}\left[\mathbb{1}_{\left\{\max_{m\neq\mathsf{M}}\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\middle|\mathsf{Y}^t\right]\right]
\end{aligned} \tag{4.57}$$

Note that given Y^t, $\Gamma_t(m)$ for different m's are independent of each other. Hence,

$$\begin{aligned}
&\mathbf{E}\left[\mathbb{1}_{\left\{\max_{m\neq\mathsf{M}}\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\middle|\mathsf{Y}^t\right]\\
&\qquad= \mathbf{E}\left[\mathbb{1}_{\left\{\max_{m\neq\mathsf{M}}\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\middle|\mathsf{Y}^t\right]\mathbf{E}\left[\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\middle|\mathsf{Y}^t\right]\\
&\qquad= \mathbf{E}\left[\sum\nolimits_{m\neq\mathsf{M}}\mathbb{1}_{\left\{\frac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t}\geq\epsilon\right\}}\middle|\mathsf{Y}^t\right]\mathbf{E}\left[\mathbb{1}_{\left\{\frac{\Gamma_t(\mathsf{M})}{e^{\mathsf{n}R}(1-\epsilon_1)^t}<\epsilon\right\}}\Gamma_t(\mathsf{M})^{-\rho}\middle|\mathsf{Y}^t\right]\\
&\qquad\leq \mathbf{E}\left[\sum\nolimits_{m\neq\mathsf{M}}(\tfrac{\Gamma_t(m)}{e^{\mathsf{n}R}(1-\epsilon_1)^t\epsilon})^{\rho_b}\middle|\mathsf{Y}^t\right]\mathbf{E}\left[(\tfrac{e^{\mathsf{n}R}(1-\epsilon_1)^t\epsilon}{\Gamma_t(\mathsf{M})})^{\rho_c}\Gamma_t(\mathsf{M})^{-\rho}\middle|\mathsf{Y}^t\right]
\end{aligned} \tag{4.58}$$

Thus using equations (4.57) and (4.58) we get,

$$\begin{aligned}
P_{\mathbf{e}tb} &\leq \tfrac{\epsilon^{\rho_c-\rho_b}(1+\epsilon_1)^{\rho t}}{(1-\epsilon_1)^{(\rho_c-\rho_b)t}}e^{\mathsf{n}(\rho+\rho_c-\rho_b)R}\mathbf{E}\left[\sum\nolimits_{m\neq\mathsf{M}}\Gamma_t(m)^{\rho_b}\Gamma_t(\mathsf{M})^{-(\rho+\rho_c)}\right]e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}\\
&\leq \tfrac{\epsilon^{\rho_c-\rho_b}(1+\epsilon_1)^{\rho t}}{(1-\epsilon_1)^{(\rho_c-\rho_b)t}}e^{\mathsf{n}(\rho+\rho_c-\rho_b+1)R}\left(\sum\nolimits_x \mathsf{P}_a(x)\eta_x(\mathsf{P}_a,\rho+\rho_c,\lambda_a,\rho_b)\right)^t e^{-(\mathsf{n}-t)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)}
\end{aligned} \tag{4.59}$$

where

$$\eta_x(\mathsf{P}_a,\rho,\lambda_a,\rho_b) = \sum\nolimits_y W(y|x)\tfrac{[\sum_{\tilde{x}}\mathsf{P}_a(\tilde{x})W(y|\tilde{x})^{\lambda_a}]^{\rho}}{W(y|x)^{\lambda_a\rho}}\tfrac{\sum_{\tilde{x}}\mathsf{P}_a(\tilde{x})W(y|\tilde{x})^{\lambda_a\rho_b}}{[\sum_{\tilde{x}}\mathsf{P}_a(\tilde{x})W(y|\tilde{x})^{\lambda_a}]^{\rho_b}} \tag{4.60}$$

Parametric Error Bounds

We use equations (4.30), (4.49), (4.51), (4.55), (4.56), and (4.59) to bound the error probability in terms of ϵ, ϵ_1, ρ, λ_a, λ_b, P_a, P_b, ρ_0, ρ_a, ρ_b and ρ_c as follows:

Theorem 10 *For any block length* n*, rate* $R>0$ *there exists fixed-length block code with feedback with* $|\mathcal{M}| = \lfloor e^{\mathsf{n}R} \rfloor$ *messages whose error probability is upper bounded as,*

$$
\begin{aligned}
P_{\mathbf{e}} &\leq P_{\mathbf{e}}{}^{*} + P_{\mathbf{e}_{\mathsf{n}}}^{*} + \sum\nolimits_{t=t_0}^{\mathsf{n}} (P_{\mathbf{e}ta} + P_{\mathbf{e}tb}) \\
t_0 &= \left\lceil \frac{\mathsf{n}(R+\ln(1-\epsilon_1))+\ln\epsilon}{\max_{x,y} \frac{W(y|x)^{\lambda_a}}{\sum_{\tilde{x}} W(y|\tilde{x})^{\lambda_a} \mathsf{P}_a(\tilde{x})}} \right\rceil \\
P_{\mathbf{e}}{}^{*} &\leq 2|\mathcal{X}|\mathsf{n} e^{-\frac{\epsilon_1^2}{2\epsilon}\min_{x:\mathsf{P}_a(x)>0} \mathsf{P}_a(x)} \\
P_{\mathbf{e}_{\mathsf{n}}}^{*} &\leq \epsilon^{\rho_0}(1+\epsilon_1)^{\mathsf{n}\rho_0} e^{\mathsf{n}(\rho_0 R + \ln \sum_x \mathsf{P}_a(x)\mu_x(\mathsf{P}_a,\rho_0,\lambda_a))} \\
P_{\mathbf{e}ta} &\leq \frac{(1+\epsilon_1)^{\rho t}}{\epsilon^{\rho_a}(1-\epsilon_1)^{\rho_a t}} e^{\mathsf{n}((\rho-\rho_a)R + \frac{t}{\mathsf{n}} \ln \sum_x \mathsf{P}_a(x)\mu_x(\mathsf{P}_a,\rho-\rho_a,\lambda_a) - (1-\frac{t}{\mathsf{n}})\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b))} \\
P_{\mathbf{e}tb} &\leq \frac{\epsilon^{\rho_c-\rho_b}(1+\epsilon_1)^{\rho t}}{(1-\epsilon_1)^{(\rho_c-\rho_b)t}} e^{\mathsf{n}((\rho+\rho_c-\rho_b+1)R + \frac{t}{\mathsf{n}} \ln \sum_x \mathsf{P}_a(x)\eta_x(\mathsf{P}_a,\rho+\rho_c,\lambda_a,\rho_b) - (1-\frac{t}{\mathsf{n}})\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b))}
\end{aligned}
$$

where $\mu_x(P,\rho,\lambda)$*,* $\eta_x(\mathsf{P}_a,\rho,\lambda_a,\rho_b)$ *and* $\mathsf{G}(P,\rho,\lambda)$ *are given in equations (4.12), (4.60) and (4.19) respectively.*

Note that if $\frac{\epsilon \mathsf{n}}{\epsilon_1{}^2} = o(1/\mathsf{n})$ then $P_{\mathbf{e}}{}^{*}$ decays super exponentially with block length n. Furthermore if $\epsilon_1 = o(1)$ and ϵ decays to zero subexponentially with n then terms in front of the exponential functions in $P_{\mathbf{e}_{\mathsf{n}}}^{*}$, $P_{\mathbf{e}ta}$ and $P_{\mathbf{e}tb}$ diverges to infinity subexponentially with n. Thus for a (ϵ,ϵ_1) pair satisfying both conditions, like $\epsilon = \frac{1}{\mathsf{n}(\ln \mathsf{n})^3}$ and $\epsilon_1 = \frac{1}{\ln \mathsf{n}}$, the exponential decay rate of the error is determined by the worst exponent among $P_{\mathbf{e}_{\mathsf{n}}}^{*}$, $P_{\mathbf{e}ta}$'s and $P_{\mathbf{e}tb}$'s.

In order to generalize results of [7] for binary symmetric channels to general DMCs we assume that $\rho_a \in [0,\rho]$ and set $\rho_0 = \rho - \rho_a$, $\rho_b = 1+\rho_a$, $\rho_c = 0$ and $\lambda_a = \frac{1}{1+\rho}$. Consequently;

$$
\begin{aligned}
-\frac{\ln P_{\mathbf{e}_{\mathsf{n}}}^{*}}{\mathsf{n}} &= \mathsf{H}(\mathsf{P}_a,\rho,\rho_a) - (\rho-\rho_a)R + o(1) \\
-\frac{\ln P_{\mathbf{e}ta}}{\mathsf{n}} &= \frac{t}{\mathsf{n}}\mathsf{H}(\mathsf{P}_a,\rho,\rho_a) + (1-\frac{t}{\mathsf{n}})\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b) - (\rho-\rho_a)R + o(1) \\
-\frac{\ln P_{\mathbf{e}tb}}{\mathsf{n}} &= \frac{t}{\mathsf{n}}\mathsf{H}(\mathsf{P}_a,\rho,\rho_a) + (1-\frac{t}{\mathsf{n}})\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b) - (\rho-\rho_a)R + o(1)
\end{aligned}
$$

where

$$
\mathsf{H}(\mathsf{P}_a,\rho,\rho_a) = -\ln \sum\nolimits_{x,y} \mathsf{P}_a(x) W(y|x) \left(\frac{\sum_{\tilde{x}} \mathsf{P}_a(\tilde{x}) W(y|\tilde{x})^{\frac{1}{1+\rho}}}{W(y|x)^{\frac{1}{1+\rho}}} \right)^{\rho-\rho_a} \tag{4.61}
$$

Thus

$$
\begin{aligned}
\mathcal{E}(R) \geq F(R) &= \max_{\mathsf{P}_a,\rho} \min_{\alpha\in[\alpha_0(\rho,\mathsf{P}_a),1]} \max_{\rho_a\in[0,\rho],\mathsf{P}_b,\lambda_b} \alpha\mathsf{H}(\mathsf{P}_a,\rho,\rho_a) + (1-\alpha)\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b) - (\rho-\rho_a)R \\
&= \max_{\mathsf{P}_a,\rho} \min_{\alpha\in[\alpha_0(\rho,\mathsf{P}_a),1]} \max_{\rho_a\in[0,\rho]} \left[\alpha\mathsf{H}(\mathsf{P}_a,\rho,\rho_a) - (\rho-\rho_a)R + (1-\alpha)\max_{\mathsf{P}_b,\lambda_b} \mathsf{G}(\mathsf{P}_b,\rho,\lambda_b) \right]
\end{aligned} \tag{4.62}
$$

where $\mathsf{H}(\mathsf{P}_a,\rho,\rho_a)$,$\mathsf{G}(\mathsf{P}_b,\rho,\lambda_b)$ are given in equations (4.61), (4.19) and $\alpha_0(\rho,\mathsf{P}_a)$ is given by

$$
\alpha_0(\rho,\mathsf{P}_a) \triangleq \min\left\{ \frac{R}{\max_{x,y} \ln \frac{W(y|x)^{\frac{1}{1+\rho}}}{\sum_{\tilde{x}} \mathsf{P}_a(\tilde{x}) W(y|\tilde{x})^{\frac{1}{1+\rho}}}}, 1 \right\}.
$$

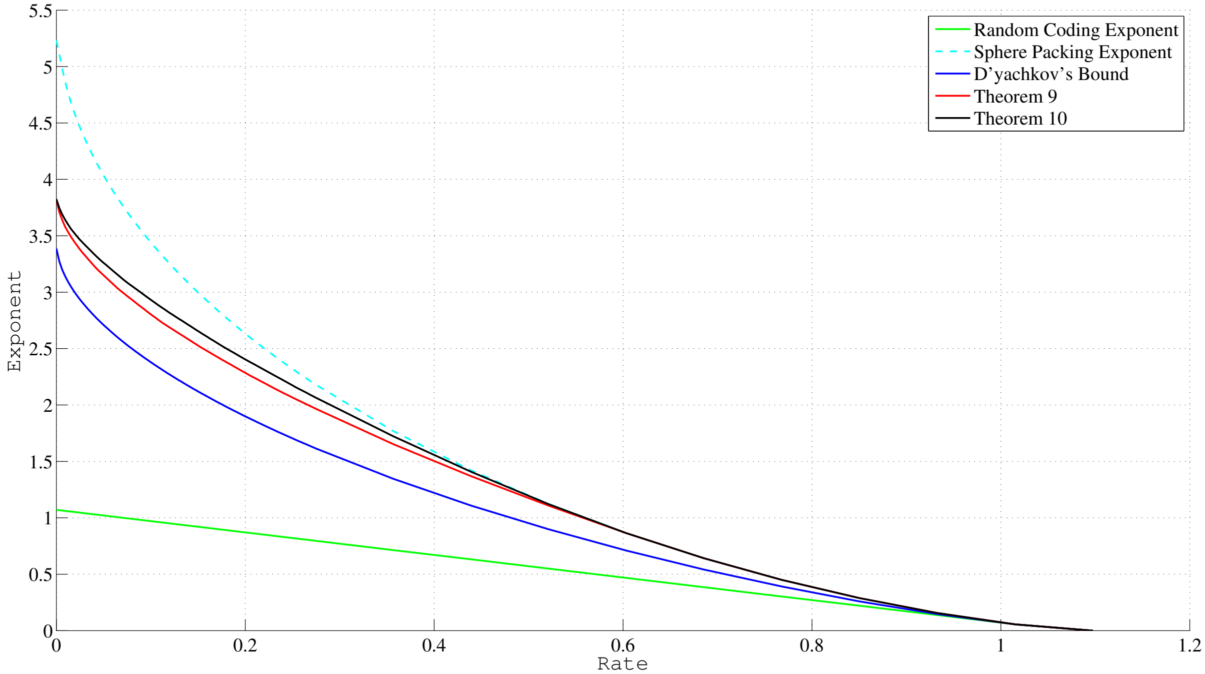

Figure 4-2: Improvement resulting from Theorem 10 is plotted together with the previous results for a ternary symmetric channel with $\delta = 10^{-4}$ given in Figure 4-1.

In binary symmetric channels the lower bound in (4.62) is equal to the one in [7, Theorem 1], which was derived specifically for binary symmetric channels.

One can further simplify the lower bound $F(R)$ by noting the concavity of $\mathsf{H}(\mathsf{P}_a, \rho, \rho_a)$ in ρ_a. As a result of Holder's inequality[7] and the definition of $\mathsf{H}(\mathsf{P}_a, \rho, \rho_a)$ given in (4.61) we have

$$\beta \mathsf{H}(\mathsf{P}_a, \rho, \rho_{a1}) + (1-\beta)\mathsf{H}(\mathsf{P}_a, \rho, \rho_{a2}) \leq \mathsf{H}(\mathsf{P}_a, \rho, \beta\rho_{a1} + (1-\beta)\rho_{a2}).$$

Consequently $\mathsf{H}(\mathsf{P}_a, \rho, \rho_a)$ is convex in ρ_a for a given (P_a, ρ). Hence we can change order of the minimization over α and maximization over ρ_a in the definition of $F(R)$.

$$\begin{aligned} F(R) &= \max_{\mathsf{P}_a, \rho} \max_{\rho_a \in [0,\rho]} \min_{\alpha \in [\alpha_0(\rho, \mathsf{P}_a), 1]} \left[\alpha \mathsf{H}(\mathsf{P}_a, \rho, \rho_a) - (\rho - \rho_a) R + (1-\alpha) \max_{\mathsf{P}_b, \lambda_b} \mathsf{G}(\mathsf{P}_b, \rho, \lambda_b) \right] \\ &= \max_{\mathsf{P}_a, \rho} \max_{\rho_a \in [0,\rho]} \left[\mathsf{H}(\mathsf{P}_a, \rho, \rho_a) - (\rho - \rho_a) R + (1 - \alpha_0(\rho, \mathsf{P}_a)) \min\{\mathsf{H}(\mathsf{P}_a, \rho, \rho_a), \max_{\mathsf{P}_b, \lambda_b} \mathsf{G}(\mathsf{P}_b, \rho, \lambda_b)\} \right] \end{aligned} \tag{4.63}$$

In Figure 4-2 we have plotted lower bound on $\mathcal{E}_{\mathbf{e}}(R)$ given in equation (4.63) for a ternary symmetric channel[8] with $\delta = 10^{-4}$, together with the previously derived lower bounds, random coding exponent and sphere packing bound. Recall that sphere packing bound in an upper bound to $\mathcal{E}_{\mathbf{e}}(R)$ for all symmetric channels.

---

[7] For any two positive random variables u, v and for any $\alpha \in [0,1]$, $\mathbf{E}[uv] \leq \mathbf{E}\left[u^{\frac{1}{\alpha}}\right]^{\alpha} \mathbf{E}\left[v^{\frac{1}{1-\alpha}}\right]^{1-\alpha}$.

[8] Transition probabilities of a k-ary symetric channel is given in equation (4.20).

4.3 Conclusions

We have presented an alternative method, for analyzing the error probability of feedback encoding schemes. We have recovered or improved all previously known results [44], [14], [4]. Our results suggest that if matching schemes are used at a rate below capacity then the encoding should be damped down via some kind of tilting in order to be optimal in terms of error performance. This not only explains why so called posterior matching schemes [39] are suboptimal in terms of error performance, but also suggests a way to make them optimal in terms of error performance.

Recall that posterior matching schemes are the counter part of Schalkwijk-Kailath scheme in DMC's. Schalkwijk Kailath scheme was designed for memoryless additive Gaussian noise channels. It is already known that, [32], [44], [17] that Schalkwijk and Kailath scheme is suboptimal in terms of its error performance in Gaussian channels. What stands as curious question is whether or not a tilting mechanism can be used to improve the error performance of Schalkwijk-Kailath scheme in additive Gaussian noise channels.

Appendix A

The Error Exponent Tradeoff for Two Messages:

Lemma 14 $\Gamma(T,\Pi)$ *defined in equation (2.34) is equal to*

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

$$\textit{where} \qquad U_s(y|x,\tilde{x}) = \left\{ \begin{array}{lll} \frac{\mathbb{1}_{\{W(y|\tilde{x})>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|\tilde{x})>0} W(\tilde{y}|x)} W(y|x) & \textit{if} & s=0 \\ \frac{W(y|x)^{1-s} W(y|\tilde{x})^{s}}{\sum_{\tilde{y}} W(\tilde{y}|x)^{1-s} W(\tilde{y}|\tilde{x})^{s}} & \textit{if} & s \in (0,1) \\ \frac{\mathbb{1}_{\{W(y|x)>0\}}}{\sum_{\tilde{y}:W(\tilde{y}|x)>0} W(\tilde{y}|\tilde{x})} W(y|\tilde{x}) & \textit{if} & s=1 \end{array} \right\}$$

Proof:

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

where (a) follows from convexity of $\mathsf{D}\left(U\| W_r|\Pi\right)+\lambda(\mathsf{D}\left(U\| W_a|\Pi\right)-T)$ in U and linearity (concavity) of it in λ; (b) holds because minimizing U is $U_{\frac{1}{1+\lambda}}$. The function on the right hand side of (A.1) is maximized at a positive and finite λ iff there is a λ such that $\mathsf{D}\left(U_{\frac{1}{1+\lambda}} \middle\| W_a|\Pi\right) = T$. Thus by substituting $\lambda = \frac{1-s}{s}$ we get

$$\Gamma(T,\Pi) = \left\{ \begin{array}{ccl} \infty & \text{if} & T < \lim_{s\to 0^+} \mathsf{D}\left(U_s \| W_a|\Pi\right) \\ \lim_{s\to 0^+} \mathsf{D}\left(U_s \| W_r|\Pi\right) & \text{if} & T = \lim_{s\to 0^+} \mathsf{D}\left(U_s \| W_a|\Pi\right) \\ \mathsf{D}\left(U_s \| W_r|\Pi\right) & \text{if} & T = \mathsf{D}\left(U_s \| W_a|\Pi\right) \quad \text{for some } s \in (0,1) \\ \lim_{s\to 1^-} \mathsf{D}\left(U_s \| W_r|\Pi\right) & \text{if} & T = \lim_{s\to 1^-} \mathsf{D}\left(U_s \| W_a|\Pi\right) \\ \lim_{s\to 1^-} \mathsf{D}\left(U_s \| W_r|\Pi\right) & \text{if} & T > \lim_{s\to 1^-} \mathsf{D}\left(U_s \| W_a|\Pi\right) \end{array} \right\} \tag{A.2}$$

Lemma follows from the definition U_s at $s=0$, $s=1$ and equation (A.2).

QED

Proof [Lemma 3]:
Proof is very much like the one for the converse part of [38, Theorem 5], except few modifications that allow us to handle the fact that encoding schemes in consideration are feedback encoding schemes. Like [38, Theorem 5] we construct a probability measure $P_T[\cdot]$ on $\mathcal{Y}^{\mathsf{n}}$ 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 $P_T[\cdot]$.

For any $T \geq T_0$ and Π, let $S_{T,\Pi}$ be

$$S_{T,\Pi} \triangleq \left\{ \begin{array}{lll} 0 & \text{if} & T < \mathsf{D}(U_0 \| W_a|\Pi) \\ s & \text{if} & \exists s \in [0,1] \text{s.t. } \mathsf{D}(U_s \| W_a|\Pi) = T \\ 1 & \text{if} & T > \mathsf{D}(U_1 \| W_a|\Pi) \end{array} \right\}. \tag{A.3}$$

Recall that

$$T_0 = \max_{x,\tilde{x}} - \ln \sum_{y:W(y|\tilde{x})>0} W(y|x) \quad \text{and} \quad \mathsf{D}(U_0 \| W_a|\Pi) = -\sum_{x,\tilde{x}} \Pi(x,\tilde{x}) \ln \sum_{y:W(y|\tilde{x})>0} W(y|x).$$

Thus

$$T_0 \geq \mathsf{D}(U_0 \| W_a|\Pi) \tag{A.4}$$

As a result of definition of $S_{T,\Pi}$ given in (A.3) and equation (A.4)

$$\mathsf{D}\left(U_{S_{T,\Pi}} \middle\| W_a|\Pi\right) \leq T \qquad \forall T \geq T_0. \tag{A.5}$$

Using Lemma 14, definition of $S_{T,\Pi}$ given in (A.3) and equation (A.4)

$$\mathsf{D}\left(U_{S_{T,\Pi}} \middle\| W_r|\Pi\right) = \Gamma(T,\Pi) \leq \Gamma(T) \qquad \forall T \geq T_0. \tag{A.6}$$

For a feedback encoding schemes with two messages at time t, $\Phi_t(\cdot) : \{m_1, m_2\} \times \mathcal{Y}^{t-1}$, given the the past channel outputs, y^{t-1}, channel input for both of the messages are fixed. Thus there is a corresponding Π:

$$\Pi(x,\tilde{x}) = \left\{ \begin{array}{lll} 0 & \text{if} & (x,\tilde{x}) \neq (\Phi_t(m_1, y^{t-1}), \Phi_t(m_2, y^{t-1})) \\ 1 & \text{if} & (x,\tilde{x}) = (\Phi_t(m_1, y^{t-1}), \Phi_t(m_2, y^{t-1})) \end{array} \right\}. \tag{A.7}$$

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

$$P_T\left[y_t \middle| y^{t-1}\right] = U_{S_{T,\Pi}}(y_t|\Phi_t(m_1, y^{t-1}), \Phi_t(m_2, y^{t-1})). \tag{A.8}$$

As a result of equation (A.5) and equation (A.6) we have,

$$\sum\nolimits_{y_t} P_T\left[y_t \middle| y^{t-1}\right] \ln \frac{P_T\left[y_t | y^{t-1}\right]}{\mathbf{P}[y_t|\mathsf{M}=m_1, y^{t-1}]} \leq T \tag{A.9}$$

$$\sum\nolimits_{y_t} P_T\left[y_t \middle| y^{t-1}\right] \ln \frac{P_T\left[y_t | y^{t-1}\right]}{\mathbf{P}[y_t|\mathsf{M}=m_2, y^{t-1}]} \leq \Gamma(T). \tag{A.10}$$

Now we make a standard measure change argument,

$$\mathbf{P}[y^{\mathsf{n}}|\,\mathsf{M} = m_1] = e^{\ln \frac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=m_1]}{P_T[y^{\mathsf{n}}]}} P_T[y^{\mathsf{n}}] \tag{A.11}$$

Note that

$$\ln \frac{\mathbf{P}[y^{\mathsf{n}}|\mathsf{M}=m_1]}{P_T[y^{\mathsf{n}}]} = \sum_{t=1}^{\mathsf{n}} \ln \frac{P_T[y_t|y^{t-1}]}{\mathbf{P}[y_t|\mathsf{M}=m_1,y^{t-1}]}$$

$$= \sum_{t=1}^{\mathsf{n}} \left[-Z_{t,1}(y_t|y^{t-1}) + \sum_{\tilde{y}_t} P_T\left[\tilde{y}_t|\, y^{t-1}\right] \ln \frac{P_T[\tilde{y}_t|y^{t-1}]}{\mathbf{P}[\tilde{y}_t|\mathsf{M}=m_1,y^{t-1}]} \right] \tag{A.12}$$

where $Z_{t,1}(y_t|y^{t-1})$ is given by

$$Z_{t,1}(y_t|y^{t-1}) = \sum_{\tilde{y}_t} P_T\left[\tilde{y}_t|\, y^{t-1}\right] \left(\ln \frac{P_T[\tilde{y}_t|y^{t-1}]}{\mathbf{P}[\tilde{y}_t|\mathsf{M}=m_1,y^{t-1}]} - \ln \frac{P_T[y_t|y^{t-1}]}{\mathbf{P}[y_t|\mathsf{M}=m_1,y^{t-1}]} \right). \tag{A.13}$$

Thus using equation (A.9), (A.11), (A.12) and (A.13) we get

$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_1] \geq e^{-\mathsf{n}T} e^{\sum_{t=1}^{\mathsf{n}} Z_{t,1}(y_t|y^{t-1})} P_T\left[y^{\mathsf{n}}\right]. \tag{A.14}$$

Following a similar line of reasoning and using equation (A.10) instead of (A.9) we get

$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_2] \geq e^{-\mathsf{n}\Gamma(T)} e^{\sum_{t=1}^{\mathsf{n}} Z_{t,2}(y_t|y^{t-1})} P_T\left[y^{\mathsf{n}}\right] \tag{A.15}$$

where $Z_{t,2}(y_t|y^{t-1})$ is given by

$$Z_{t,2}(y_t|y^{t-1}) = \sum_{\tilde{y}_t} P_T\left[\tilde{y}_t|\, y^{t-1}\right] \left(\ln \frac{P_T[\tilde{y}_t|y^{t-1}]}{\mathbf{P}[\tilde{y}_t|\mathsf{M}=m_2,y^{t-1}]} - \ln \frac{P_T[y_t|y^{t-1}]}{\mathbf{P}[y_t|\mathsf{M}=m_2,y^{t-1}]} \right). \tag{A.16}$$

Note that $\forall m \in \{m_1, m_2\}$, $t \in \{1, 2, \ldots, \mathsf{n}\}$, $y^{t-1} \in \mathcal{Y}^{t-1}$ and $k \in \{1, 2, \ldots, t-1\}$ we have

$$\sum_{y_t} P_T\left[y_t|\, y^{t-1}\right] Z_{t,m}(y_t|y^{t-1}) = 0 \tag{A.17a}$$

$$(Z_{t,m}(y_t|y^{t-1}))^2 \leq 4(\ln P_{min})^2 \tag{A.17b}$$

$$\sum_{y_t} P_T\left[y_t|\, y^{t-1}\right] Z_{t,m}(y_t|y^{t-1}) Z_{t-k,m}(y_t|y^{t-1-k}) = 0 \tag{A.17c}$$

Thus as a result of equation (A.17), for all $m = \{m_1, m_2\}$

$$\sum_{y^{\mathsf{n}}} P_T\left[y^{\mathsf{n}}\right] \sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1}) = 0 \tag{A.18a}$$

$$\sum_{y^{\mathsf{n}}} P_T\left[y^{\mathsf{n}}\right] \left(\sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1}) \right)^2 \leq 4\mathsf{n}(\ln P_{min})^2 \tag{A.18b}$$

For $m \in \{m_1, m_2\}$ let $\mathcal{Z}_m$ be

$$\mathcal{Z}_m = \left\{ y^{\mathsf{n}} : |\sum_{t=1}^{\mathsf{n}} Z_{t,m}(y_t|y^{t-1})| \leq 4\sqrt{\mathsf{n}} \ln \tfrac{1}{P_{min}} \right\}$$

Using equation (A.18) and Chebychev's inequality we conclude that,

$$P_T\left[\mathcal{Z}_m\right] \geq 3/4 \qquad m = m_1, m_2 \Rightarrow P_T\left[\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}\right] \geq 1/2.$$

Thus either the probability of intersection of $\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}$ with the decoding region of the first message is strictly larger than 1/4 or the probability of intersection of $\mathcal{Z}_{m_1} \cap \mathcal{Z}_{m_2}$ with the decoding region of the second message is equal to or larger than 1/4. Then Lemma 3 follows from equations (A.14) and (A.15).

QED

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

Lemma 15 *For any feedback encoding scheme with the message set* $\mathcal{M} = \{m_1, m_2\}$ *if* $P_{\mathbf{e}|m_2} = 0$ *then* $P_{\mathbf{e}|m_1} \geq e^{-\mathsf{n}T_0}$*. Furthermore there does exist an encoding scheme such that* $P_{\mathbf{e}|m_2} = 0$ *and* $P_{\mathbf{e}|m_1} = e^{-\mathsf{n}T_0}$*.*

Proof:
In order to prove the outer bound i.e. the first part of the lemma we use a construction similar to the one used in the proof of Lemma 3,

$$P_T\left[y_t|\, y^{t-1}\right] = U_0(y_t|\, \Phi_t(m_1, y^{t-1}), \Phi_t(m_2, y^{t-1})).$$

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

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

Then

$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_1] \geq e^{-\mathsf{n}T_0} P_T\left[y^{\mathsf{n}}\right] \tag{A.19}$$
$$\mathbf{P}[y^{\mathsf{n}}|\, \mathsf{M} = m_2] \geq e^{\mathsf{n}\ln P_{min}} P_T\left[y^{\mathsf{n}}\right] \tag{A.20}$$

where P_{min} is the minimum non-zero element of W. Since $P_{\mathbf{e}|m_2} = 0$ equation (A.20) implies that $P_T\left[\hat{\mathsf{M}} = m_2\right] = 1$. Using this fact and equation (A.19) one gets

$$P_{\mathbf{e}|m_1} \geq e^{-\mathsf{n}T_0}. \tag{A.21}$$

In order to prove the inner bound i.e. the second part of the lemma, let us consider the following encoder decoder pair

- The encoder sends x_1^* for the first message and x_2^* for the second message where (x_1^*, x_2^*) is the maximizing input letter pair satisfying $T_0 = -\ln \sum_{y:W(y|x_2^*)>0} W(y|x_1^*)$.
- The decoder decodes to the second message unless $\exists t \in \{1, 2, \ldots, \mathsf{n}\}$ and $\mathsf{Y}_t = y^*$ where y^* satisfies $W(y^*|x_2^*) = 0$.

Then the conditional error probabilities $P_{\mathbf{e}|m_1}$ and $P_{\mathbf{e}|m_2}$ are given by

$$P_{\mathbf{e}|m_1} = e^{-\mathsf{n}T_0}. \qquad\qquad P_{\mathbf{e}|m_2} = 0$$

QED

Appendix B

Certain Results on $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$

B.1 Convexity of $E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi)$ in α:

Lemma 16 *For any input distribution P, $\zeta(P, Q, R)$ defined in equation (2.33) is convex in (Q, R) pair.*

Proof:
Note that

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

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

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

where $V_\gamma = \gamma V_a + (1-\gamma)V_b$.
If the set that a minimization is done over 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,

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

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

QED

Lemma 17 *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[1] of the equation $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$.*

Proof:
For a given input distribution P, the function $E_r(R, P)$ is non-negative, convex and decreasing in R on the interval $R \in [0, \mathsf{I}(P, W)]$. Thus for an (R, P) pair such that $R \in [0, \mathsf{I}(P, W)]$, function $\alpha E_r(\frac{R}{\alpha}, P)$ is strictly increasing and continuous in α on the

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

interval $\alpha \in [\frac{R}{\mathsf{I}(P,W)}, 1]$. Furthermore $\alpha E_r(\frac{R}{\alpha}, P)\big|_{\alpha=\frac{R}{\mathsf{I}(P,W)}} = 0$ and $\alpha E_r(\frac{R}{\alpha}, P)\big|_{\alpha=1} \geq E_{\mathbf{x}}$. Thus for an (R, P) pair such that $R \in [0, \mathsf{I}(P, W)]$ the equation $\alpha E_r(\frac{R}{\alpha}, P) = E_{\mathbf{x}}$ has a unique solution for α.
Note that for any $\gamma \in [0, 1]$

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

where α_γ, T_γ, Q_γ, $R_{1\gamma}$ and $R_{2\gamma}$ are given by,

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

The inequality follows from convexity arguments analogous to the ones used in the proof of Lemma 16.

QED

B.2 $\max_\Pi E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_\Pi E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi)$, $\forall P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha)$

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,

$$
E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi_P) > E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi_P) \qquad \forall P \in \mathcal{P}(R, E_{\mathbf{x}}, \alpha) \tag{B.1}
$$

First consider

$$
\begin{aligned}
\mathsf{D}\left(U \| W_a | \Pi_P\right) &= \tfrac{1}{1-\sum_x (P(x))^2} \sum_{x_1, x_2 : x_1 \neq x_2} P(x_1) P(x_2) \sum\nolimits_y U(y|x_1, x_2) \log \tfrac{U(y|x_1, x_2)}{W(y|x_1)} \\
&= \tfrac{1}{1-\sum_x (P(x))^2} \sum_{x_1, x_2 : x_1 \neq x_2} P(x_1) P(x_2) \sum\nolimits_y U(y|x_1, x_2) \left[\log \tfrac{U(y|x_1, x_2)}{V_U(y|x_1)} + \log \tfrac{V_U(y|x_1)}{W(y|x_1)} \right] \\
&\geq \tfrac{1}{1-\sum_x (P(x))^2} \left[\mathsf{I}\left(P, \hat{V}_U\right) + \mathsf{D}\left(V_U \| W | P\right) \right]
\end{aligned} \tag{B.2}
$$

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

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

Using a similar line of reasoning we get,

$$\mathsf{D}\left(U\|\, W_r|\Pi_P\right) \geq \tfrac{1}{1-\sum_x (P(x))^2}\left[\mathsf{D}\left(\hat{V}_U\middle\|\, W|P\right) + \mathsf{I}\left(P, V_U\right)\right] \tag{B.3}$$

Furthermore $\forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$ using the inequalities (B.2) and (B.3) together the definition of $E_{\mathbf{e}}$ given in (2.13) and (2.18) 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 $\forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$, equation (B.1) holds.

Note that $\forall \Pi$ and $\forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right)$

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

Thus,

$$\max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, \alpha, P, \Pi) > \max_{\Pi} E_{\mathbf{e}}(R, E_{\mathbf{x}}, 1, P, \Pi) \qquad \forall P \in \mathcal{P}\left(R, E_{\mathbf{x}}, \alpha\right). \tag{B.4}$$

Bibliography

[1] E. R. Berlekamp. *Block Coding with Noiseless Feedback*. Ph.d. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering, 1964. http://dspace.mit.edu/handle/1721.1/14783.

[2] P. Berlin, B. Nakiboğlu, B. Rimoldi, and E. Telatar. A simple converse of Burnashev's reliability function. *Information Theory, IEEE Transactions on*, 55(7):3074–3080, Jul. 2009.

[3] S. Borade, B. Nakiboğlu, and L. Zheng. Unequal error protection: An information-theoretic perspective. *Information Theory, IEEE Transactions on*, 55(12):5511–5539, Dec. 2009.

[4] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problems of Information Transmission*, 12(4):10–30, 1976.

[5] M. V. Burnashev. Note: On the article "Data transmission over a discrete channel with feedback, random transmission time". *Problems of Information Transmission*, 13(1):108, 1977.

[6] M. V. Burnashev. Sequential discrimination of hypotheses with control of observations. *Math. USSR Izvestija*, 15(3):419–440, 1980.

[7] M. V. Burnashev. On the reliability function of a binary symmetrical channel with feedback. *Problems of Information Transmission*, 24(1):3–10, 1988.

[8] M. V. Burnashev and H. Yamamoto. On the zero-rate error exponent for a BSC with noisy feedback. *Problems of Information Transmission*, 44(3):198–213, 2008.

[9] M. V. Burnashev and H. Yamamoto. On the reliability function for a BSC with noisy feedback. *Problems of Information Transmission*, 46(2):103–121, 2010.

[10] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Math*, 2006.

[11] I. Csiszár. Joint source-channel error exponent. *Problems of Control and Information Theory*, 9(5):315–328, 1980.

[12] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982.

[13] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, 8:161–168, 1962.

[14] A. G. D'yachkov. Upper bounds on the error probability for discrete memoryless channels with feedback. *Problems of Information Transmission*, 11(4):13–28, 1975.

[15] G. Jr. Forney. Exponential error bounds for erasure, list, and decision feedback schemes. *Information Theory, IEEE Transactions on*, 14(2):206–220, Mar. 1968.

[16] R. G. Gallager. A simple derivation of the coding theorem and some applications. *Information Theory, IEEE Transactions on*, 11(1):3–18, 1965.

[17] R. G. Gallager and B. Nakiboğlu. Variations on a theme by Schalkwijk and Kailath. *Information Theory, IEEE Transactions on*, 56(1):6–17, Jan. 2010.

[18] P. K. Gopala, Y-H. Nam, and H. El Gamal. On the error exponents of ARQ channels with deadlines. *Information Theory, IEEE Transactions on*, 53(11):4265–4273, Nov. 2007.

[19] S. K. Gorantla, B. Nakiboğlu, T. P. Coleman, and L. Zheng. Multi-layer bit-wise unequal error protection for variable length blockcodes with feedback. arXiv:1101.1934v1 [cs.IT], `http://arxiv.org/abs/1101.1934`.

[20] E. A. Haroutunian. A lower bound on the probability of error for channels with feedback. *Problems of Information Transmission*, 13(2):36–44, 1977.

[21] E. Hof, I. Sason, and S. Shamai. Performance bounds for erasure, list, and decision feedback schemes with linear block codes. *Information Theory, IEEE Transactions on*, 56(8):3754–3778, Aug. 2010.

[22] B. D. Kudryashov. On message transmission over a discrete channel with noiseless feedback. *Problems of Information Transmission*, 15(1):3–13, 1973.

[23] N. Merhav. Error exponents of erasure/list decoding revisited via moments of distance enumerators. *Information Theory, IEEE Transactions on*, 54(10):4439–4447, Oct. 2008.

[24] N. Merhav and M. Feder. Minimax universal decoding with an erasure option. *Information Theory, IEEE Transactions on*, 53(5):1664–1675, May 2007.

[25] P. Moulin. A Neyman–Pearson approach to universal erasure and list decoding. *Information Theory, IEEE Transactions on*, 55(10):4462–4478, Oct 2009.

[26] B. Nakiboğlu. Variable Block Length Coding for Channels with Feedback and Cost Constraints. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, September 2005. `http://dspace.mit.edu/handle/1721.1/33802`.

[27] B. Nakiboğlu and R. G. Gallager. Error exponents for variable-length block codes with feedback and cost constraints. *IEEE Transactions on Information Theory*, 54(3):945–963, Mar. 2008.

[28] B. Nakiboğlu and L. Zheng. Error-and-erasure decoding for block codes with feedback. arXiv:0903.4386v3 [cs.IT], `http://arxiv.org/abs/0903.4386`.

[29] B. Nakiboğlu and L. Zheng. Upper bounds to error probability with feedback. In *Information Theory, 2009. IEEE International Symposium on*, pages 1515–1519, June 28 -July 3 2009.

[30] B. Nakiboğlu and L. Zheng. Upper bounds to error probability with feedback. In *Communication, Control, and Computing, 2009. 47th Annual Allerton Conference on*, pages 865–871, Sep. 30 - Oct. 2 2009.

[31] B. Nakiboğlu and L. Zheng. Upper bounds to error probability with feedback. In *Information Theory and Applications Workshop, 2010.*, Jan. 2010.

[32] M. S. Pinsker. The probability of error in block transmission in a memoryless Gaussian channel with feedback. *Problems of Information Transmission*, 4(4):1–4, 1968.

[33] E. Sabbag and N. Merhav. Achievable error exponents for channels with side informationerasure and list decoding. *Information Theory, IEEE Transactions on*, 56(11):5424–5431, Nov. 2010.

[34] J. P. M. Schalkwijk. A coding scheme for additive noise channels with feedback–II: Band-limited signals. *Information Theory, IEEE Transactions on*, 12(2):183–189, Apr. 1966.

[35] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback–I: No bandwidth constraint. *IEEE Transactions on Information Theory*, 12(2):172–182, Apr. 1966.

[36] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[37] C. E. Shannon. The zero error capacity of a noisy channel. *Information Theory, IEEE Transactions on*, 2(3):8–19, Sep. 1956.

[38] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. *Information and Control*, 10(1):65–103, Jan. 1967.

[39] O. Shayevitz and M. Feder. Optimal feedback communication via posterior matching. arXiv:0909.4828v2 [cs.IT], `http://arxiv.org/abs/0909.4828`.

[40] Albert N. Shiriaev. *Probability*. Springer-Verlag Inc., New York, NY, USA, 1996.

[41] İ. E. Telatar. *Multi-Access Communications with Decision Feedback Decoding*. Ph.d. thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, May 1992. `http://dspace.mit.edu/handle/1721.1/13237`.

[42] İ. E. Telatar and R. G. Gallager. New exponential upper bounds to error and erasure probabilities. In *Information Theory, 1994. IEEE International Symposium on*, June 27- July 1 1994.

[43] H. Yamamoto and K. Itoh. Asymptotic performance of a modified Schalkwijk-Barron scheme for channels with noiseless feedback. *Information Theory, IEEE Transactions on*, 25(6):729–733, Nov. 1979.

[44] K. Sh. Zigangirov. Upper bounds for the error probability for channels with feedback. *Problems of Information Transmission*, 6(2):87–92, 1970.

[45] K. Sh. Zigangirov. Optimum zero rate date transmission through binary symmetric channel with feedback. *Problems of Control and Information Theory*, 7(3):21–35, 1978.