12-2011

New Analytical Bounds on the Probability of Code-Word Error for Convolution Codes with Viterbi Decoding

Shweta Tomar
Clemson University, stomar@g.clemson.edu

New Analytical Bounds on the Probability of Code-Word Error for Convolution Codes with Viterbi Decoding

---

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Electrical and Computer Engineering

---

by
Shweta Tomar
December 2011

Accepted by:
Dr. Daniel L. Noneaker, Committee Chair
Dr. Michael B. Pursley
Dr. Harlan B. Russell

Abstract

New analytical bounds are developed for the probability of code-word error in a communication system with convolutional coding and soft-decision Viterbi decoding. The bounds are applicable to communications in channels with asynchronous interferers which are modeled as independent, partial-time white Gaussian interference sources. This model is often used in simulations to reflect the circumstances encountered in many packet radio communication networks. The new results include both purely analytical bounds and offline-simulation-aided bounds that permit implementation of accurate communication-link models with much lower online computational and storage requirements than are required with traditional Monte Carlo simulations of link performance. They significantly improve the trade-off that has previously existed between model fidelity and simulation complexity in Monte Carlo simulations of large-scale wireless communication networks with links employing convolutional coding.

Acknowledgments

I would like to express my sincere gratitude to my advisor Dr. Noneaker for his continuous support and guidance through my gradute studies. His patience and immense knowledge of the subject have helped me overcome all difficulties that I experienced during the research phase; and his feedback and direction were indispensable to the development of this thesis. I would also like to thank Dr. Pursley and Dr. Russell for serving on my committee and for sharing their knowledge and experience through their teaching.

I also thank my colleagues at the Clemson CommNets group for the valuable discussions and advice and for enriching my graduate school experience. I would like to extend a special thanks to my friends, both here and in India, for always being there to share my stories.

Most importantly, I thank my parents for their guidance through all these years and for their unwavering support and encouragement in my pursuits.

Table of Contents

List of Figures

Chapter 1

Introduction

Research in protocols for mobile ad hoc radio networks typically uses simulation of the network as a necessary tool in performance evaluation due to the analytical intractability of the evaluation. The large-scale Monte Carlo simulations require severe tradeoffs between the computation time required and the fidelity of the network model; consequently, compromises are often made in the model that lead to uncertainty about the usefulness of the simulation results. The focus of the tradeoffs in most simulations is the model of the radio link. Bit-level modeling of each radio link in the network leads to highly accurate simulation results but is generally infeasible. Computationally efficient, high-fidelity network simulation is thus achievable only if we can develop computationally simple, accurate means of modeling each radio link under the range of channel conditions considered in the simulation.

In this thesis, we investigate the link-level performance of a communication system using a binary convolutional code, binary antipodal modulation, and soft-decision Viterbi decoding. Our goal is the development of simple, but accurate, bounds on the probability of code-word error in the link which will permit computationally feasible simulation of mobile ad hoc networks with higher fidelity than has been possible until now. Convolutional codes enjoy widespread use in digital communication systems, including satellite communications and mobile wireless communications. The Viterbi algorithm is the most commonly used decoding algorithm for convolutional codes in noisy communication channels. Convolutional

codes and Viterbi decoding are used in CDMA and GSM digital cellular networks, deep-space satellite communications, 802.11 wireless LANs, and mobile ad hoc packet radio networks.

Interference is the dominant factor in link errors in many wireless communication systems, including many mobile ad hoc packet radio networks. Each transmission may be subjected to partial-time jamming and unintentional interference from sources outside the network. It may also be subjected to multiple-access interference from sources within the network [1], and the level of multiple-access interference can vary across the duration of a single packet if the network employs an asynchronous channel-access protocol. Intentional jammers often employ a Gaussian noise source. Under some conditions, the effect of multiple-access interference or non-network interference on the probability of error at the receiver can be approximated with reasonable accuracy using the decision statistics that would arise from appropriately chosen Gaussian interference [2]. We consider a channel in which one or more independent, partial-time white Gaussian interference sources partially overlap the interval of the transmission of interest, and we consider the effect of the channel on the probability of code-word error in the link. (The accuracy of the Gaussian approximation for any particular form of interference is not part of the investigation.)

In the thesis, we focus on the development of bounds and approximations for the probability of code-word error in a link that is subjected to non-stationary additive Gaussian noise. It builds in some respects on the earlier work of others. Bounds on the probability of code-word error for convolution codes and Viterbi decoding are developed in [3] which are based of the union bound on the first-event error probability, P_u (which is presented in [4]. Two bounds are presented in [3]: one which is a linear function of P_u, and another which is a concave function of P_u. The proof of the latter depends heavily on combinatorial arguments. The results in [3] are applicable to hard-decision Viterbi decoding and communications over an independent, identically distributed (i.i.d.) Gaussian noise channel. It has been used in large-scale simulations of direct-sequence spread-spectrum packet radio networks (e.g., [5]).

Some of the new bounds we develop here are expressed in the same functional form

as the two bounds derived in [3]. The new results are applicable to *soft-decision* Viterbi decoding, however, and they are obtained using different analytical techniques than were employed in the earlier work. Moreover, they are applicable to a broader class of Gaussian noise channels (which includes the i.i.d. Gaussian channel as a special case). We also present bounds based on the actual (simulated) first-event error probability. Furthermore, we examine the accuracy of the bounds if the time-varying multiple-access interference is approximated by an i.i.d. Gaussian channel. The collection of results provides a wide range of options for bounds and approximations that represent various tradeoffs between offline computation prior to the network simulation, online computation during the network simulation, and the accuracy of the bound or approximation.

In Chapter 2, we introduce the system and channel models, and in Chapter 3 we develop the notation used in the subsequent chapters. In Chapters 4 and 5, we give a comparative analysis of two interleaving schemes and develop a an equivalent stationary non-Gaussian channel model that represents the effect of a pseudo-random interleaver used over a non-stationary Gaussian channel. In Chapter 6, new analytical bounds on the probability of code-word error are developed. The accuracy of the bounds is examined by their comparison with simulation results for a variety of channels. In Chapter 7, we consider the approximation of the non-Gaussian interference channel by an equivalent i.i.d. Gaussian channel, and we examine the accuracy of the approximation.

Chapter 2

System Description

The system we consider in the thesis is shown in Fig. 2.1. It employs binary convolutional encoding and soft-decision Viterbi decoding. Two models of the additive noise channel are considered in the thesis.

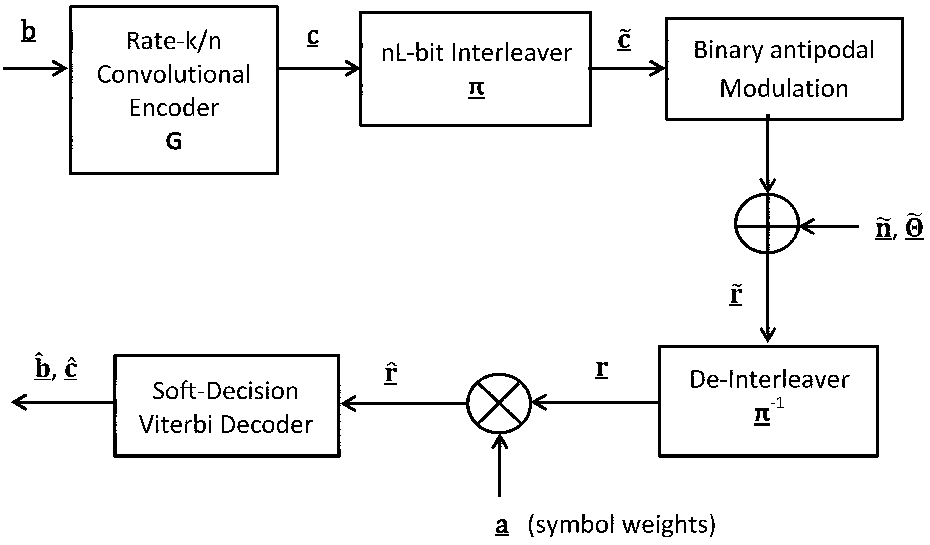

Figure 2.1: System model.

2.1 System Model

The transmitter employs a rate-$\frac{k}{n}$ binary convolutional encoder of memory order m. Information bits form an information word

$$\underline{\mathbf{b}} = (b_0, \ldots, b_{k-1}, b_k, \ldots, b_{(L-m)k-1})$$

which is encoded into the code word denoted by

$$\underline{\mathbf{c}} = (c_0, \ldots, c_{n-1}, c_n, \ldots, c_{nL-1})$$

according to the generator polynomial $\mathbf{G}$ [6] where L is the *block length* of the transmission. Tail bits are used to force the encoder to the all-zeros state at the end of the transmission [6]. We use two rate-$\frac{1}{2}$ convolutional codes as examples in this thesis: the memory-order-three encoder with $\mathbf{G} = (13, 15)$ (octal representation) [6], and the memory-order-six CCSDS (NASA standard) encoder [7] with $\mathbf{G} = (171, 133)$.

The encoded bits of the code word are interleaved using an nL-bit interleaver to mitigate the effect of time-varying channel quality. The interleaver is defined by the vector

$$\underline{\pi} = (\pi_0, \pi_1, \ldots, \pi_{nL-1})$$

such that $\tilde{c}_i = c_{\pi_i}$ for $0 \leq i \leq nL - 1$, where

$$\underline{\tilde{\mathbf{c}}} = (\tilde{c}_0, \ldots, \tilde{c}_{nL-1})$$

denotes the interleaved codeword.

The interleaved code word $\underline{\tilde{\mathbf{c}}}$ is transmitted using binary antipodal modulation with symbol energy E_c; the energy per bit of information is thus $E_b = E_c \cdot \frac{n}{k} \cdot \frac{L}{L-m}$. We consider

a discrete-time additive channel with noise process

$$\underline{\tilde{\mathbf{n}}} = (\tilde{n}_0, \ldots, \tilde{n}_{nL-1})$$

and parameter process

$$\underline{\tilde{\mathbf{\Theta}}} = (\tilde{\Theta}_0, \ldots, \tilde{\Theta}_{nL-1})$$

such that $\{\tilde{n}_0, \tilde{n}_1, \ldots, \tilde{n}_{nL-1}\}$ are conditionally independent, Gaussian random variables given $\underline{\tilde{\mathbf{\Theta}}}$. The received sequence

$$\underline{\tilde{\mathbf{r}}} = (\tilde{r}_0, \ldots, \tilde{r}_{nL-1})$$

is given by

$$\tilde{r}_i = \sqrt{E_c}\,(-1)^{\tilde{c}_i} + \tilde{n}_i$$

for $0 \leq i \leq nL-1$.

The received word $\underline{\tilde{\mathbf{r}}}$ is passed through a de-interleaver defined by the vector

$$\underline{\tilde{\pi}} = (\tilde{\pi}_0, \tilde{\pi}_1, \ldots, \tilde{\pi}_{nL-1})$$

where $\tilde{\pi}_{\pi_i} = i$ for $0 \leq i \leq nL-1$ and the de-interleaved received word

$$\underline{\mathbf{r}} = (r_0, \ldots, r_{nL-1})$$

is given by $r_i = \tilde{r}_{\tilde{\pi}_i}$ for $0 \leq i \leq nL-1$. We assume that the receiver is able to estimate the channel quality for each received symbol accurately, and it weights the received symbols according to maximal-ratio combining [8]. The weighted, de-interleaved received word

$$\underline{\hat{\mathbf{r}}} = (\hat{r}_0, \ldots, \hat{r}_{nL-1})$$

is thus given by

$$\begin{aligned}\hat{r}_i &= a_i r_i \\ &= a_i \sqrt{E_c}\,(-1)^{c_i} + a_i n_i\end{aligned}$$

for $0 \le i \le nL-1$, where $n_i = \tilde{n}_{\tilde{\pi}_i}$, $a_i = [\mathrm{Var}(n_i \mid \Theta_i)]^{-1}$, and $\Theta_i = \tilde{\Theta}_{\tilde{\pi}_i}$.

The code word is decoded by applying the soft-decision Viterbi algorithm with the squared-Euclidean distance metric [6] to $\underline{\hat{\mathbf{r}}}$ so that the detected code word is given by

$$\underline{\hat{\mathbf{c}}} = \arg \max_{\underline{\mathbf{c}} \in \mathcal{C}} Z(\underline{\mathbf{c}})$$

where

$$Z(\underline{\mathbf{c}}) = \sum_{i=0}^{nL-1} \hat{r}_i (-1)^{c_i}$$

is the correlator form of the path-metric [6] and $\mathcal{C}$ is the set of valid code words of length nL code symbols. The receiver thus provides maximum-likelihood detection of the code word (and the information word) based on the continuous-valued channel outputs [6] with perfect channel estimates for each symbol interval. The detected code word is denoted by $\underline{\hat{\mathbf{c}}}$, and the detected information word is denoted by $\underline{\hat{\mathbf{b}}}$.

2.2 Channel Models

We consider two statistical models of the discrete-time additive channel in the thesis. The first is a block-interference model that reflects explicitly the effect of one or more independent, partial-time white Gaussian interference sources that partially overlap the interval of the transmission of interest. The second model approximates the behavior of the channel in the first model but is more amenable than the first model to tractable analysis of the decoder's performance.

2.2.1 Block-interference channel model

The block-interference channel model is a discrete-time channel with time-varying noise power. The noise process in the block-interference channel is independent of the parameter vector $\underline{\tilde{\mathbf{\Theta}}}$. The (non-stationary) noise random process $\underline{\tilde{\mathbf{n}}}$ thus consists of independent, zero-mean Gaussian random variables with variances that differ in general.

The channel exhibits J interference epochs over the interval of the transmission, with the noise variance remaining constant within a given epoch. The duration of the j^{th} epoch is a fraction η_j of the transmission duration, where $0 < \eta_j \leq 1$ and $nL\eta_j$ is an integer for $0 \leq j \leq J-1$, and $\sum_{j=0}^{J-1} \eta_j = 1$. The noise variance within the j^{th} epoch is $\frac{N_j}{2}$. Consequently,

$$\mathrm{Var}(\tilde{n}_i) = \begin{cases} \frac{N_0}{2} & \text{if } 0 \leq i < l_0, \\ \frac{N_k}{2} & \text{if } l_{k-1} \leq i < l_k, \text{for } 1 \leq k \leq J-1. \end{cases}$$

where $l_k = \sum_{j=0}^{k} \eta_j nL$.

Two special cases of the block-interference channel are considered as examples in the thesis, though the analytical results apply to the general block-interference channel. If $J = 1$, the channel is an independent, identically distributed (i.i.d.) additive Gaussian noise channel. If $J = 2$, the noise variance is N_0 for the first $\eta_0 nL$ code symbols and is N_1 for the remaining $\eta_1 nL$ code symbols. In this case, we simplify the notation to $\eta = \eta_0$ so that η_1 is given by $1 - \eta$. Without loss of generality, $N_1 \geq N_0$ in the two-epoch channel.

The two-epoch channel can be considered as a channel in which only ambient white Gaussian noise is present for part of the transmission and a single white Gaussian interferer is also present for the remainder of the transmission. Under this premise, we refer to η as the *interference-free fraction* (of the transmission). The relative noise power in the two epochs is referred to as the *noise power ratio* of the channel; it is given by

$$\gamma = 10 \log_{10} \left(\frac{N_1}{N_0} \right) \ (dB).$$

The (ambient) *signal-to-noise ratio* in the received signal is defined as $\frac{E_b}{N_0}$.

2.2.2 Mixed-distribution channel model

The mixed-distribution channel is a (stationary) i.i.d. discrete-time channel that is non-Gaussian in general. It is intended to approximate the effect of the block-interference channel on a communication system with interleaving. The channel is thus characterized by the parameters $J, \eta_0, \ldots, \eta_{J-1}, N_0, \ldots,$ and N_{J-1} as defined in the previous subsection. It is additionally characterized by the parameter process $\underline{\tilde{\mathbf{\Theta}}}$ with

$$\Pr(\tilde{\Theta}_i = j) = \eta_j, \qquad 0 \leq j \leq J-1,\ 1 \leq i \leq nL-1$$

for the independent (and identically distributed) random variables $\tilde{\Theta}_0, \ldots, \tilde{\Theta}_{nL-1}$. The random variables $\tilde{n}_0, \tilde{n}_1, \ldots, \tilde{n}_{nL-1}$ are conditionally zero-mean, independent Gaussian random variables given $\underline{\tilde{\mathbf{\Theta}}}$ with

$$\mathrm{Var}(\tilde{n}_i \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) = \mathrm{Var}(\tilde{n}_i \mid \tilde{\Theta}_i = j) = \frac{N_j}{2}.$$

Two special cases of the mixed-distribution channel are considered as examples in this thesis, though the analytical results apply to the general mixed-distribution channel. If $J = 1$, the channel is an i.i.d. additive Gaussian noise channel. If $J = 2$, each random variable $\tilde{\Theta}_i$ follows a Bernoulli distribution with parameter η and

$$\mathrm{Var}(\tilde{n}_i) = \begin{cases} \frac{N_0}{2} & \text{with probability } \eta, \\ \frac{N_1}{2} & \text{with probability } 1-\eta \end{cases}.$$

The mixed-distribution channel with parameters $J, \eta_0, \ldots, \eta_{J-1}, N_0, \ldots,$ and N_{J-1} is said to be *equivalent to* the block-interference channel characterized by the same parameters. The terms interference-free fraction, noise power ratio, and signal-to-noise ratio (and their respective notation) are adopted from the terminology for the block-interference channel

for the special case $J = 2$. Note that the choice of the interleaver $\underline{\pi}$ has no effect on the performance of the system in the mixed-distribution channel model since the channel symbols are subjected to i.i.d. noise.

Chapter 3

Statistics used in Performance Analysis

The performance of the system under consideration does not depend on the transmitted code word with either the block-interference channel or the mixed-distribution channel. Consequently, in the remainder of the thesis, we assume without loss of generality that the all-zeros code word is transmitted. Thus

$$\hat{r}_i = a_i\sqrt{E_c} + a_i n_i \qquad \text{for all } i.$$

Two types of error statistic arise in our analysis: one related to error events, and the other related to code-word errors. Suppose

$$\underline{\mathbf{c}}_l^{l+p} = (c_{ln}, \ldots, c_{(l+p)n-1}),$$

for some $0 \leq l \leq l+p \leq L$, denotes a subsequence of a code word which corresponds to a single excursion from the all-zeros path in the code trellis. (I.e., the encoder is in state zero at the start of the l^{th} time step, it is in state zero after the $(l+p)^{th}$ time step, and its state is non-zero in the interim.) The corresponding *error event* at time l [6] is given by $\{X < 0\}$

where the pairwise error-event statistic for $\underline{\mathbf{c}}_l^{l+p}$ is given by

$$X = \sum_{ln \le i \le (l+p)n-1 : c_i = 1} \hat{r}_i \ .$$

Suppose E_l denotes the set of ν_l error events starting at time l, $0 \le l \le L-1$, and E denotes the set of all ν error events (where $\nu = \nu_0 + \ldots + \nu_{L-1}$). The error events in E_l are indexed by i, $1 \le i \le \nu_l$. The *first event error* at time l is given by

$$F_l = [\cup_{j=0}^{l-1} \cup_{i=1}^{\nu_j} \{X_{i,j} > 0\}] \cap [\cup_{i=1}^{\nu_l} \{X_{i,l} < 0\}] \tag{3.1}$$

where $X_{i,l}$ is the pairwise error-event statistic of the i^{th} error event in E_l. Similarly, the *event error* at time l is given by

$$G_l = [\cup_{i=1}^{\nu_l} \{X_{i,l} < 0\}] \ . \tag{3.2}$$

(Note that $F_0 = G_0$.) In many instances we focus on error events and first event errors at time zero in which case we omit the time reference if there is no resulting ambiguity. The *pairwise error-event probability* of $\underline{\mathbf{c}}_l^{l+p}$ is given by $\Pr(X < 0)$. Similarly, the *first-event error probability* at time l is given by $\Pr(F_l)$, and the *event error probability* at time l is given by $\Pr(G_l)$. The event error probability at time zero and the first-event error probability at time zero are equal, and they are denoted by p for simplicity in the development in later chapters.

Similarly, if $\underline{\mathbf{c}}$ is a valid non-zero code word, the *pairwise code-word error probability* is given by

$$\Pr(X(\underline{\mathbf{c}}) < 0)$$

where the code-word error statistic is given by

$$\begin{aligned} X(\underline{\mathbf{c}}) &= \frac{1}{2}\left(Z(\underline{\mathbf{0}}) - Z(\underline{\mathbf{c}})\right) \\ &= \sum_{0 \le i \le nL-1 : c_i = 1} \hat{r}_i \ . \end{aligned}$$

(Note that a non-zero code word may contain several error events.) The probability of code-word error, P_e, can be expressed in either of two ways:

$$P_e = \Pr(\cup_{\underline{\mathbf{c}} \in \mathcal{C} : \underline{\mathbf{c}} \neq \underline{\mathbf{0}}} \{X(\underline{\mathbf{c}}) < 0\}), \tag{3.3}$$

or

$$P_e = \Pr(\cup_{l=0}^{L-1} \cup_{i=1}^{\nu_l} \{X_{i,l} < 0\}). \tag{3.4}$$

Under the condition that the all-zeros code word is transmitted, the random variables $\{X(\underline{\mathbf{c}}) | \underline{\mathbf{c}} \neq \underline{\mathbf{0}}\}$ are jointly Gaussian and the random variables $\{X_{i,l} : 0 \le l \le L-1, 1 \le i \le \nu_l\}$ are also jointly Gaussian. For both collections of random variables, each pair of random variables has a non-negative covariance.

Chapter 4

Interleaver Designs for the Block-Interference Channel

Many time-varying channels of practical interest (including the block-interference channels) produce a sequence of several consecutive channel outputs which result from poorer than average channel conditions. This leads to clustering of low-quality channel outputs at the decoder's input, which in turn results in a high probability of error at the decoder's output if the system uses convolutional coding and Viterbi decoding without interleaving. Interleaving is intended to distribute the low-quality channel outputs evenly across the sequence of inputs to the Viterbi decoder, which results in better decoder performance in general.

The choice of interleaver in Fig. 2.1 is an important design decision that can have a significant impact on the system's performance. In this chapter, we consider the effect of the interleaver design on the probability of code-word error, P_e, at the output of the decoder. Two commonly used interleaver designs are considered in this chapter: the rectangular interleaver [6] and the pseudo-random interleaver [9]. The performance with each interleaver is illustrated using examples of the block-interference channel with two interference epochs.

4.1 Rectangular Interleaver

The *rectangular interleaver* re-orders code symbols based on their placement in an M-by-N rectangular matrix, where $MN = nL$. The code symbols are written column-wise into the interleaving matrix (beginning with the left-most column) in the order in which they are produced by the encoder. They are read row-wise out of the matrix (beginning with the top row) and transmitted across the channel in that order. We investigate the probability of code-word error of convolutional codes on the block-interference channel for different dimensions of the interleaving matrix.

The performance of the system using the rate-$\frac{1}{2}$, memory-order-three convolutional encoder is shown in Figs. 4.1-4.3 for each of several choices of the interleaver's row dimension and for various two-epoch channels. The block length of each code word is 1000 (so that the code word contains 2000 code symbols), and the noise power ratio of each channel is $\gamma = 3$ dB.

The performance of the system is highly sensitive to the choice of the interleaver if the interference-free fraction $\eta = 0.75$ (i.e., 25% of the symbols are subjected to interference), as illustrated in Fig. 4.1. For example, there is a difference in performance of approximately 2 dB between the best and the poorest choice of interleaver at a probability of code-word error of $P_e = 10^{-3}$. The relationship between the row dimension (M) and the performance is not simple, however. The poorest performance results from no interleaving. It improves as M is increased to 5, then degrades as M is increased to 10, improves again as M is increased to 20, and degrades with further increase in M. The small multiplicity of low-weight error events (all of similar spans) in the memory-order-three code appears to be responsible for this complicated sensitivity to the precise interleaver structure.

The effect of the interleaver's structure on performance diminishes as the interference-free fraction is reduced. This is illustrated in Figs. 4.2 and 4.3 for $\eta = 0.5$ and $\eta = 0.25$, respectively. The difference in performance using the best and the poorest rectangular interleavers is 1 dB if $\eta = 0.5$, but the difference is only 0.5 dB if $\eta = 0.25$. Furthermore, the

occurrence of dual local minima in P_e as a function of M is less pronounced if $\eta = 0.5$ than if $\eta = 0.75$, and the second local minimum does not occur if $\eta = 0.25$.

The performance of the system using the NASA-standard code is shown in Figs. 4.4-4.6 for the channels considered in Figs. 4.1-4.3, respectively. The difference in the probability of code-word error using the best and the worst rectangular interleavers with each channel is similar to the corresponding difference with the memory-order-three code. There is no occurrence of the dual local minima observed with the weaker code, however; in fact, the performance changes very little as M is varied between 5 and 20. The large multiplicity of low-weight error events (many with very different spans) in the more powerful code appears to be responsible for the reduced sensitivity to the precise interleaver structure.

4.2 Pseudo-Random Interleaver

The *pseudo-random interleaver* re-orders the code symbols before transmission on the channel according to a sequence generated by a pseudo-random-number generator algorithm. The receiver has the knowledge of the pseudo-random interleaving pattern used in the transmission of each code word; thus it can perform the de-interleaving operation on the received word. We consider several example pseudo-random interleavers for a code-word of length 2000 symbols, each obtained via a pseudo-random-number generator. The probability of code-word error is evaluated as a function of the signal-to-noise ratio for each interleaver, each of the two codes and each of the three channels considered in the previous section. For a given code and channel, the performance plots with the different interleavers are indistinguishable. This indicates that, with a high probability, the performance obtained with a pseudo-random interleaver is close to the average performance obtained with a *uniform interleaver*. A uniform interleaver is an interleaver that is selected at random among all possible interleavers (of the relevant block size) according to a uniform distribution [10].

4.3 Comparison of the Two Interleaving Schemes

A pseudo-random interleaver typically yields performance comparable to the performance obtained with the optimal rectangular interleaver. This is illustrated in Figs. 4.7 and 4.8 for an example of the block-interference channel and the two encoders with a block length of 1000. The channel has an interference-free fraction of one-half, and the noise power ratio is 3 dB. The pseudo-random interleaver is one of those mentioned in the previous section.

The probability of code-word error as a function of the signal-to-noise ratio is shown in Fig. 4.7 for the memory-order-three code using the pseudo-random interleaver and two choices of the rectangular interleaver: $M = 5$ (the best choice), and $M = 10$. The performance using the pseudo-random interleaver is within 0.15 dB of the performance using the best rectangular interleaver if $P_e = 10^{-3}$. The same comparison is illustrated in Fig. 4.8 for the system using the NASA-standard code. The pseudo-random interleaver results in better performance than the best rectangular interleaver.

The pseudo-random interleaver has the further advantage that it can be used effectively with any encoder (and by truncation, with any block length), whereas the optimal row dimension of the rectangular interleaver differs between codes and for different block lengths in general. The pseudo-random interleaver and de-interleaver have the disadvantage of requiring unstructured memory accesses (in comparison with structured accesses for a rectangular interleaver). De-interleaver memory access is unlikely to be the limiting factor in decoder performance, however. In addition, the uniform interleaver allows for tractable analytical results in the performance evaluation of the system; hence, the subsequent work deals with the analysis of the performance of convolutional codes with a uniform interleaver.

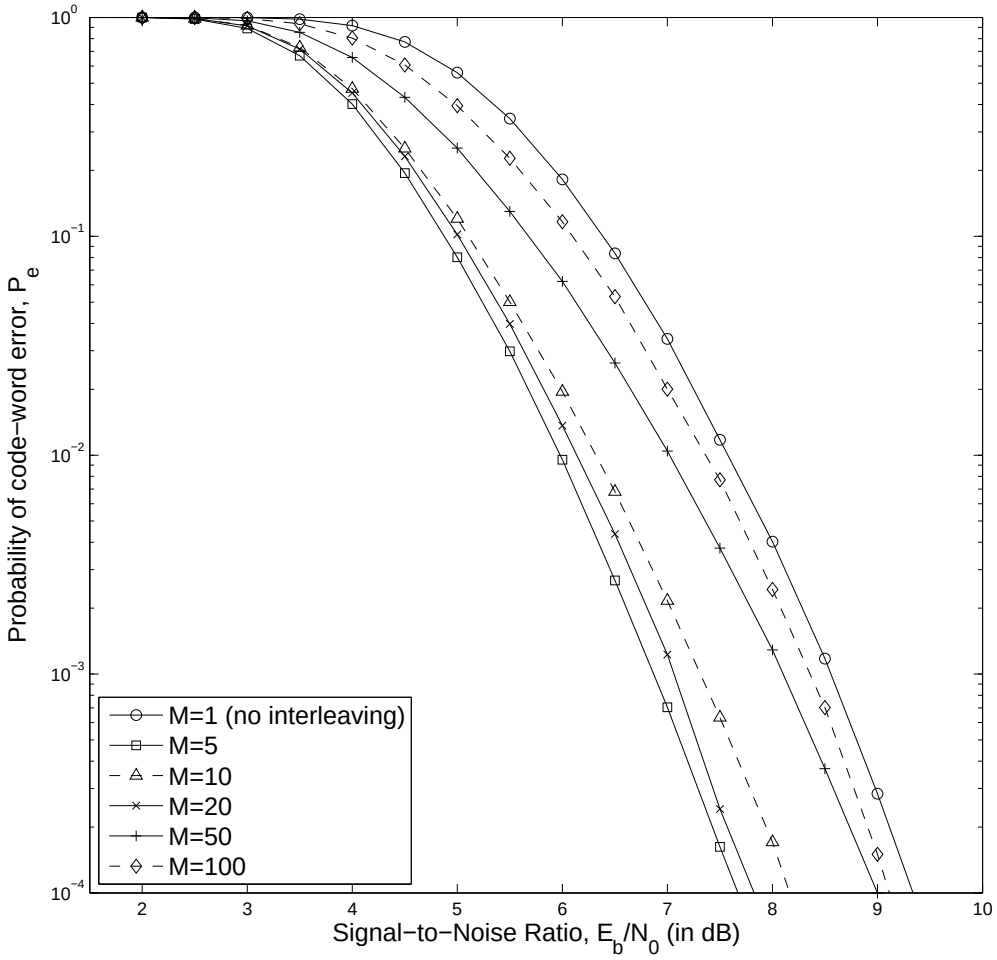

Figure 4.1: Performance with the memory-order-three code and rectangular interleaving ($\eta = 0.75$).

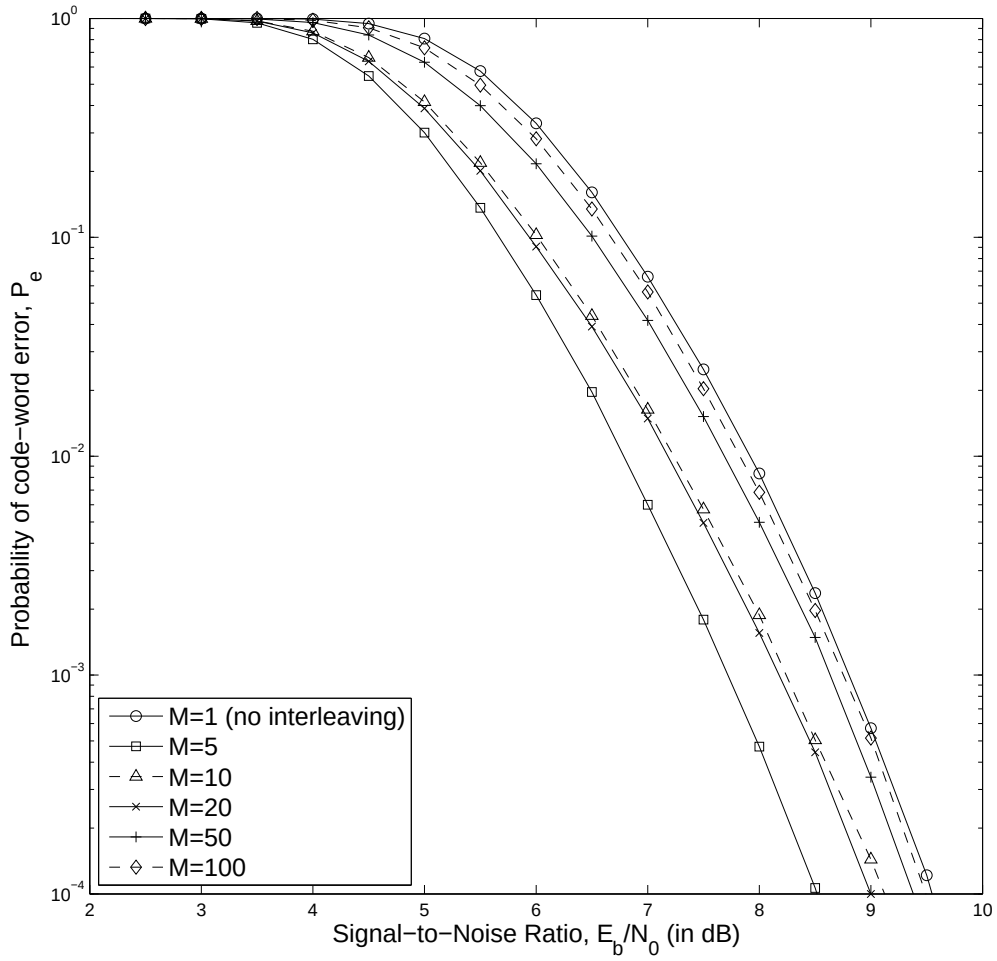

Figure 4.2: Performance with the memory-order-three code and rectangular interleaving ($\eta = 0.5$).

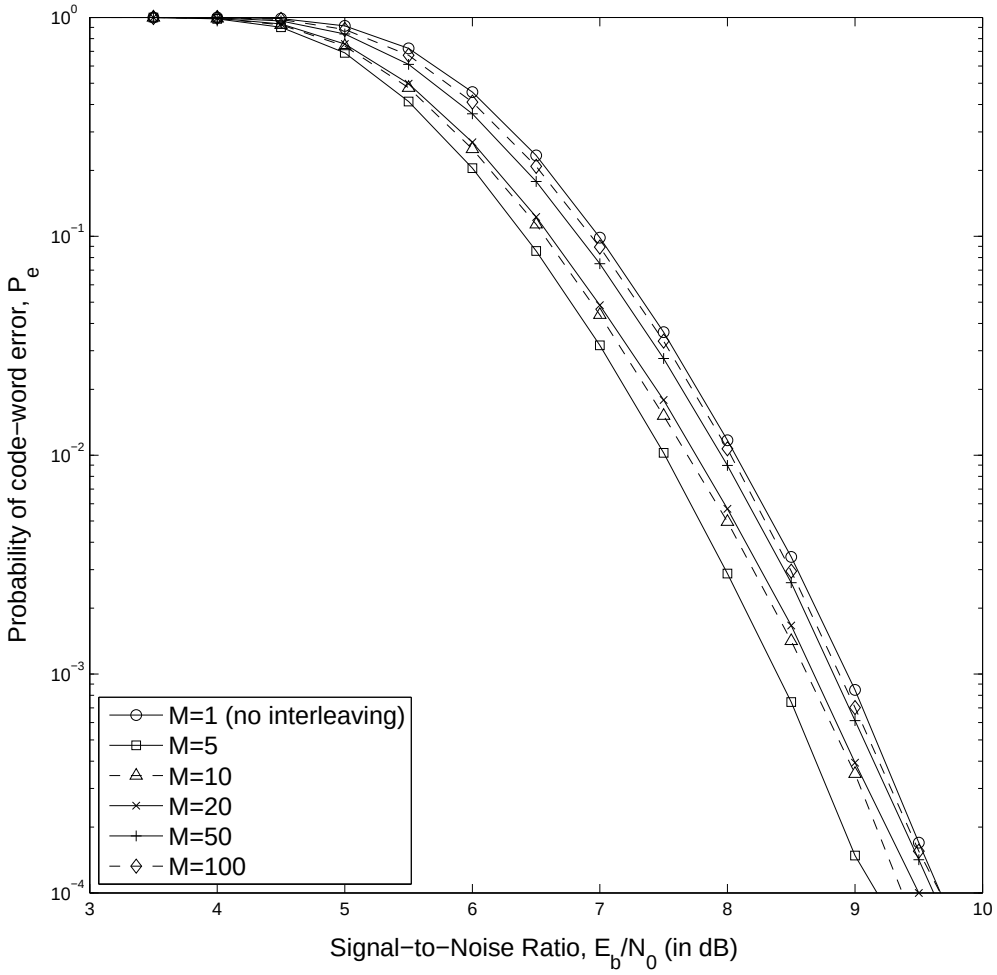

Figure 4.3: Performance with the memory-order-three code and rectangular interleaving ($\eta = 0.25$).

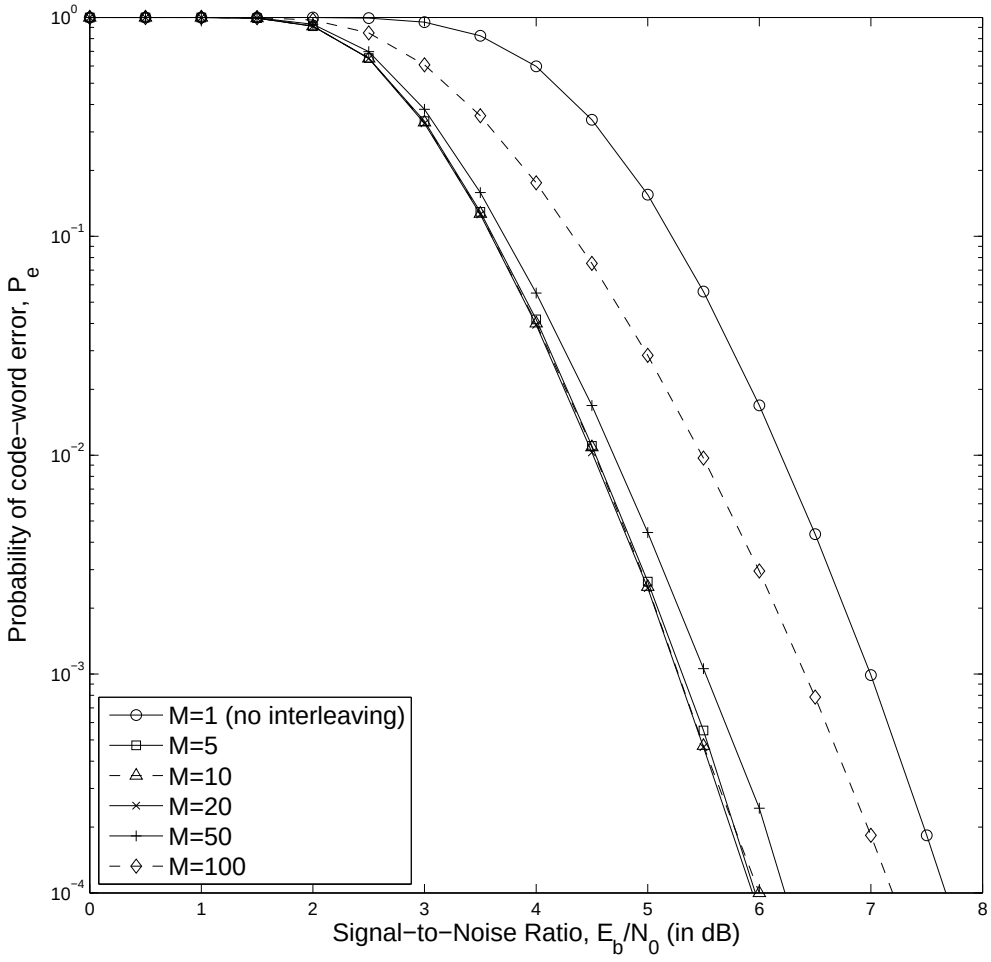

Figure 4.4: Performance with the NASA-standard code and rectangular interleaving ($\eta = 0.75$).

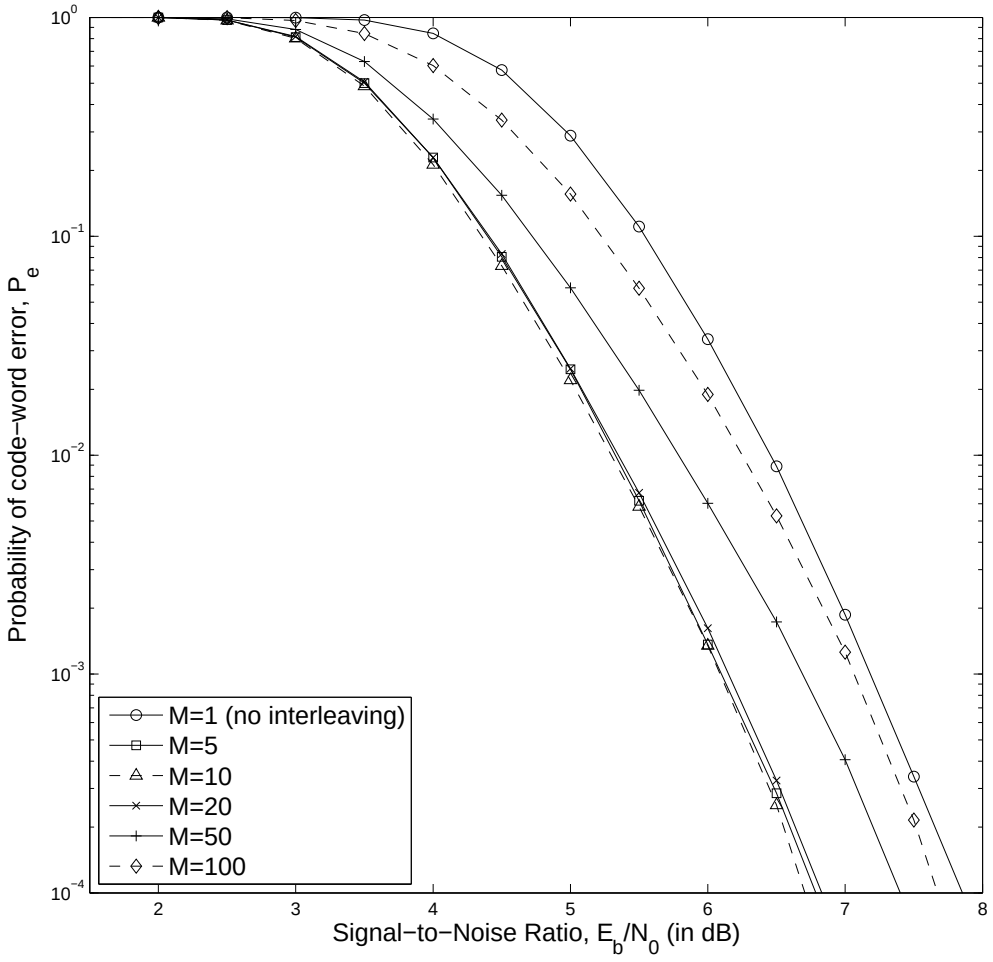

Figure 4.5: Performance with the NASA-standard code and rectangular interleaving ($\eta = 0.5$).

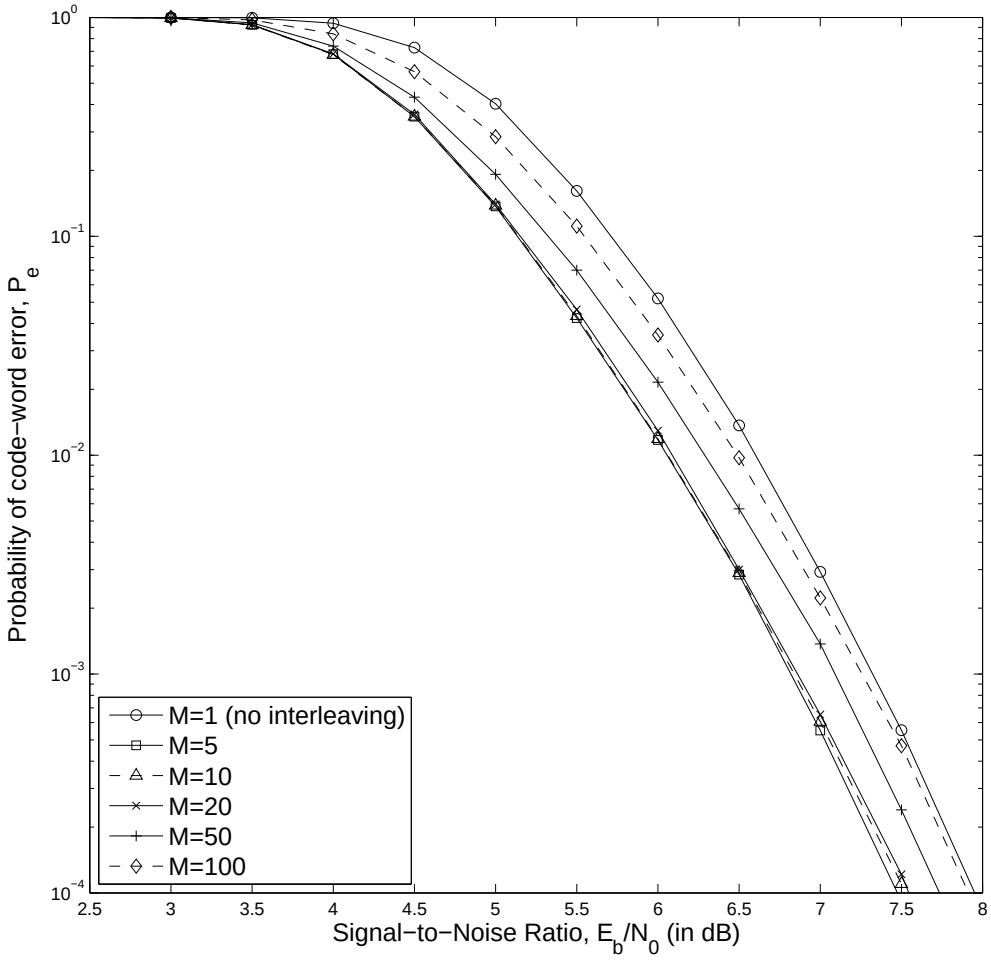

Figure 4.6: Performance with the NASA-standard code and rectangular interleaving ($\eta = 0.25$).

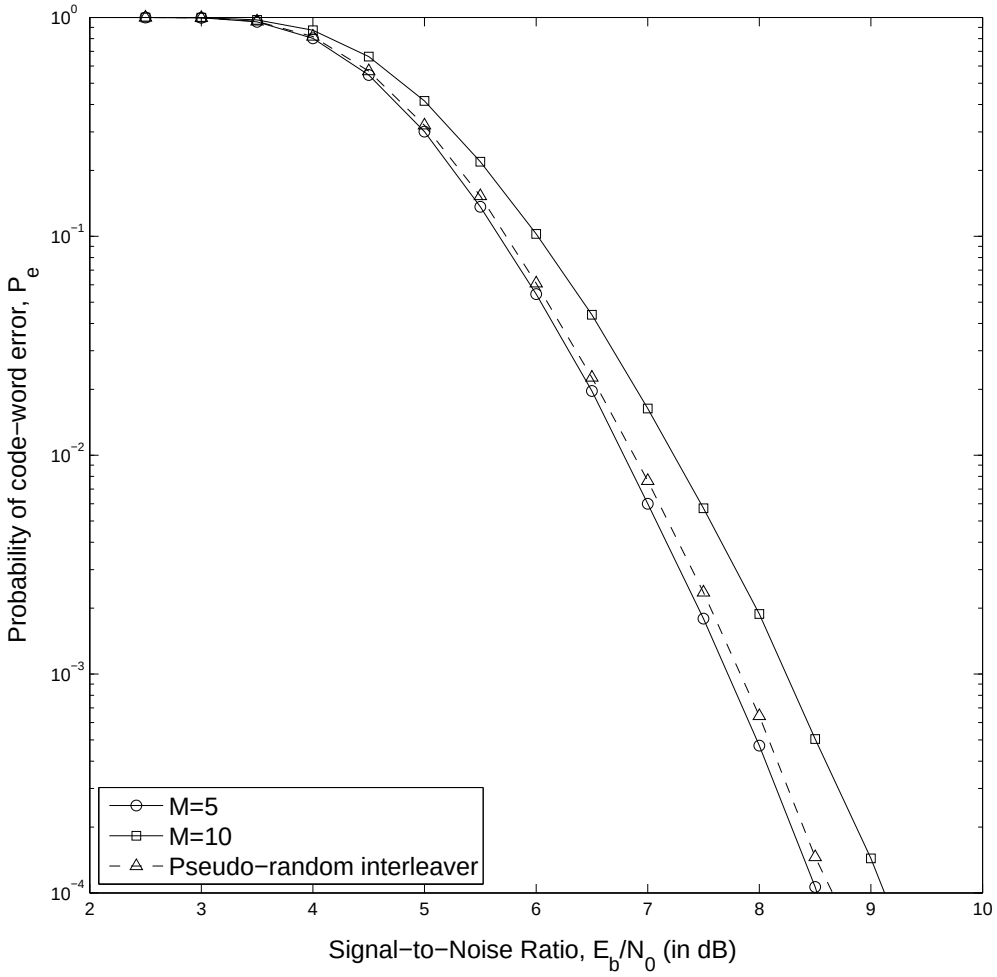

Figure 4.7: Performance with rectangular and pseudo-random interleavers for the memory-order-three code ($\eta = 0.5$).

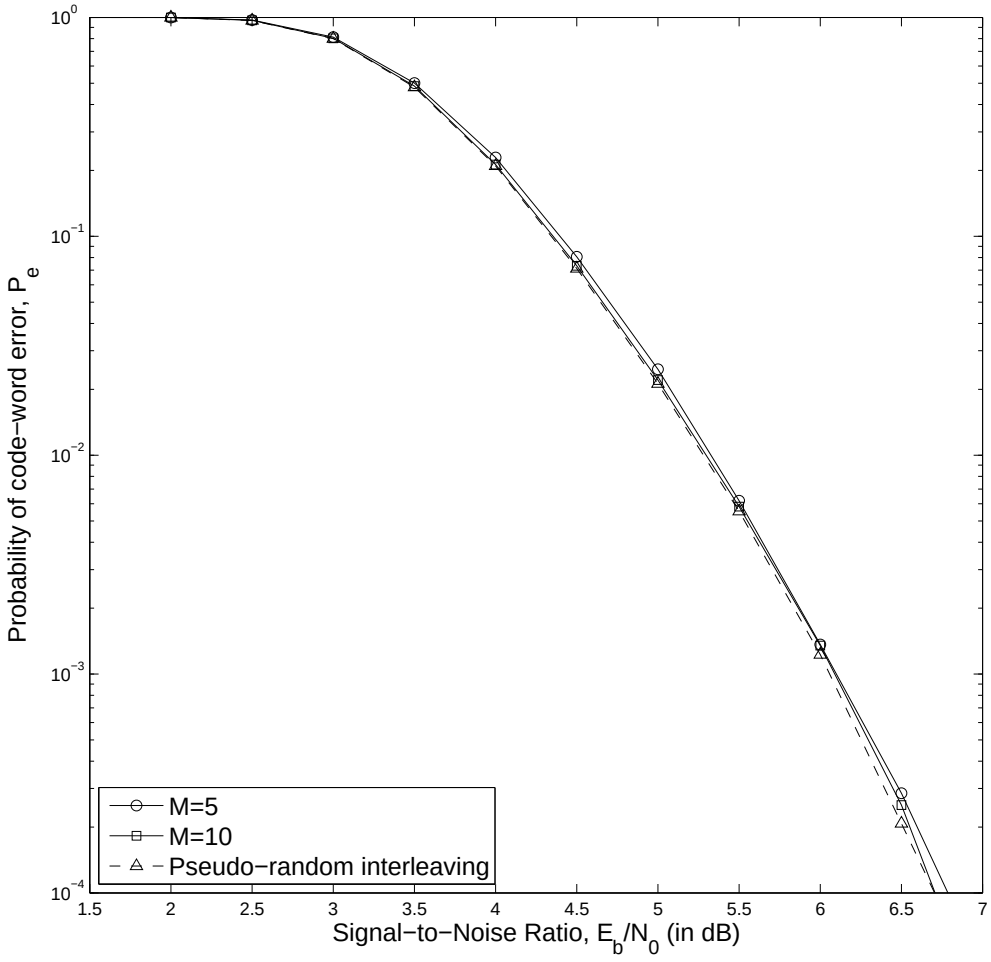

Figure 4.8: Performance with rectangular and pseudo-random interleavers for the NASA-standard code ($\eta = 0.5$).

Chapter 5

Mixed-Distribution Approximation to the Block-Interference Channel

In this chapter we show the asymptotic equality of the performance of a system with a block-interference channel and a uniform interleaver and the performance of a system with the equivalent mixed-distribution channel. We define the *channel state* for each code symbol in a code word as the interference epoch in which it lies in the block-interference channel. We show that the joint distribution of the channel states for any fixed set of code symbols with a uniform interleaver approaches their joint distribution with the equivalent mixed-distribution channel. A similar result is shown for each pair-wise error-event probability as well. In addition, it is shown by simulation that the performance with the block-interference channel and a typical pseudo-random interleaver is close to the performance in the equivalent mixed-distribution channel.

5.1 Joint Distribution of Channel States

Consider the J-epoch block-interference channel with parameters $\eta_0, \ldots, \eta_{J-1}$ and $N_0, \ldots, N_{J-1}$ and any d symbols from the de-interleaved received word $\underline{\mathbf{r}}$ of length nL symbols. Let the interference epoch in which the channel experiences noise density N_j be

represented by A_j for $0 \leq j \leq J-1$. Also, let $q_0, \ldots, q_{J-1}$ be non-negative integers such that $q_j \leq \eta_j nL$ for all $0 \leq j \leq J-1$ and $\sum_{j=0}^{J-1} q_j = d$. With the uniform interleaver,

$$\Pr(q_0 \text{ symbols} \in A_0, \ldots, q_{J-1} \text{ symbols} \in A_{J-1}) = \frac{\binom{\eta_0 nL}{q_0} \cdots \binom{\eta_{J-1} nL}{q_{J-1}}}{\binom{nL}{d}} \quad ,$$

which is a multivariate hypergeometric distribution [11].

For a given d, as the block length L increases,

$$\begin{aligned}
&\lim_{L\to\infty} \Pr(q_0 \text{ symbols} \in A_0, \ldots, q_{J-1} \text{ symbols} \in A_{J-1}) \\
&\quad = \lim_{L\to\infty} \frac{\prod_{j=0}^{J-1} \binom{\eta_j nL}{q_j}}{\binom{nL}{d}} \\
&\quad = \lim_{L\to\infty} \frac{\prod_{j=0}^{J-1} \frac{\eta_j nL \cdots (\eta_j nL - q_j + 1)}{q_j(q_j-1)\cdots 1}}{\frac{nL(nL-1)\cdots(nL-d+1)}{d(d-1)\cdots 1}} \\
&\quad = \lim_{L\to\infty} \frac{\prod_{j=0}^{J-1} \frac{(\eta_j nL)^{q_j}(1\cdots(1-\frac{q_j-1}{\eta_j nL}))}{q_j!}}{\frac{(nL)^d(1\cdot(1-\frac{1}{nL})\cdots(1-\frac{d-1}{nL}))}{d!}} \\
&\quad = \frac{d!}{q_0!\ldots q_{J-1}!} \frac{\prod_{j=0}^{J-1}(\eta_j nL)^{q_j}}{(nL)^d} \lim_{L\to\infty} \frac{\prod_{j=0}^{J-1} 1\cdots(1-\frac{q_j-1}{\eta_j nL})}{1\cdot(1-\frac{1}{nL})\cdots(1-\frac{d-1}{nL})} \\
&\quad = \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} \qquad (5.1)
\end{aligned}$$

which is a multinomial distribution [11].

The latter is the expression for the corresponding distribution in the equivalent mixed-distribution channel. Hence, as the block length approaches infinity, the joint distribution of the channel states for any fixed set of symbol positions in the block-interference channel with uniform interleaving approaches their joint distribution in the equivalent mixed-distribution channel. For the special case of the two-epoch ($J = 2$) channel with the interference-free duration η, the above distribution simplifies to

$$\begin{aligned}\Pr(q \text{ symbols} \in A_0, \ldots, d-q \text{ symbols} \in A_1) &= \frac{d!}{q!(d-q)!}\eta^q(1-\eta)^{(d-q)} \\ &= \textbf{Binomial}(d, \eta).\end{aligned}$$

5.2 Pairwise Error-Event Probability

The pairwise error-event probability of an error event with Hamming weight d in an AWGN channel with two-sided power spectral density $\frac{N_0}{2}$ is given by [4],

$$P_d = Q\left(\sqrt{\frac{2dE_c}{N_0}}\right).$$

In case of the J-epoch block-interference channel, the pairwise error-event probability also depends on how the non-zero code symbols in the error event are distributed among the J epochs. Suppose the weighted code subsequence $\underline{\mathbf{c}}_l^{l+p}$ represents an excursion from state zero with weight q_j among its code symbols transmitted in the j^{th} epoch, $0 \leq j \leq J-1$, so that $\sum_{j=0}^{J-1} q_j = d$. If X denotes the pairwise error-event statistic for $\underline{\mathbf{c}}_l^{l+p}$, the pairwise error-event probability is

$$P_{q_0,\ldots,q_{J-1}} = \Pr(X < 0) = \Pr\Big(\sum_{i:c_i=1} \hat{r}_i < 0\Big).$$

The mean and variance of X are given by

$$\begin{aligned} E[X] &= a_0 q_0 \sqrt{E_c} + a_1 q_1 \sqrt{E_c} + \ldots + a_{J-1} q_{J-1} \sqrt{E_c} \\ &= \sqrt{E_c} \left(\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}} \right) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(X) &= \frac{a_0^2 q_0 N_0}{2} + \ldots + \frac{a_{J-1}^2 q_{J-1} N_{J-1}}{2} \\ &= \frac{1}{2} \Big(\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}} \Big). \end{aligned}$$

Since X is a Gaussian random variable,

$$\begin{aligned} P_{q_0, \ldots q_{J-1}} &= Q\left(\frac{E[X]}{\sqrt{\mathrm{Var}(X)}} \right) \\ &= Q\left(\sqrt{2E_c} \sqrt{\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}}} \right). \end{aligned} \tag{5.2}$$

Thus for a given error event of Hamming weight d and block length L, the pairwise error-event probability with a uniform interleaver is

$$P_d(L) = \sum_{q_0=0}^{\min\{\eta_0 nL, d\}} \sum_{q_1=0}^{\min\{\eta_1 nL, d-q_0\}} \cdots \sum_{q_{J-2}=0}^{\min\{\eta_{J-2} nL, d-q_0-\ldots-q_{J-3}\}} \frac{\binom{\eta_0 nL}{q_0} \cdots \binom{\eta_{J-1} nL}{q_{J-1}}}{\binom{nL}{d}} P_{q_0, \ldots q_{J-1}}.$$

The quantity $P_{q_0,\ldots,q_{J-1}}$ is not a function of the block length. Therefore,

$$
\begin{aligned}
P_d &= \lim_{L\to\infty} P_d(L) \\
&= \lim_{L\to\infty} \sum_{q_0=0}^{\min\{\eta_0 nL,d\}} \sum_{q_1=0}^{\min\{\eta_1 nL,d-q_0\}} \cdots \sum_{q_{J-2}=0}^{\min\{\eta_{J-2} nL,d-q_0-\ldots-q_{J-3}\}} \frac{\binom{\eta_0 nL}{q_0}\cdots\binom{\eta_{J-1} nL}{q_{J-1}}}{\binom{nL}{d}} P_{q_0,\ldots q_{J-1}} \\
&= \sum_{q_0=0}^{d}\sum_{q_1=0}^{d-q_0}\cdots\sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} P_{q_0,\ldots q_{J-1}} \lim_{L\to\infty} \frac{\binom{\eta_0 nL}{q_0}\cdots\binom{\eta_{J-1} nL}{q_{J-1}}}{\binom{nL}{d}} \\
&= \sum_{q_0=0}^{d}\sum_{q_1=0}^{d-q_0}\cdots\sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} P_{q_0,\ldots q_{J-1}} \binom{d}{q_0\ldots q_{J-1}} \prod_{j=0}^{J-1} \eta_j^{q_j} \ , \qquad (5.3)
\end{aligned}
$$

where the last step follows from equation (5.1). Hence, the pairwise error-event probability for a given error event with the block-interference channel and the uniform interleaver approaches the pairwise error-event probability for the same error event in the equivalent mixed-distribution channel as the block length approaches infinity. For the special case of the two-epoch channel with interference-free duration η, the pairwise error-event probability simplifies to

$$
\begin{aligned}
P_d &= \sum_{i=0}^{d} \Pr(q=i) P_{q,d-q} \\
&= \sum_{i=0}^{d} \binom{d}{i} \eta^i (1-\eta)^{d-i} Q\left(\sqrt{\frac{2E_c}{N_0}}\sqrt{i+(d-i)\frac{N_0}{N_1}}\right). \qquad (5.4)
\end{aligned}
$$

Equations (5.2), (5.3), and (5.4) are equally applicable to the pairwise code-word error probability for a code word of Hamming weight d. We will assume in the analytical development in Chapters 5 and 6 that the block length is sufficiently large that the asymptotic result in equation (5.3) is highly accurate. We will use it in developing the bounds in both chapters.

5.3 Probability of Code-Word Error

Figs. 5.1 and 5.2 compare the probability of code-word error for one choice of pseudo-random interleaver in the block-interference channel with the performance in the equivalent mixed-distribution channel for the memory-order-three encoder and the NASA-standard encoder, respectively, with a block length of 1000. The pseudo-random interleaver is one of the randomly generated interleavers discussed in Section 4.2. The two-epoch channel considered in the simulation has an interference-free fraction of 0.5 and the noise power ratio γ is 3 dB. The figures also include the performance with a rectangular interleaver using a 1000-by-2 interleaving matrix. As observed in Section 4.1, this choice of dimension of the interleaving matrix gives approximately the worst performance among all rectangular interleaver matrices; hence, it serves as an example of the performance resulting from a poor choice of the interleaver. As suggested by the analysis in the previous sections, the simulation results show that performance in the mixed-distribution channel matches closely the performance in the block-interference channel with a typical pseudo-random interleaver.

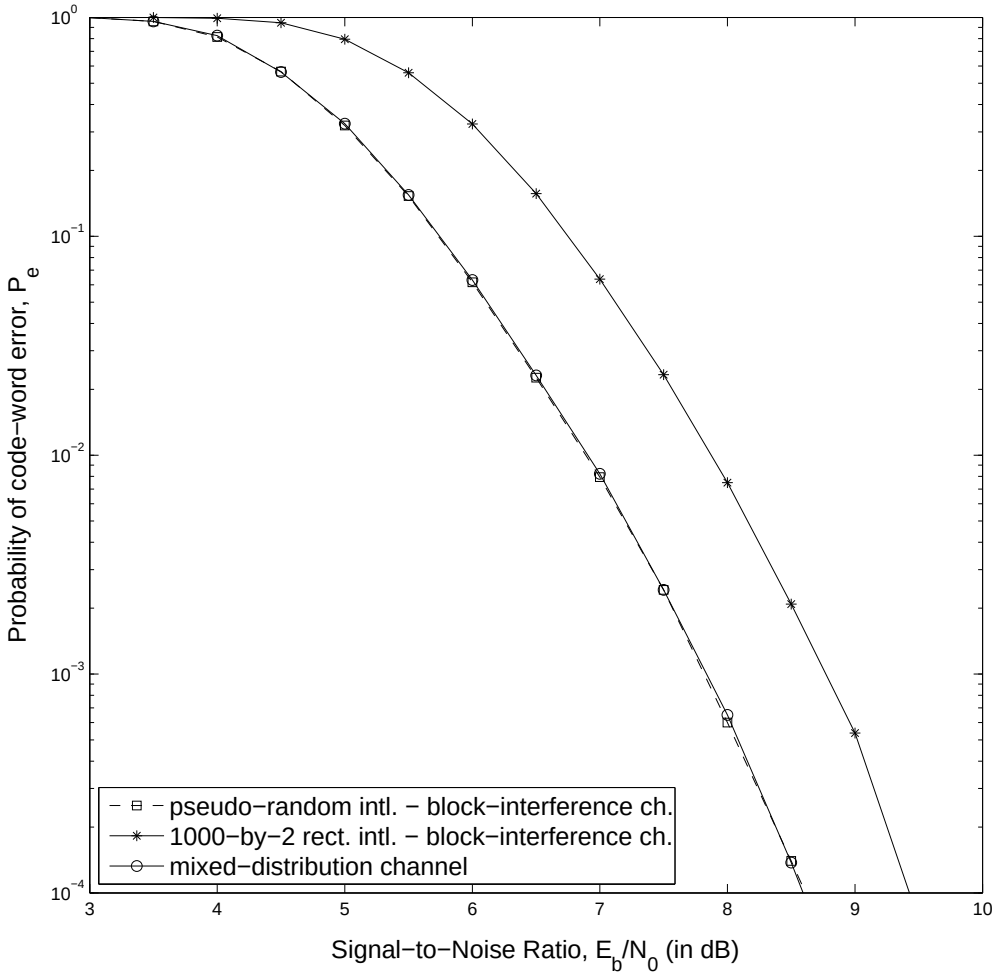

Figure 5.1: Performance in the block-interference and mixed-distribution channels for the memory-order-three encoder ($\eta = 0.5$).

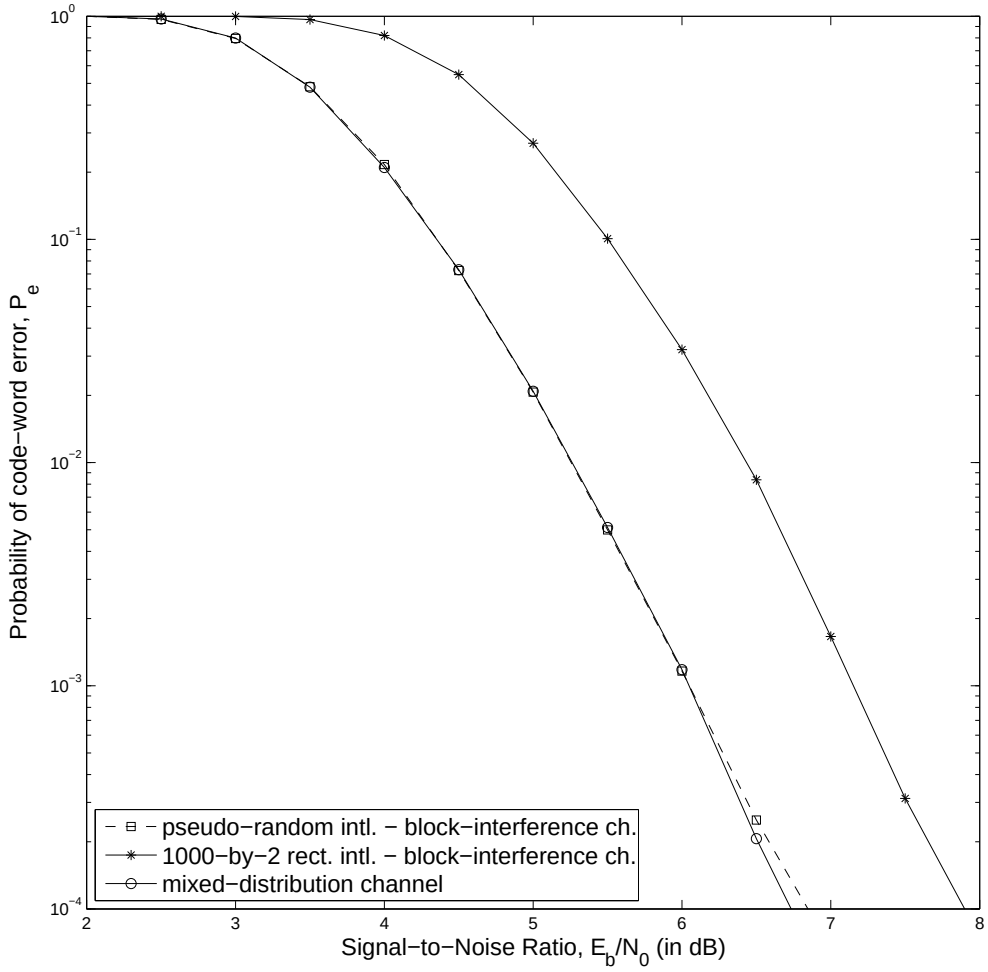

Figure 5.2: Performance in the block-interference and mixed-distribution channels for the NASA-standard encoder ($\eta = 0.5$).

Chapter 6

Bounds on the Probability of Code-word Error with Soft-Decision Viterbi Decoding

In this chapter, we derive tight bounds on the probability of code-word error for the system of Fig. 2.1 in the mixed-distribution channel. As illustrated in Chapter 5, the performance in the mixed-distribution channel closely approximates the performance in the block-interference channel with a uniform interleaver. The bounds derived in this chapter thus also serve approximately as bounds on the performance in the block-interference channel with a typical pseudo-random interleaver.

6.1 Union Bounds Using the Code-Word Weight-Enumerating Function

The probability of code-word error can be bounded in terms of the pairwise code-word error probabilities. From equation (3.3),

$$\begin{aligned} P_e &= \Pr(\cup_{\underline{\mathbf{c}} \in \mathcal{C}: \underline{\mathbf{c}} \neq \underline{\mathbf{0}}} \{X(\underline{\mathbf{c}}) < 0\}) \\ &\leq \sum_{\underline{\mathbf{c}} \in \mathcal{C}: \underline{\mathbf{c}} \neq \underline{\mathbf{0}}} \Pr(X(\underline{\mathbf{c}}) < 0) \\ &= \sum_{d} A_d^{'} P_d \end{aligned}$$

where $A_d^{'}$ is the number of code words of length nL and Hamming weight d and P_d is the pairwise code-word error probability for a code word of Hamming weight d. This union bound on P_e is the (code word) *weight-enumerator bound*.

The number of code words of each Hamming weight is obtained by noting that each possible non-zero detected code word results from one or more error events. The error-event weight-span enumerating function of the code, $A(W, L)$, gives the number of error events of each Hamming weight W and span L that start at time zero [6]. It can be determined from Mason's Theorem applied to the encoder's modified state-event diagram with each branch labeled with a bivariate label that denotes the Hamming weight and span (always one) associated with the corresponding state transition [12].

Using the technique in [10], we can determine the number of code words with multiple error events using the corresponding weight-enumerating functions. Specifically, an auxiliary code-word weight-span enumerating function

$$A^{(i)}(W, L) = [A(W, L)]^i$$

is used which enumerates the non-zero code words of each Hamming weight and total error-event span that consist of i error events that occur consecutively beginning at time zero.

Accounting for non-zero code words in which the error events do not all occur consecutively, there are

$$c[l,i] = \binom{nL-l+i}{i}$$

code words consisting of i error events of total error-event span l for each code word counted in $A^{(i)}(W,L)$. Thus the weight-enumerating function is given by

$$\begin{aligned} A(W) &= \sum A_d' W^d \\ &= \left(\sum_i \sum_l c[l,i] A^{(i)}(W,L) \right) \Bigg|_{L=1} \\ &= \left(\sum_l c[l,1] A^{(1)}(W,L) + \sum_l c[l,2] A^2(W,L) + \ldots \right) \Bigg|_{L=1} \end{aligned} \tag{6.1}$$

The weight-enumerating function has finitely many terms since the block length L is finite. Unless the block length is very small, however, it is impractical to determine A_d' for large values of d using this method. Thus in practice, the weight-enumerator bound is approximated by considering only terms with small values of i in equation (6.1). For sufficiently large signal-to-noise ratios,

$$P_e \lessapprox \sum_{d=d_{free}}^{d'} A_d' P_d$$

where d_{free} is the minimum free Hamming distance of the code [6] and d' is somewhat larger than d_{free}. Thus enough terms in equation (6.1) are utilized to account for all code words of Hamming weight d' or less in practice. Note that the approximation is not guaranteed to provide an upper bound on P_e.

Even if all the terms in $A(W)$ are known, the resulting expression for the weight-enumerator bound using the exact expression for P_d is difficult to evaluate. Two bounds on P_d and one alternative form of expression for P_d can be used to simplify the evaluation of the weight-enumerator bound or its approximation.

6.1.1 Chernoff Weight-Enumerator Bound

An upper bound on the pairwise error-event probability or the code-word error-event probability P_d given by equation (5.3) can be obtained using the Chernoff bound [6]. It is given by

$$\begin{aligned}
P_d &= \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} Q\left(\sqrt{2E_c}\sqrt{\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}}}\right) \\
&\leq \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} \exp\Big(-E_c\big(\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}}\big)\Big) \\
&= \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \Big(\eta_0 \exp\Big(-\frac{E_c}{N_0}\Big)\Big)^{q_0} \ldots \Big(\eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}}\Big)\Big)^{q_{J-1}} \\
&= \left(\eta_0 \exp\Big(-\frac{E_c}{N_0}\Big) + \ldots + \eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}}\Big)\right)^d .
\end{aligned}$$

Consequently, the *Chernoff-weight-enumerator* bound on the probability of code-word error is

$$\begin{aligned}
P_e &\leq \sum_d A_d' P_d \\
&\leq \sum_d A_d' \left(\eta_0 \exp\Big(-\frac{E_c}{N_0}\Big) + \ldots + \eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}}\Big)\right)^d \\
&= A(W)\Big|_{W=\eta_0 \exp\left(-\frac{E_c}{N_0}\right)+\ldots+\eta_{J-1}\exp\left(-\frac{E_c}{N_{J-1}}\right)} .
\end{aligned}$$

In practice, the bound is approximated by truncating the infinite series.

6.1.2 Tighter Chernoff Weight-Enumerator Bound

The bound on P_d can be improved by using the improved Chernoff bound [13],

$$Q(\sqrt{x+y}) \leq Q(\sqrt{x}) \exp\Big(-\frac{y}{2}\Big) .$$

Then

$$P_d = \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} Q\left(\sqrt{2E_c} \sqrt{\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}}} \right)$$
$$\leq Q\left(\sqrt{\frac{2d_{free}E_c}{N_{max}}} \right) \exp\left(\frac{E_c d_{free}}{N_{max}} \right) \left(\eta_0 \exp\Big(-\frac{E_c}{N_0} \Big) + \ldots + \eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}} \Big) \right)^d$$

where $N_{max} = \max_{0 \leq j \leq J-1} N_j$. Consequently, the *tighter Chernoff-weight-enumerator* bound on the probability of code-word error is

$$P_e \leq \sum_d A_d' P_d$$
$$\leq \sum_d A_d' Q\left(\sqrt{\frac{2d_{free}E_c}{N_{max}}} \right) \exp\left(\frac{E_c d_{free}}{N_{max}} \right) \left(\eta_0 \exp\Big(-\frac{E_c}{N_0} \Big) + \ldots + \eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}} \Big) \right)^d$$
$$= Q\left(\sqrt{2d_{free}\frac{E_c}{N_{max}}} \right) \exp\left(d_{free}\frac{E_c}{N_{max}} \right) A(W)\Big|_{W=\eta_0 \exp\left(-\frac{E_c}{N_0}\right)+\ldots+\eta_{J-1}\exp\left(-\frac{E_c}{N_{J-1}}\right)} .$$

In practice, the bound is approximated by truncating the infinite series.

6.1.3 Integral Form of the Weight-Enumerator Bound

The expression for the pairwise error-event probability or the code-word error-event probability P_d can be rewritten using an identity from [14],

$$Q(x) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \exp\Big(-\frac{x^2}{2\sin^2\theta} \Big) d\theta \ .$$

Thus

$$P_d = \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} Q\left(\sqrt{2E_c} \sqrt{\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}}} \right)$$
$$= \frac{1}{\pi} \sum_{q_0=0}^{d} \sum_{q_1=0}^{d-q_0} \cdots \sum_{q_{J-2}=0}^{d-q_0-\ldots-q_{J-3}} \binom{d}{q_0 \ldots q_{J-1}} \eta_0^{q_0} \ldots \eta_{J-1}^{q_{J-1}} \int_0^{\frac{\pi}{2}} \exp\left(\frac{-E_c}{\sin^2\theta} \Big(\frac{q_0}{N_0} + \ldots + \frac{q_{J-1}}{N_{J-1}} \Big) \right) d\theta$$
$$= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \left(\eta_0 \exp\Big(-\frac{E_c}{N_0 \sin^2\theta} \Big) + \ldots + \eta_{J-1} \exp\Big(-\frac{E_c}{N_{J-1}\sin^2\theta} \Big) \right)^d d\theta \ .$$

Note that the above expression gives the exact pairwise error-event probability. Hence this expression along with the code-word weight-enumerator function yields an exact integral-form expression for the union bound on the code-word error probability. We refer to it as the *integral weight-enumerator* bound. It is given by

$$\begin{aligned} P_e &\leq \sum_d A_d' P_d \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \sum_d A_d' \Big(\eta_0 \exp\Big(- \frac{E_c}{N_0 \sin^2\theta} \Big) + \ldots + \eta_{J-1} \exp\Big(- \frac{E_c}{N_{J-1} \sin^2\theta} \Big) \Big)^d d\theta \\ &= \frac{1}{\pi} \int_0^{\frac{\pi}{2}} A(W) \Big|_{W = \eta_0 \exp\left(-\frac{E_c}{N_0 \sin^2\theta}\right) + \ldots + \eta_{J-1} \exp\left(-\frac{E_c}{N_{J-1} \sin^2\theta}\right)} d\theta \ . \end{aligned} \tag{6.2}$$

In practice, the bound is approximated by truncating the infinite series in the integrand.

6.1.4 Simulation Results

In Fig. 6.1, the simulated probability of code-word error is compared with approximations to each of the three weight-enumerator bounds for the memory-order-three encoder and the two-epoch channel with $\eta = 0.5$. The block length is 1000, and the noise power ratio γ is 3 dB. Each approximated bound is determined by truncating the bound to include only code words composed of error events of Hamming weight sixteen or less. The approximation accounts for all code words of total error-event span of ten or less (as well as some with a larger total error-event span), but it excludes all code words composed of three or more error events.

As seen in Fig. 6.1, the approximated Chernoff-weight-enumerator bound is quite loose. The approximated tighter Chernoff-weight-enumerator bound is approximately 1 dB better than the approximated Chernoff-weight-enumerator bound, and the approximated integral weight-enumerator bound is even better. The latter two approximated bounds are accurate for a probability of code-word error of 10^{-2} or less. No corresponding results are available for the NASA-standard code since calculation of the weight-span enumerating function of the code requires substantial effort [12].

6.2 Linear Bounds Using the First-Event Error Probability and Its Union Bound

The probability of code-word error can be expressed in terms of the first-event errors defined by equation (3.1). Two bounds are expressed in the following two theorems. An analogous result to Theorem 6.2 for hard-decision Viterbi decoding in i.i.d. Gaussian noise is given in [3].

Theorem 6.1: Suppose p denotes the first-event error probability at time zero for a transmission of block length L. Then

$$P_e \leq Lp.$$

Proof: See Appendix A.

The probability of code-word error is thus bounded by a union bound in terms of the first-event error probability at time zero. The latter probability can be determined only through simulation, however. An upper bound on the first-event error probability is given by the *first-event union bound*

$$P_u = \sum_{d_{free}}^{\infty} A_d P_d$$

where the coefficient A_d denotes the number of error events of Hamming weight d that begin at time zero. The coefficients $\{A_d\}$ together determine the *code-generating function* [6] of the convolutional code given by

$$T(W) = \sum_{d=d_{free}}^{\infty} A_d W^d.$$

A closed-form expression for the code-generating function can be determined by using Mason's Theorem [6]. The union bound on the probability of code-word error is given by the

following theorem.

Theorem 6.2:

$$P_e \leq LP_u. \tag{6.3}$$

Proof: See Appendix A.

The first-event union bound can be bounded or expressed exactly in closed-form using the respective bounds and exact expression for P_d presented in the previous section.

6.2.1 Union-Chernoff Bound

The Chernoff bound on P_d results in

$$P_e \leq LP_{ch} \tag{6.4}$$

where

$$P_{ch} = T(W)\big|_{W=\eta_0 \exp\left(-\frac{E_c}{N_0}\right)+\ldots+\eta_{J-1}\exp\left(-\frac{E_c}{N_{J-1}}\right)}. \tag{6.5}$$

We refer to equation (6.4) as the *union-Chernoff bound* on P_e.

6.2.2 Tighter Union-Chernoff Bound

The tighter Chernoff bound on P_d results in

$$P_e \leq LP_{t\text{-}ch} \tag{6.6}$$

where

$$P_{t\text{-}ch} = Q\left(\sqrt{\frac{2d_{free}E_c}{N_{max}}}\right)\exp\left(\frac{E_c d_{free}}{N_{max}}\right)T(W)\Big|_{W=\eta_0 \exp\left(-\frac{E_c}{N_0}\right)+\ldots+\eta_{J-1}\exp\left(-\frac{E_c}{N_{J-1}}\right)}. \tag{6.7}$$

We refer to equation (6.6) as the *tighter union-Chernoff bound* on P_e.

6.2.3 Integral Form of the Union Bound

The integral expression for P_d can be used to obtain an exact expression for the union bound on the first-event error probability, resulting in equation (6.3), where the first-event union-bound P_u is expressed as

$$P_u = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} T(W)\Big|_{W=\eta_0 \exp\left(-\frac{E_c}{N_0}\right)+\ldots+\eta_{J-1} \exp\left(-\frac{E_c}{N_{J-1}}\right)} d\theta \ . \tag{6.8}$$

We refer to equation (6.3) expressed in this form as the *union-integral bound* on P_e. Note that this is a union-bound that differs from the union bound in equation (6.2).

6.2.4 Simulation Results

The accuracy of each union bound is illustrated in Figs. 6.2 and 6.3 for the memory-order-three encoder and the NASA-standard encoder, respectively, and the two-epoch channel with $\eta = 0.5$. The block length is 1000 and the noise power ratio is 3 dB. Once again, the use of the integral bound leads to the tightest bound on the probability of code-word error. The union-integral bound is almost 0.4 dB better than the tighter union-Chernoff bound at $P_e = 10^{-3}$ for the memory-order-three encoder. The tighter union-Chernoff bound is almost as tight as the union-integral bound for the NASA-standard encoder. The union-integral bound provides approximately the same accuracy as the approximate integral weight-enumerator bound for the memory-order-three encoder, as seen by comparing Figs. 6.1 and 6.2.

6.3 New Concave Bounds Using the First-Event Union Bound

The best of the union bounds in the previous two sections are loose if the probability of code-word error is much above 10^{-1} for the memory-order-three encoder. The union-integral bound is loose if the probability of code-word error is much above 5×10^{-2} for the NASA-standard code, and the integral weight-enumerator bound is difficult to obtain for

convolutional codes of complexity equal to or greater than the NASA-standard code. In this section we consider new bounds which are uniformly tighter than the corresponding union bounds and can be determined in practice for codes of the complexity of the NASA-standard code. The bounds are analogous in form to one derived earlier for hard-decision Viterbi decoding [3], though the proof requires a different approach than the combinatorial argument used in [3].

The new bound is given by the following theorem.

Theorem 6.3:

$$P_e \leq 1 - (1 - P_u)^L \ . \tag{6.9}$$

Proof: See Appendix B.

The bound given by equation (6.9) is a concave function of the first-event union bound, P_u, in contrast with the union bound of equation (6.3), which is a linear function of P_u. It is easily shown that the concave bound is strictly tighter than the linear bound. Several variants of the bound can be obtained using bounds on P_d.

6.3.1 Concave-Chernoff Bound

The Chernoff bound on each pairwise error-event probability results in

$$P_e \leq 1 - (1 - P_{ch})^L \tag{6.10}$$

where P_{ch} is given by equation (6.5). We refer to the bound of equation (6.10) as the *concave-Chernoff bound*. It is uniformly tighter than the union-Chernoff bound.

6.3.2 Tighter Concave-Chernoff Bound

The tighter Chernoff bound on each pairwise error-event probability results in

$$P_e \leq 1 - (1 - P_{t\text{-}ch})^L \tag{6.11}$$

where $P_{t\text{-}ch}$ is given by equation (6.7). We refer to the bound of equation (6.11) as the *tighter concave-Chernoff bound*. It is uniformly tighter than the tighter union-Chernoff bound.

6.3.3 Integral Form of the Concave Bound

The integral expression for P_d can be used to represent the bound of Theorem 6.3 exactly as

$$P_e \leq 1 - (1 - P_u)^L \tag{6.12}$$

where P_u is expressed as in equation (6.8). We refer to the bound of equation (6.12) as the *concave-integral bound*. It is uniformly tighter than the union-integral bound.

6.3.4 Simulation Results

The probability of code-word error is shown in Figs. 6.4 and 6.5 for the memory-order-three encoder and the NASA-standard encoder, respectively, both with the two-epoch channel with $\eta = 0.5$. The block length is 1000 and the noise power ratio is 3 dB. For both the encoders, the concave-integral bound gives a tight bound on the probability of code-word error. The bound is in error by only 0.1 dB at a probability of code-word error of 0.1 for the memory-order-three encoder, and it is in error by only 0.2 dB if the probability of error is 0.5. The concave-integral bound is not as tight for the NASA-standard code. It is accurate to within 0.2 dB if the probability of code-word error is 0.1, but it is much less accurate if the probability of code-word error is 0.5. It is noticeably tighter than the union-integral bound in this range, however, as seen by comparing Figs. 6.3 and 6.5.

6.4 New Concave Bound Using the First-Event Error Probability

The use of the union bound on the first-event error probability contributes some looseness to the bounds of the previous section, since the error events making up the first event error are not disjoint. A result of Slepian [15] can be used to remove this factor in the

looseness of the concave-integral bound. The tighter bound is given in the following theorem.

Theorem 6.4:

$$P_e \leq 1 - (1-p)^L \ .$$

Proof: See Appendix C.

We refer to this bound as the *concave-first-event bound*. Theorem 6.3 and the concave-union bounds follow immediately from Theorem 6.4 since $1-(1-x)^L$ is an increasing function of x for $0 \leq x \leq 1$.

Unlike the concave-union bounds, the concave-first-event bound is not analytically tractable. Instead, the first-event error probability p must be obtained through simulation for each channel of interest. Once it is known for a given channel and code, however, it can be used to bound the probability of code-word error for any block length.

6.5 Comparison of the Derived Bounds

In this section, we present a comparison of the bounds developed in the previous sections. Four bounds are considered: the simulation-aided concave-first-event bound of Section 6.4, and the three integral form bounds or approximations from Sections 6.1-6.3. Each of the integral-form bounds is the best from its respective section, which is expected since it uses an exact expression for the pair-wise error event probability or pairwise code-word error probability as opposed to an upper bound on the same. Simulation results are shown in Figs. 6.6 - 6.9 for the memory-order-three encoder and Figs. 6.10 - 6.13 for the NASA-standard encoder. In each example, the block length is 1000 and the noise power ratio is 3 dB. Interference-free intervals of $0.25, 0.5, 0.75$, and 1.0 are included.

For the memory-order-three encoder, the concave-first-event bound is the tightest; it is accurate to within 0.05 dB for all channel conditions. The bound is particularly useful at

low signal-to-noise ratios for which the other bounds become less accurate and then diverge below a critical signal-to-noise ratio. This is a consequence of its use of the simulated (actual) first-event error probability, instead of the (infinite-series) first-event union bound used in the other bounds.

The concave-integral bound gives the best estimate of the code-word error probability among the (purely) analytical bounds, followed by the approximate integral weight-enumerator bound and the union-integral bound (both of which are equally good). Note that differences between the bounds is limited to low signal-to-noise ratios. For higher signal-to-noise ratios, the bounds match closely and all agree closely with the simulated system performance. These observations are common the all the channel conditions in Figs. 6.6 - 6.9.

For the NASA-standard encoder, the concave-first-event bound again provides the best estimate of the code-word error probability (within 0.1 dB). The effect of the divergence of other bounds at low signal-to-noise ratios is more pronounced than with the weaker code. The analytical bounds are not accurate if the channel is such that the probability of code-word error is more than 0.1.

The union-integral bound is comparable to the concave-integral bound if the signal-to-noise ratio is high, though it is slightly poorer at lower signal-to-noise ratios. Both bounds result in an error of at least 0.5 dB for a probability of code-word error of 0.5.

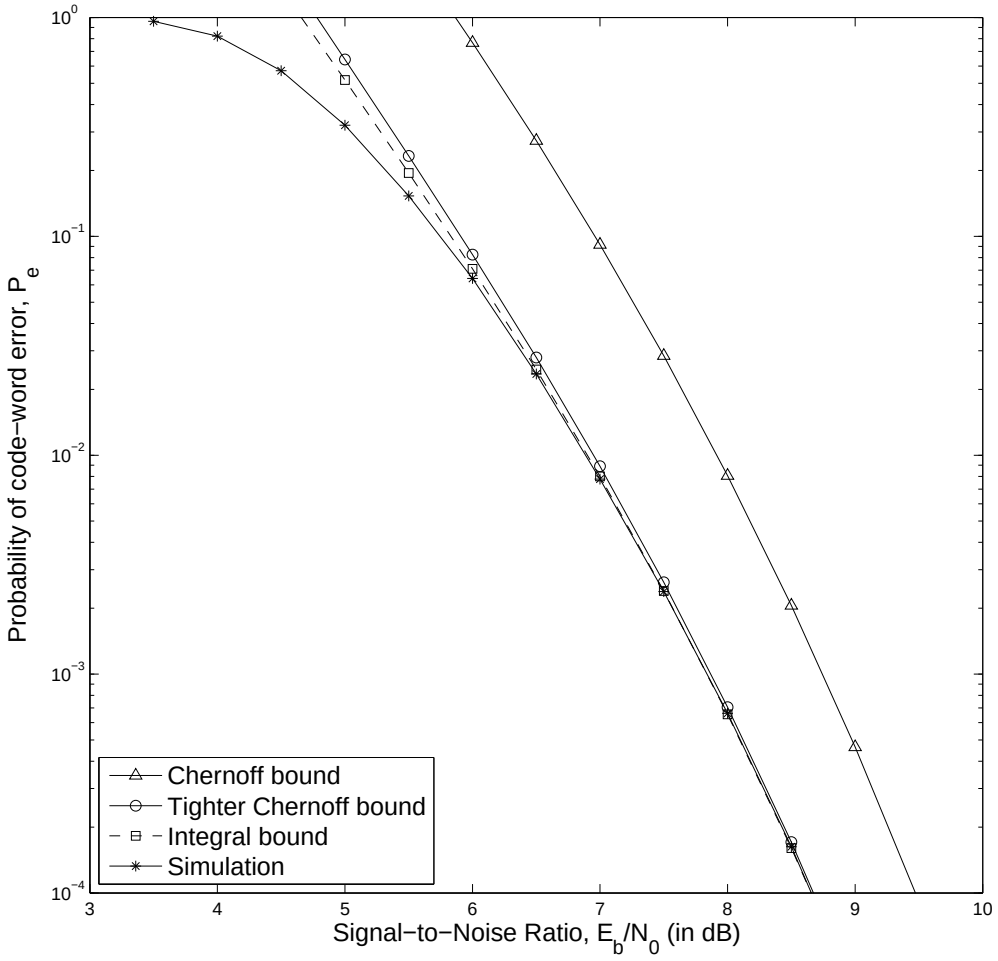

Figure 6.1: Weight-enumerator bounds for the memory-order-three encoder ($\eta = 0.5$).

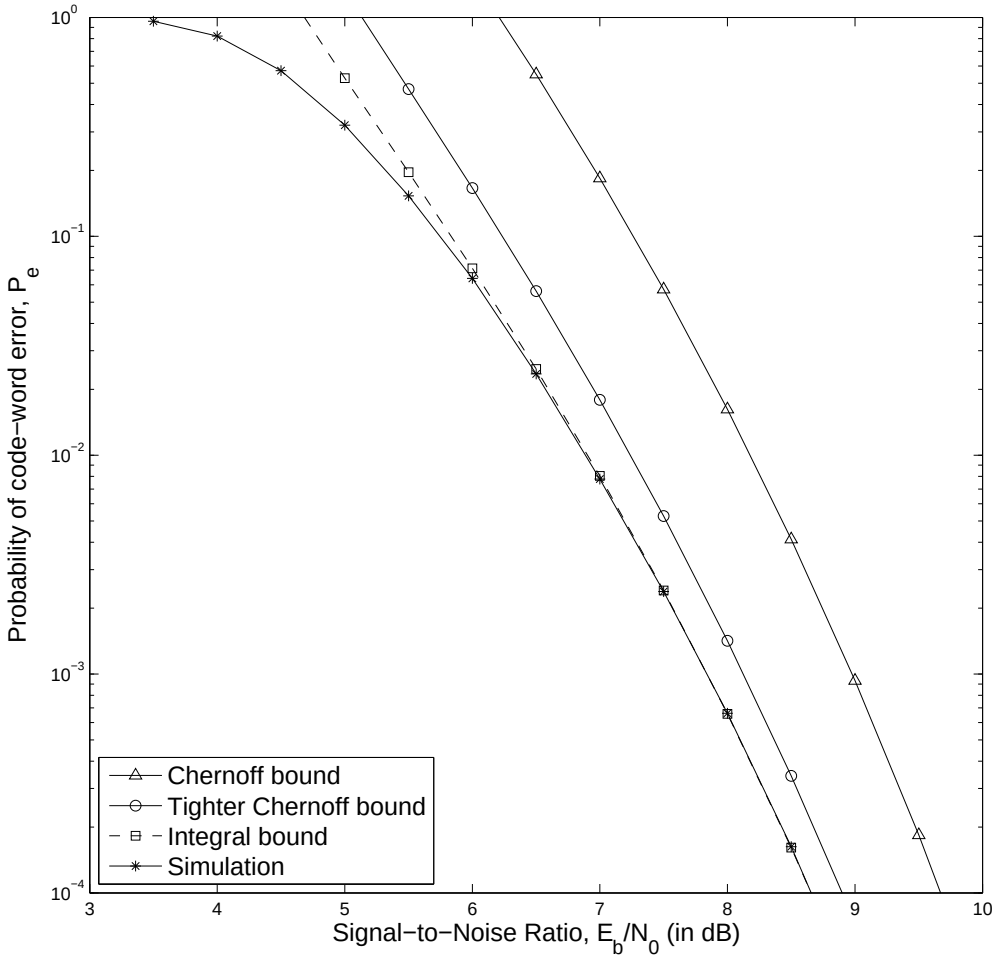

Figure 6.2: Linear bounds using the first-event union bound for the memory-order-three encoder ($\eta = 0.5$).

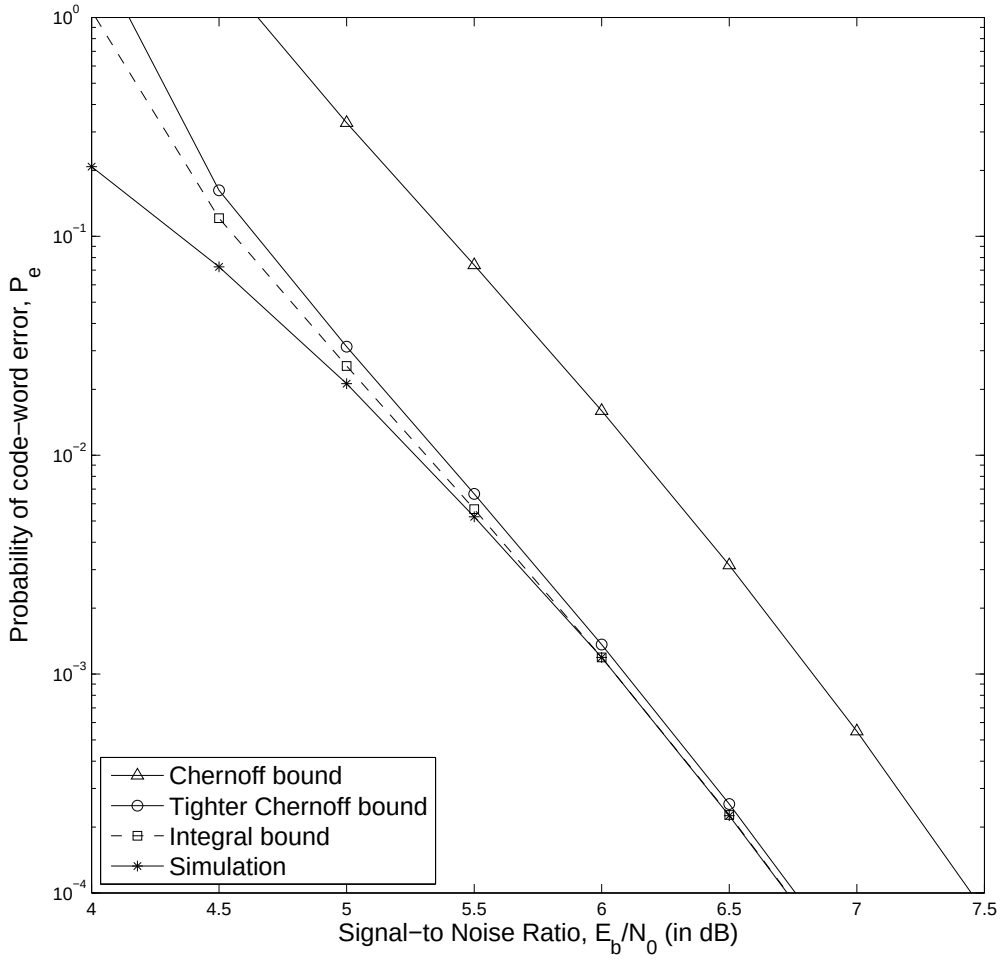

Figure 6.3: Linear bounds using the first-event union bound for the NASA-standard encoder ($\eta = 0.5$).

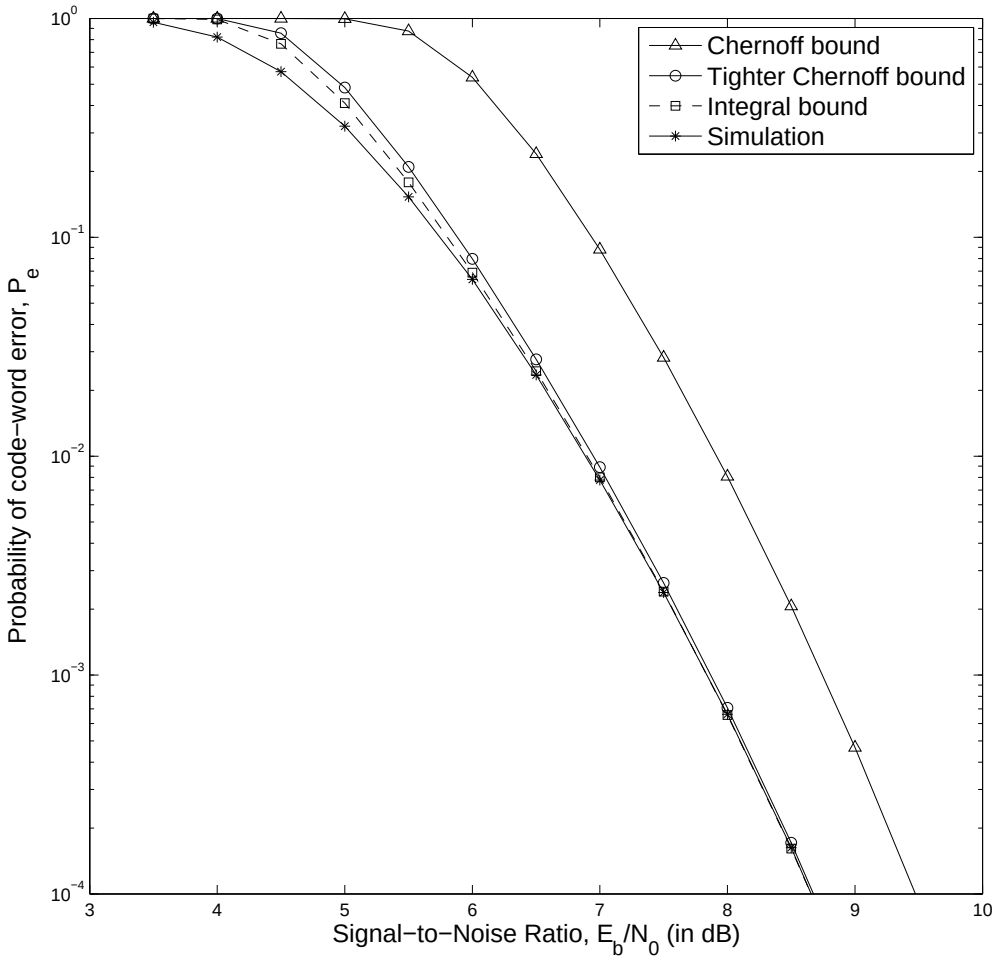

Figure 6.4: Concave bounds using the first-event union bound for the memory-order-three encoder ($\eta = 0.5$).

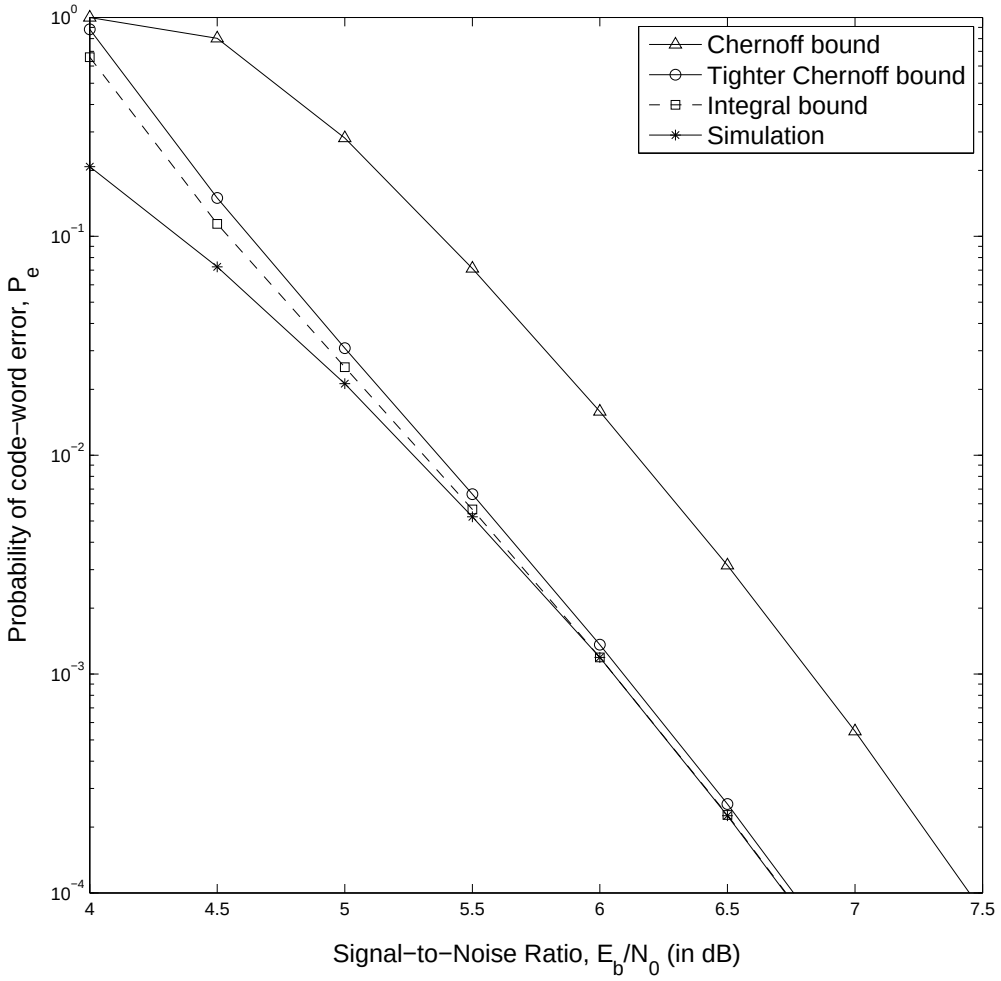

Figure 6.5: Concave bounds using the first-event union bound for the NASA-standard encoder ($\eta = 0.5$).

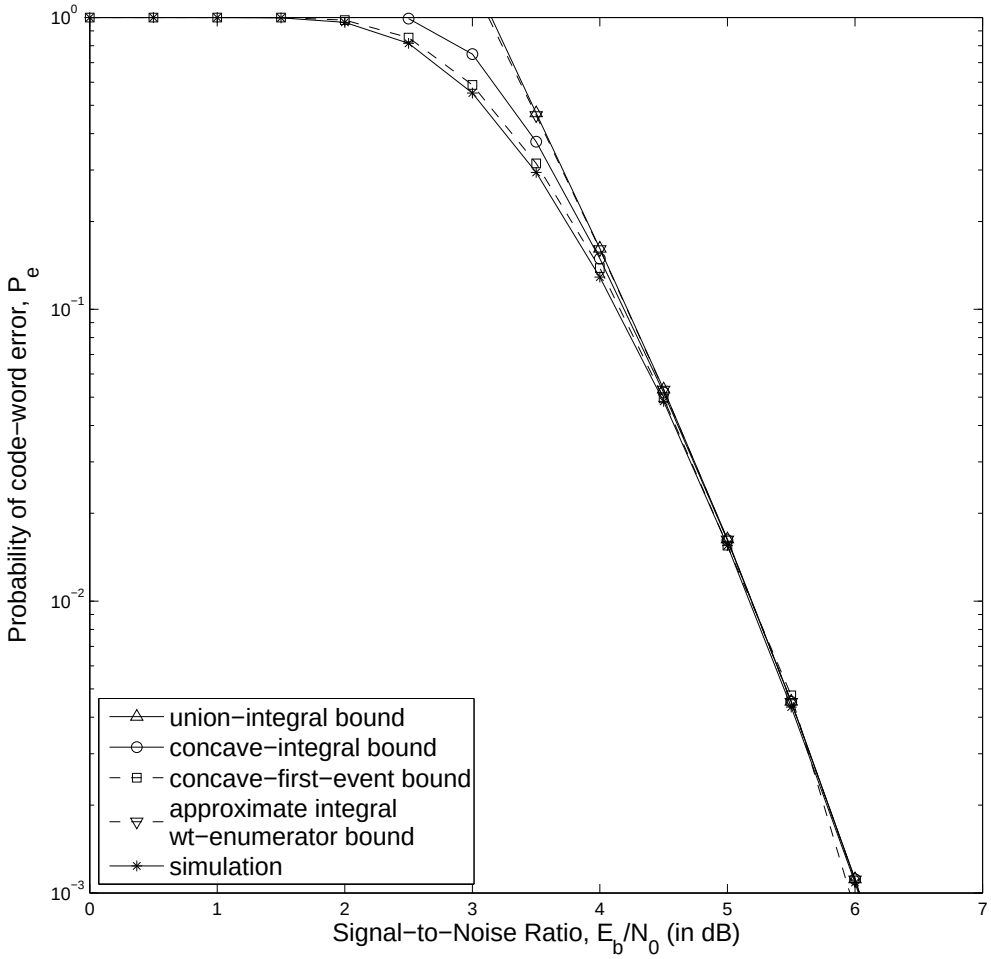

Figure 6.6: Comparison of bounds for memory-order-three encoder ($\eta = 1$).

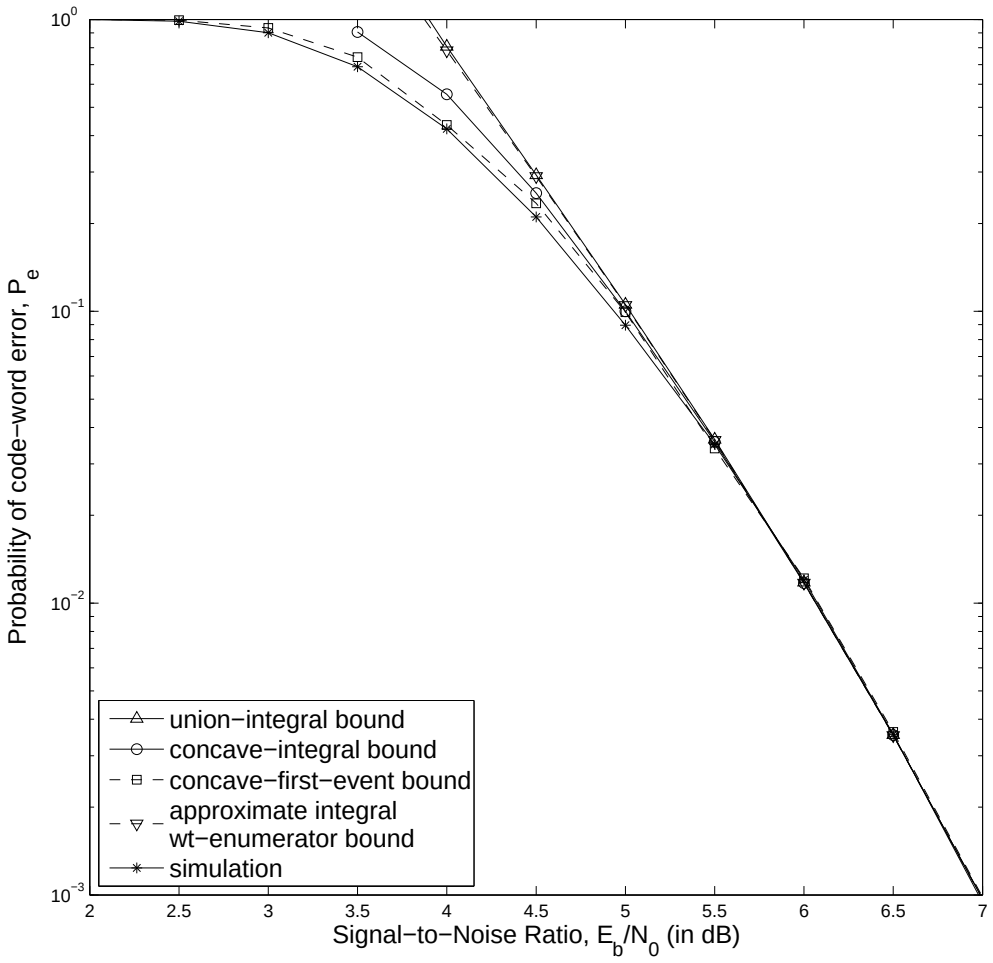

Figure 6.7: Comparison of bounds for memory-order-three encoder ($\eta = 0.75$).

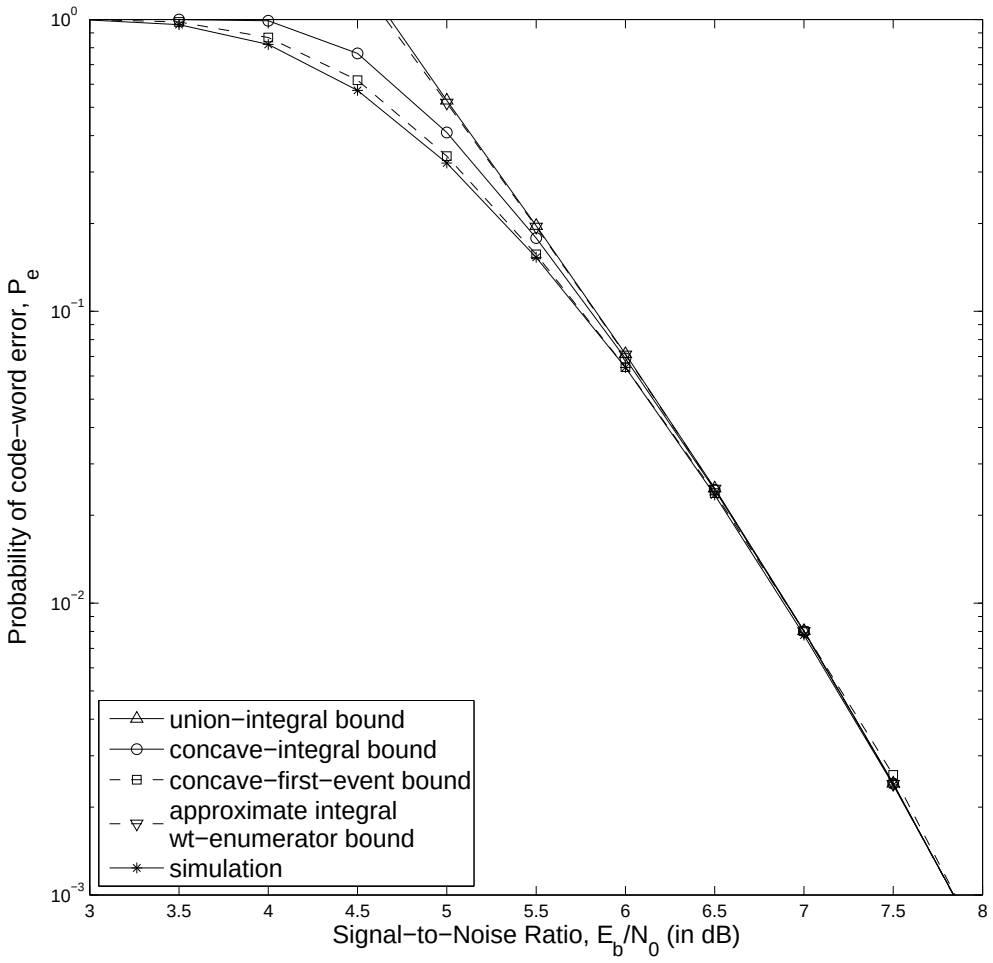

Figure 6.8: Comparison of bounds for memory-order-three encoder ($\eta = 0.5$).

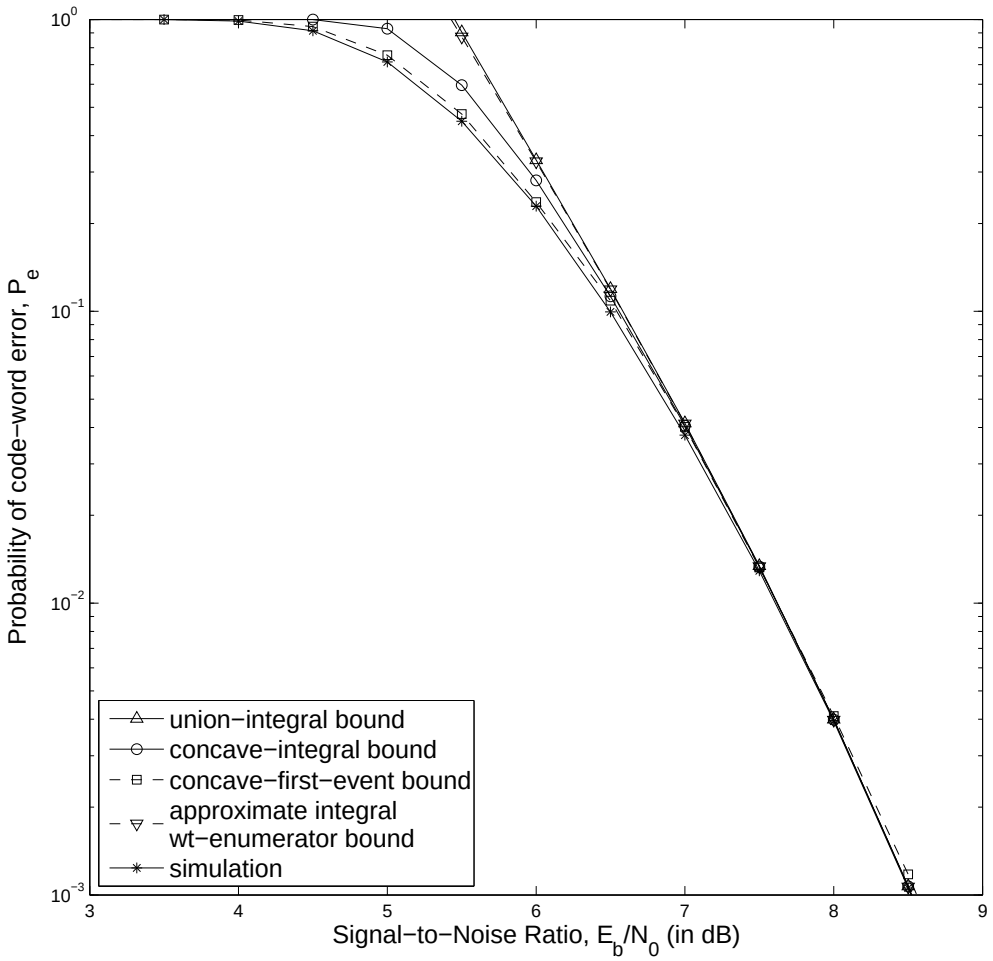

Figure 6.9: Comparison of bounds for memory-order-three encoder ($\eta = 0.25$).

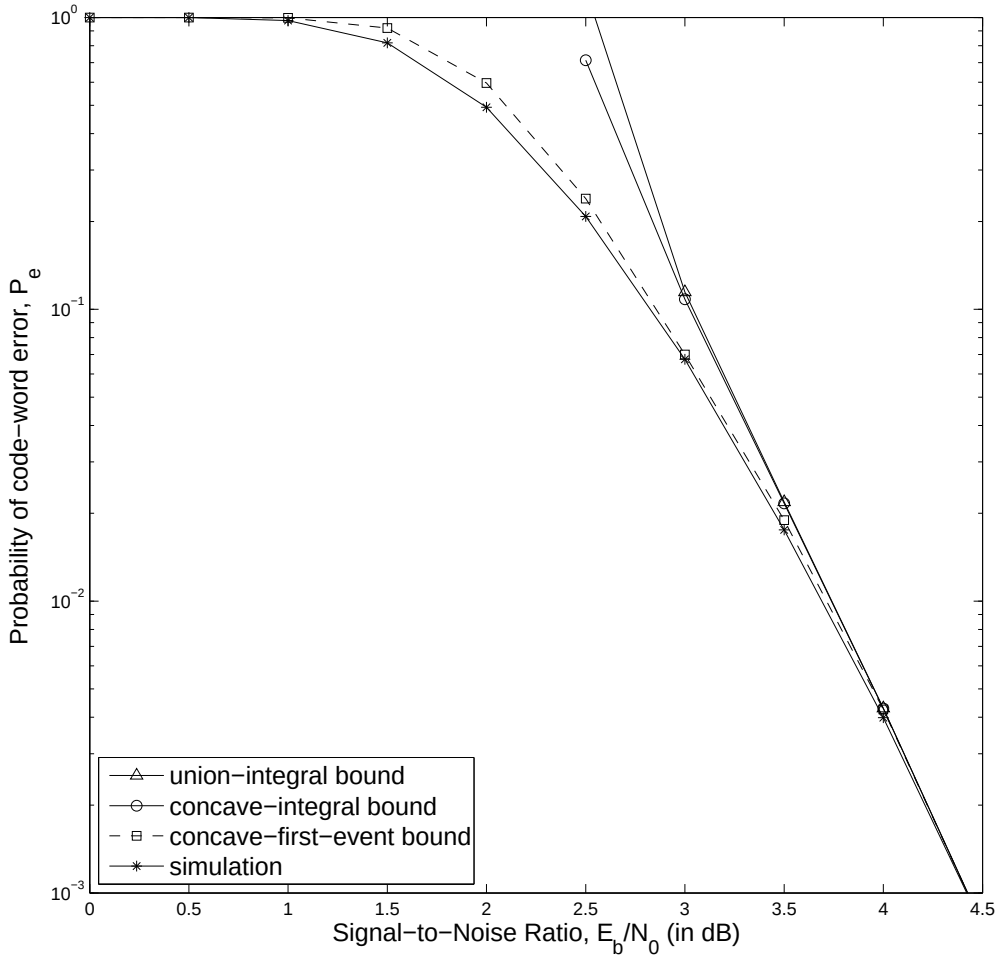

Figure 6.10: Comparison of bounds for NASA-standard encoder ($\eta = 1$).

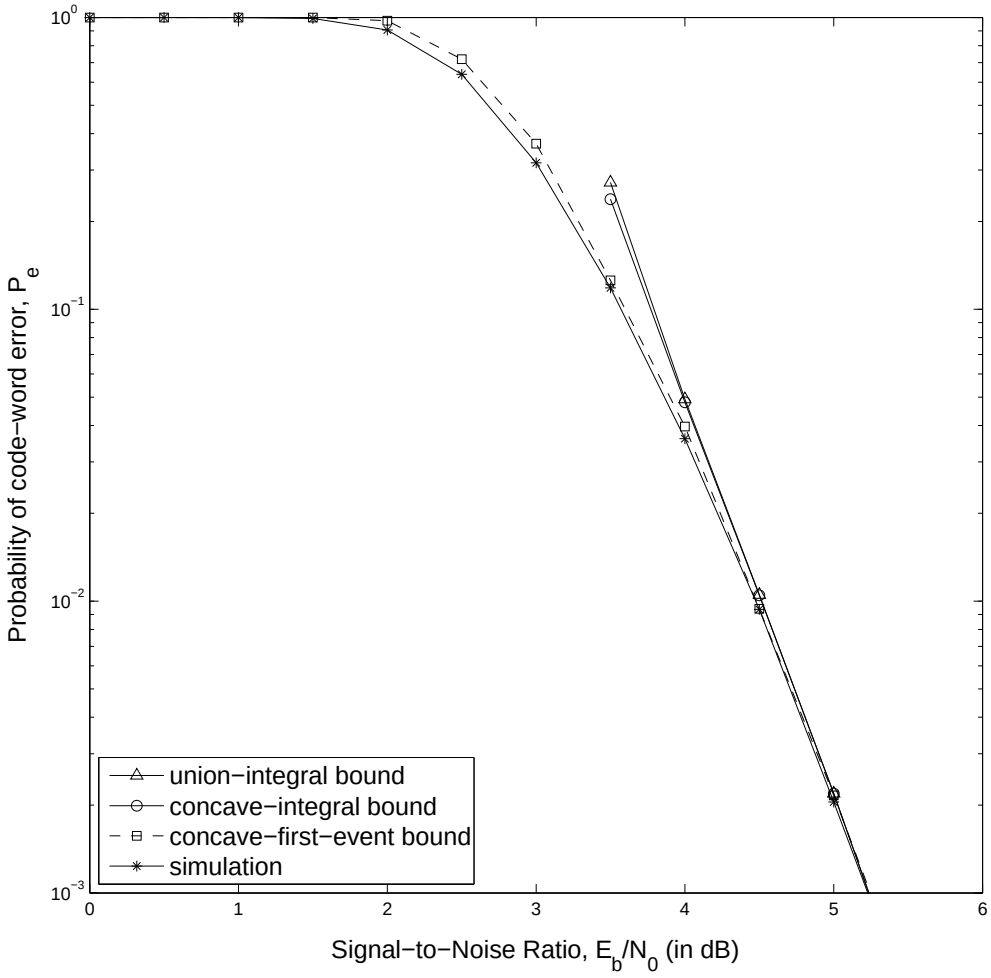

Figure 6.11: Comparison of bounds for NASA-standard encoder ($\eta = 0.75$).

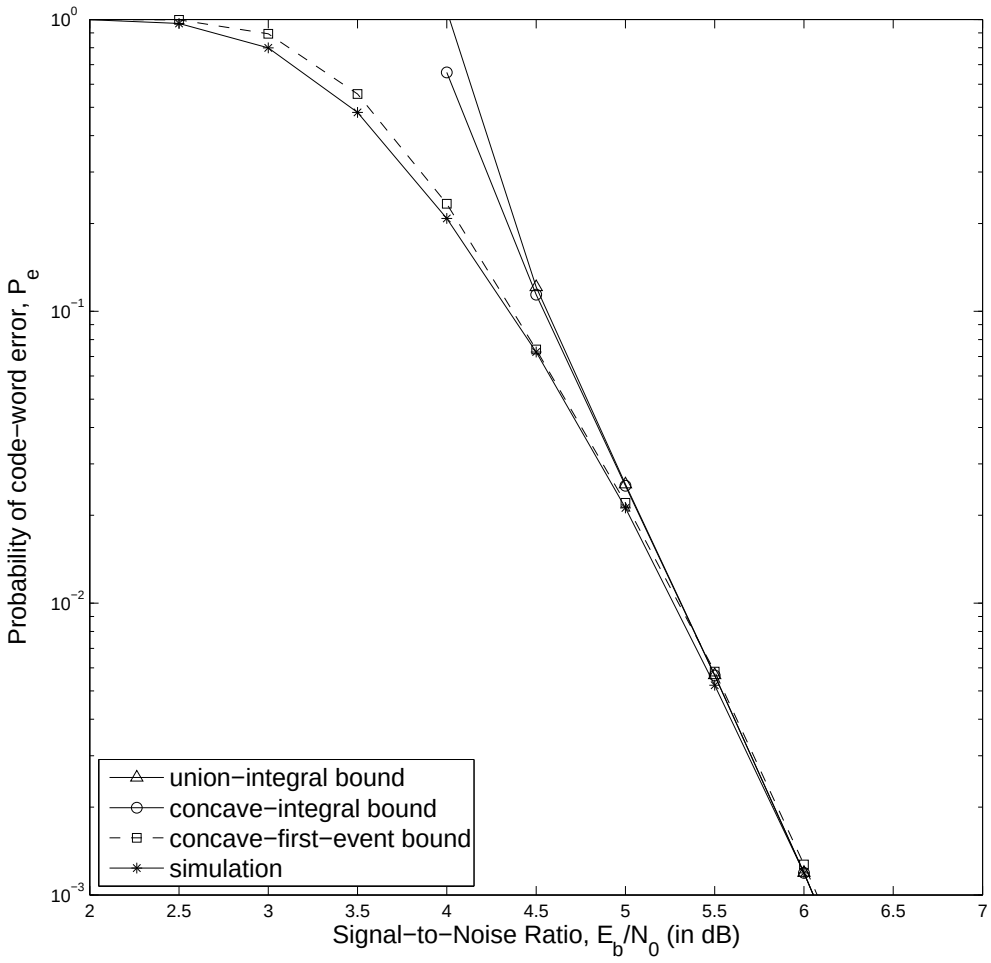

Figure 6.12: Comparison of bounds for NASA-standard encoder ($\eta = 0.5$).

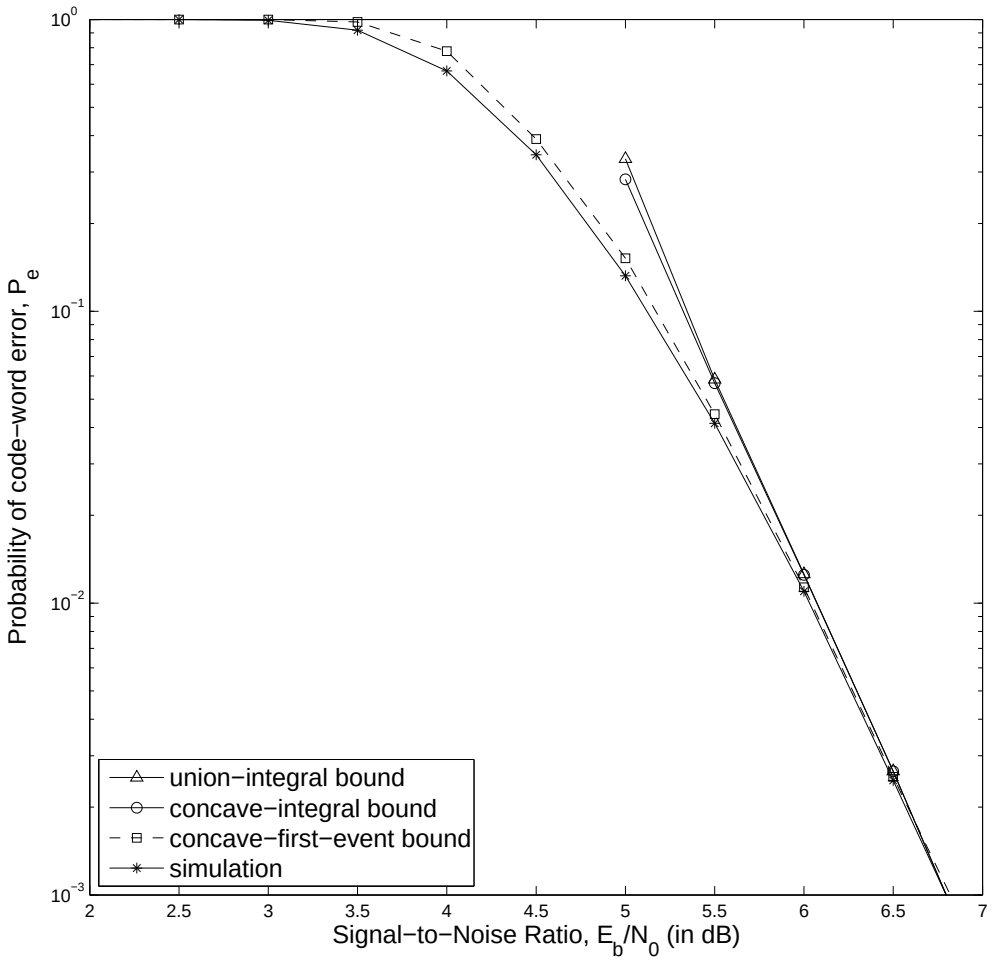

Figure 6.13: Comparison of bounds for NASA-standard encoder ($\eta = 0.25$).

Chapter 7

Stationary Gaussian Approximation to the Mixed-Distribution Channel

The concave-integral and concave-first-event bounds developed in Chapter 6 for the mixed-distribution channel provide very accurate approximations to the probability of code-word error which can be calculated very simply during a network simulation for any block length. Each requires numerous offline calculations for each code represented in the simulation and correspondingly large storage space for calculated values. High-fidelity use of the concave-integral bound requires computation of the first-event union bound for a sufficiently dense sampling of the following parameters: number of epochs, epoch durations, relative noise power levels, and signal-to-noise ratios. Greater offline computation is needed for the concave-first-event bound, which requires a simulation to obtain the first-event error probability for each combination of the channel parameters. (A strictly simulation-based performance evaluation for each combination of the channel parameters and each possible block length requires even greater offline computation and online storage.)

In this chapter, we consider an approximation to the mixed-distribution channel

which dramatically reduces the offline computation and the online storage required to model link performance in a network simulation. Specifically, the mixed-distribution channel is approximated by an i.i.d. (stationary) Gaussian noise channel in which the noise variances are given by

$$\mathrm{Var}(\tilde{n}_i) = E_{\Theta_i}[\mathrm{Var}(\tilde{n}_i \mid \Theta_i)].$$

Thus

$$\mathrm{Var}(\tilde{n}_i) = \frac{N}{2}$$

where

$$N = \sum_{i=0}^{J-1} \eta_i N_i.$$

7.1 Bounds on the Probability of Code-word Error

The channel based on the stationary Gaussian approximation is itself a mixed-distribution channel in which there is only one epoch. Thus each of the results developed in Chapter 6 is applicable. As in the earlier examples, the concave-integral bound and the concave-first-event bound are of greatest interest.

7.2 Accuracy of the Stationary Gaussian Approximation

The accuracy of the stationary Gaussian approximation is illustrated by considering an example of a two-epoch channel and its stationary approximation. The two-epoch channel has an interference-free fraction of 0.5 and a noise power ratio of 3 dB. For a given noise variance $\frac{N_0}{2}$ in the interference-free epoch, the noise variance throughout the transmission with the Gaussian approximation is thus

$$\frac{N}{2} = \frac{N_0 + 2N_0}{4} = \frac{3}{2}\frac{N_0}{2}.$$

The performance in the two channels is shown in Fig. 7.1 for the memory-order-three

encoder. The actual performance in the stationary Gaussian channel is slightly worse than the performance in the mixed-distribution channel if the probability of code-word error is high, whereas it is slightly better if the probability of code-word error is low. Across the range of error probabilities from 10^{-1} to 10^{-4}, the two differ by less than 0.2 dB. If the probability of error is above 10^{-1}, the two differ by less than 0.25 dB.

The concave-integral bound and the concave-first-event bound both result in tight bounds on the performance in the stationary Gaussian channel. The concave-first-event bound for the Gaussian channel differs from the performance of the mixed-distribution channel by no more than 0.3 dB if $P_e \geq 10^{-4}$. The concave-integral bound for the Gaussian channel differs from the performance of the mixed-distribution channel by no more than 0.4 dB if $P_e \geq 10^{-4}$.

Similar results are shown in Fig. 7.2 for the standard-NASA encoder. The actual performance in the stationary Gaussian channel differs by no more than 0.3 dB from the performance of the mixed-distribution channel if $P_e = 10^{-4}$. The concave-first-event bound for the stationary Gaussian channel is at most 0.1 dB poorer than this over the same range. The concave-integral bound for the Gaussian channel is almost 0.5 dB poorer than the actual performance in the mixed-distribution channel if $P_e = 0.1$, but the difference decreases to 0.1 dB if $P_e = 10^{-4}$.

Similar levels of accuracy using the stationary Gaussian channel approximation and its concave bounds are observed for the two encoders if $\eta = 0.25$ or $\eta = 0.75$.

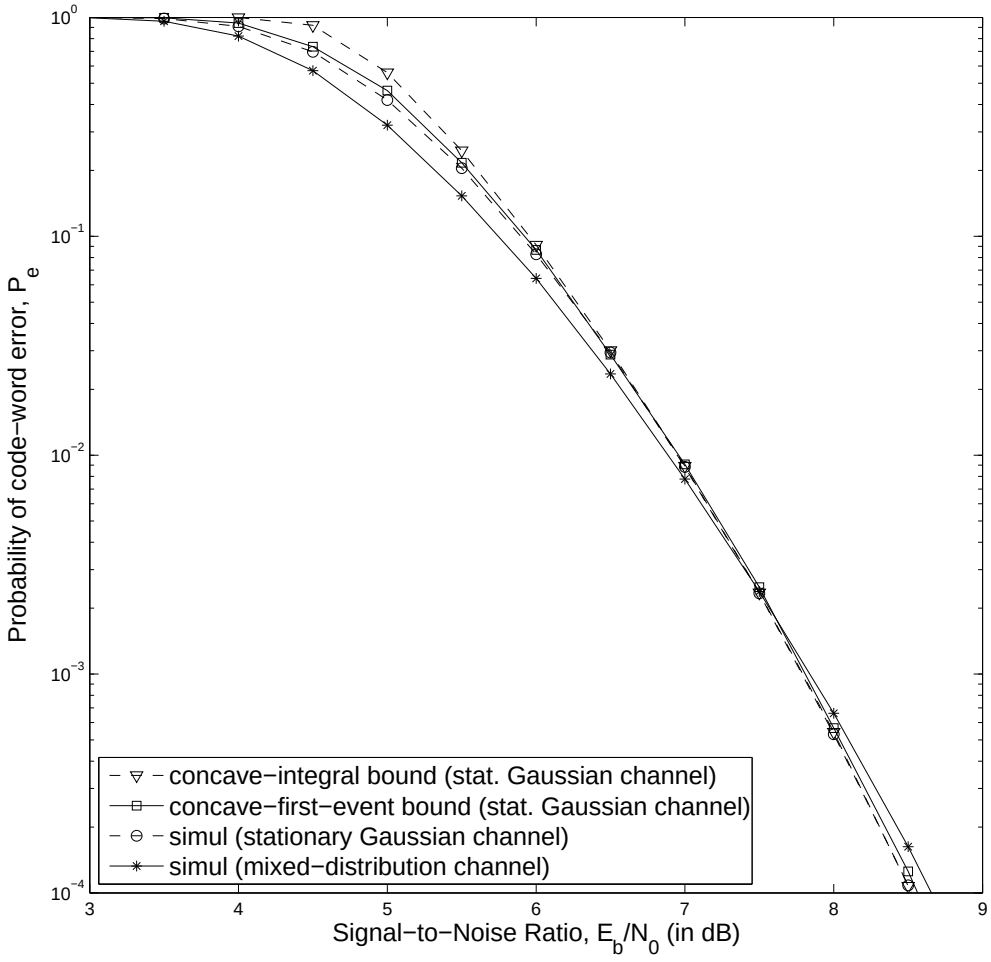

Figure 7.1: Accuracy of the stationary Gaussian approximation for the memory-order-three encoder ($\eta = 0.5$).

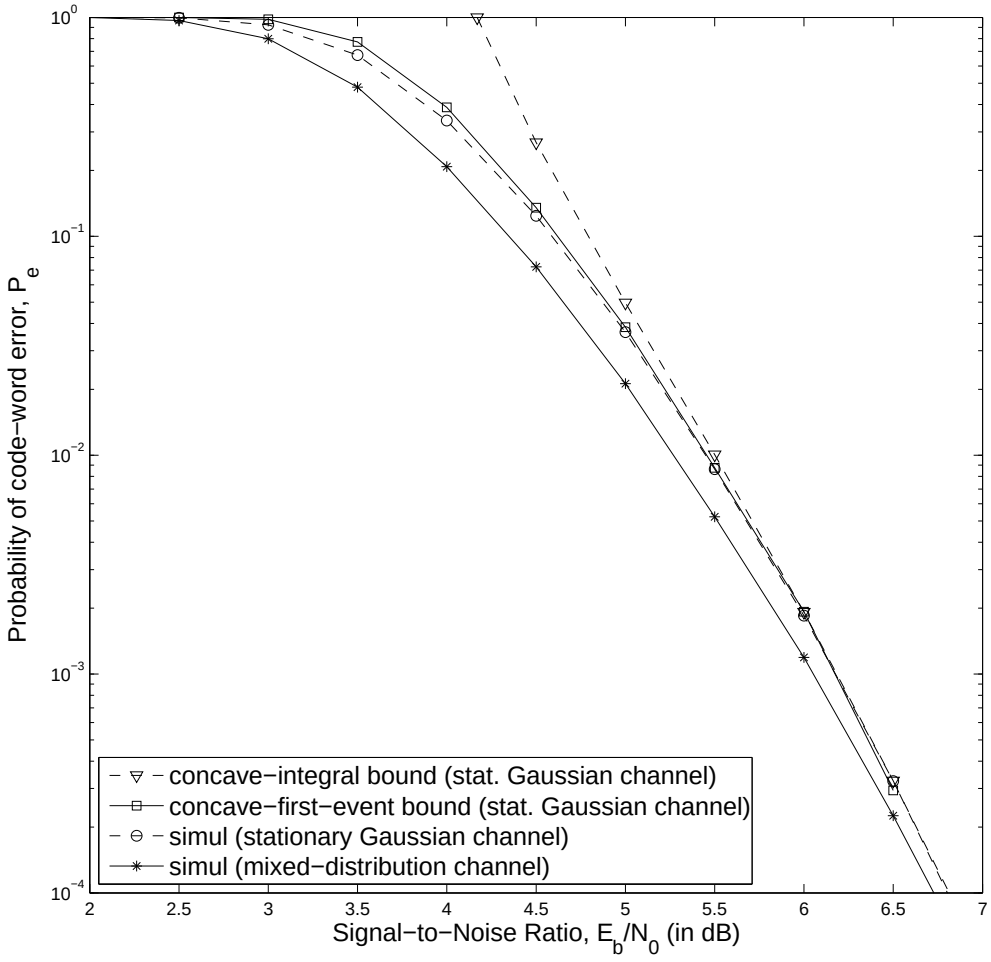

Figure 7.2: Accuracy of the stationary Gaussian approximation for the NASA-standard encoder ($\eta = 0.5$).

Chapter 8

Conclusion

The performance of convolutional codes and interleaving in the presence of block interference is investigated in this thesis. It is seen that the pseudo-random interleaver performs as well as optimal rectangular interleaving, and we develop an equivalent, analytically tractable statistical model for a system with pseudo-random interleaving and block interference. Using this model, we derive new bounds on the probability of code-word error for convolutional codes and soft-decision Viterbi decoding. The best of the bounds are highly accurate, as is illustrated for two rate-$\frac{1}{2}$ convolutional codes and various channel conditions. The new bounds permit accurate link modeling in simulations of ad hoc packet radio network that are subjected to partial-time Gaussian interference. The bounds require much less off-line computation and on-line storage than is required for off-line Monte Carlo simulation of each combination of channel conditions and packet format.

We also consider a simpler stationary Gaussian approximation for a system with pseudo-random interleaver and block interference. The approximation is useful for evaluating network performance when the memory available for storing the look-up tables of off-line simulation results is limited. The approximation leads to results which significantly reduce the memory requirements, though at the cost of a moderate penalty in the accuracy of the result.

Appendices

Appendix A Proofs of Theorems 6.1 and 6.2

The form of the bound in Theorem 6.2 has previously been shown to hold for hard-decision Viterbi decoding and an i.i.d. Gaussian noise channel [3]. In contrast, Theorems 6.1 and 6.2 are proven in this appendix to hold for soft-decision Viterbi decoding and any symmetric i.i.d. channel (including the mixed-distribution channel).

Theorem 6.1: Suppose p denotes the first-event error probability at time zero for a transmission of block length L. Then

$$P_e \leq Lp \ .$$

Proof. From equation (3.4),

$$\begin{aligned} P_e &= \Pr(\cup_{l=0}^{L-1} \cup_{i=1}^{\nu_l} \{X_{i,l} < 0\}) \\ &= \Pr(\cup_{l=0}^{L-1} F_l) \\ &= \sum_{l=0}^{L-1} \Pr(F_l) \end{aligned}$$

since the $\{F_l\}$ are disjoint.

The event error at time l, defined by equation (3.2), contains the first event error at time l. (That is, $F_l \subseteq G_l$.) Thus

$$P_e \leq \sum_{l=0}^{L-1} \Pr(G_l).$$

Moreover, $\Pr(G_0) \geq \Pr(G_l)$ for $l > 0$ since the encoder is shift invariant and the noise process is i.i.d. , and $G_0 = F_0$. Thus

$$P_e \leq \sum_{l=0}^{L-1} \Pr(F_0) = L \Pr(F_0) = Lp. \qquad \square$$

Theorem 6.2:

$$P_e \leq LP_u$$

Proof.

$$\begin{aligned} p &= \Pr(\cup_{i=1}^{\nu_0}\{X_{i,l} < 0\}) \\ &\leq \sum_{i=1}^{\nu_0} \Pr(X_{i,l} < 0) \\ &\leq \sum_{d=d_{free}}^{\infty} A_d P_d = P_u \end{aligned}$$

since $\Pr(X_{i,l} < 0)$ depends only on the Hamming weight of the corresponding code word. Thus from Theorem 6.1,

$$P_e \leq LP_u. \qquad \square$$

Appendix B Proof of Theorem 6.3

Consider n jointly Gaussian random variables, $\underline{\mathbf{X}}_n = (X_1, X_2, \ldots, X_n)$, with vector of mean values $\underline{\boldsymbol{\mu}}_n$ and covariance matrix $\boldsymbol{B}$ where

$$\underline{\boldsymbol{\mu}} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} \text{ and } \boldsymbol{B} = \begin{pmatrix} \Sigma_{n-1} & \Sigma_{n-1,1} \\ \Sigma_{1,n-1} & \Sigma_1 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_1 & \ldots & \ldots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \ldots & \rho_{2n}\sigma_2\sigma_n \\ \rho_{31}\sigma_3\sigma_1 & \ldots & \ldots & \ldots & \vdots \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \ldots & \ldots & \ldots & \sigma_n^2 \end{pmatrix},$$

Σ_{n-1} is an $(n-1) \times (n-1)$ matrix, $\Sigma_{n-1,1}$ is a column vector of length $n-1$, $\Sigma_{n-1,1} = \Sigma_{1,n-1}^T$, and $\Sigma_1 = \sigma_n^2$ is a scalar. The joint distribution of first $n-1$ Gaussian variables, $\underline{\mathbf{X}}_{n-1} = (X_1, X_2, \ldots, X_{n-1})$, is given as

$$\underline{\mathbf{X}}_{n-1} \sim \mathcal{N}\left(\underline{\boldsymbol{\mu}}_{n-1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{n-1} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_1 & \ldots & \rho_{1,n-1}\sigma_1\sigma_{n-1} \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \ldots & \rho_{2,n-1}\sigma_2\sigma_{n-1} \\ \vdots & & \ddots & \vdots \\ \rho_{n-1,1}\sigma_{n-1}\sigma_1 & \ldots & \ldots & \sigma_{n-1}^2 \end{pmatrix}\right)$$

where $\Sigma = \Sigma_{n-1}$. The joint distribution of same $n-1$ Gaussian variables conditioned on $X_n = \alpha$ is given as

$$\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) = \underline{\mathbf{X}}_{n-1} \mid (X_n = \alpha) \sim$$
$$\mathcal{N}\left(\hat{\underline{\boldsymbol{\mu}}}(\alpha) = \underline{\boldsymbol{\mu}}_{n-1} + \Sigma_{n-1,1}\Sigma_1^{-1}(\alpha - \mu_n), \hat{\Sigma} = \Sigma - \Sigma_{n-1,1}\Sigma_1^{-1}\Sigma_{1,n-1}\right).$$

First consider the case in which all of the $(n-1)$ entries in $\Sigma_{n-1,1}$ are strictly positive. (The case in which any of the entries is zero is dealt with separately.)

Lemma B.1: Suppose $\Sigma_{n-1,1}$ is strictly positive. If

$$g(\alpha) = \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) = \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid (X_n = \alpha)),$$

then $g(\alpha)$ is a strictly increasing function of α.

Proof. Let $\alpha_1 < \alpha_2$ and $\epsilon_i = \hat{\mu}_i(\alpha_2) - \hat{\mu}_i(\alpha_1)$ for $1 \leq i \leq n-1$. Note that for all i, $\epsilon_i > 0$.

$$g(\alpha_1) = \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha_1)$$

$$= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\hat{\underline{\mathbf{X}}}_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}$$

$$= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} \frac{1}{\sqrt{(2\pi)^{n-1}|\hat{\Sigma}|}} \exp\left(-\frac{1}{2}\Big(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha_1)\Big)^T \hat{\Sigma}^{-1}\Big(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha_1)\Big)\right) \mathbf{d}\underline{\mathbf{x}}$$

$$= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} \frac{1}{\sqrt{(2\pi)^{n-1}|\hat{\Sigma}|}} \exp\left(-\frac{1}{2}\Big(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha_2) + \underline{\epsilon}\Big)^T \hat{\Sigma}^{-1}\Big(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha_2) + \underline{\epsilon}\Big)\right) \mathbf{d}\underline{\mathbf{x}}.$$

Making the substitution $\underline{\mathbf{x}} + \underline{\epsilon} = \underline{\mathbf{x}}'$,

$$g(\alpha_1) = \int_{\underline{\mathbf{x}}' \geq \underline{\epsilon}} \frac{1}{\sqrt{(2\pi)^{n-1}|\hat{\Sigma}|}} \exp\left(-\frac{1}{2}\Big(\underline{\mathbf{x}}' - \underline{\hat{\mu}}(\alpha_2)\Big)^T \hat{\Sigma}^{-1}\Big(\underline{\mathbf{x}}' - \underline{\hat{\mu}}(\alpha_2)\Big)\right) \mathbf{d}\underline{\mathbf{x}}'$$

$$< \int_{\underline{\mathbf{x}}' \geq \underline{\mathbf{0}}} \frac{1}{\sqrt{(2\pi)^{n-1}|\hat{\Sigma}|}} \exp\left(-\frac{1}{2}\Big(\underline{\mathbf{x}}' - \underline{\hat{\mu}}(\alpha_2)\Big)^T \hat{\Sigma}^{-1}\Big(\underline{\mathbf{x}}' - \underline{\hat{\mu}}(\alpha_2)\Big)\right) \mathbf{d}\underline{\mathbf{x}}'$$

$$= g(\alpha_2).$$

Hence $g(\alpha)$ is a strictly increasing function of α. □

Lemma B.2: Suppose $\Sigma_{n-1,1}$ is strictly positive. Then,

$$\lim_{\alpha\to-\infty} g(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha\to\infty} g(\alpha) = 1.$$

Proof. For the joint distribution of $\underline{\hat{\mathbf{X}}}_{n-1}(\alpha)$, consider the $(n-1)$-dimensional ellipsoid of a given fixed probability density centered at $\underline{\hat{\boldsymbol{\mu}}}(\alpha)$ and let the interior of the ellipsoid be denoted by $E_c(\alpha)$ (i.e., $E_c(\alpha) = \{\underline{\mathbf{x}} : (\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha))\hat{\Sigma}^{-1}(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha)) < c\}$). Also, since the covariance matrix $\hat{\Sigma}$ is positive semi-definite, we can write $\hat{\Sigma} = AA^T$ where A is an $(n-1)\times(n-1)$ matrix. Suppose $\underline{\mathbf{n}} \sim \mathcal{N}(\underline{\mathbf{0}}, \mathbb{I}_{n-1})$ and $\underline{\hat{\mathbf{Y}}}_{n-1}(\alpha) = A\underline{\mathbf{n}} + \underline{\hat{\mu}}(\alpha)$. Then $\underline{\hat{\mathbf{Y}}}_{n-1}(\alpha) \sim \underline{\hat{\mathbf{X}}}_{n-1}(\alpha)$ (i.e. they have the same joint distribution).

Consider $\Pr(\underline{\hat{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha))$ for some c.

$$\Pr(\underline{\hat{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha)) = \int_{E_c(\alpha)} f_{\underline{\hat{\mathbf{X}}}_{n-1}(\alpha)}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}$$

$$= \int_{E_c(\alpha)} \frac{1}{(2\pi)^{\frac{n-1}{2}}|\hat{\Sigma}|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha))^T\hat{\Sigma}^{-1}(\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha))\Big)\ \mathbf{d}\underline{\mathbf{x}}\ .$$

Making the substitution $\underline{\mathbf{x}} - \underline{\hat{\mu}}(\alpha) = A\underline{\mathbf{n}}$ and letting $\hat{E}_c = \{\underline{\mathbf{n}} | \underline{\mathbf{n}}^T \underline{\mathbf{n}} < c\}$,

$$\Pr(\underline{\hat{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha)) = \int_{\hat{E}_c} \frac{|A|}{(2\pi)^{\frac{n-1}{2}} |\hat{\Sigma}|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2} (A\underline{\mathbf{n}})^T \hat{\Sigma}^{-1} (A\underline{\mathbf{n}}) \Big) \, \mathbf{d}\underline{\mathbf{n}}$$

$$= \int_{\hat{E}_c} \frac{|A|}{(2\pi)^{\frac{n-1}{2}} |AA^T|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2} (A\underline{\mathbf{n}})^T (AA^T)^{-1} (A\underline{\mathbf{n}}) \Big) \, \mathbf{d}\underline{\mathbf{n}}$$

$$= \int_{\hat{E}_c} \frac{1}{(2\pi)^{\frac{n-1}{2}}} \exp\Big(-\frac{1}{2} \underline{\mathbf{n}}^T \underline{\mathbf{n}} \Big) \, \mathbf{d}\underline{\mathbf{n}}$$

$$= \Pr(\underline{\mathbf{n}} \in \hat{E}_c)$$

$$= \Pr(\sum_{i=1}^{n-1} n_i^2 < c) \ .$$

Let $Z = \sum_{i=1}^{n-1} n_i^2$. Then Z has a central chi-square distribution with $(n-1)$ degrees of freedom [11]. Hence

$$\Pr(\underline{\hat{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha)) = \Pr(Z < c)$$
$$= \frac{1}{\Gamma(\frac{n-1}{2})} \, \gamma\left(\frac{n-1}{2}, \frac{c}{2} \right)$$

where $\Gamma(k)$ is the Gamma function and $\gamma(k, z)$ is the lower incomplete Gamma function [11].

For any $\epsilon > 0$, c can be chosen such that

$$1 - \epsilon < \frac{1}{\Gamma(\frac{n-1}{2})} \, \gamma\left(\frac{n-1}{2}, \frac{c}{2} \right) \ .$$

Choose some finite value of c which satisfies the above equation.

Let λ denote the largest eigenvalue of $\hat{\Sigma}$. Then the largest (semi-principal) axis of

the ellipsoid is $a = \sqrt{c\lambda}$. For $1 \leq i \leq n-1$, we know that

$$\hat{\mu}_i(\alpha) = \mu_i + \rho_{i,n}\frac{\sigma_i}{\sigma_n}(\alpha - \mu_n).$$

Let $\alpha_i = (\sqrt{c\lambda} - \mu_i)\frac{\sigma_n}{\sigma_i\rho_{i,n}} + \mu_n$ for all $1 \leq i \leq n-1$. Choose $\alpha^* = \max\{\alpha_1, \ldots, \alpha_{n-1}\}$. Then $\hat{\mu}_i(\alpha^*) \geq \sqrt{c\lambda}$ for all i and $E_c(\alpha^*) \subset \{\hat{\underline{\mathbf{X}}}_{n-1}(\alpha^*) \geq \underline{\mathbf{0}}\}$. Therefore, for all $\alpha > \alpha^*$,

$$\begin{aligned}
&\Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) \geq \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha)) > 1 - \epsilon \\
\Rightarrow &\lim_{\alpha\to\infty} \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) = 1 \\
\Rightarrow &\lim_{\alpha\to\infty} g(\alpha) = 1.
\end{aligned}$$

Similarly, let $\alpha_i = (-\sqrt{c\lambda} - \mu_i)\frac{\sigma_n}{\sigma_i\rho_{i,n}} + \mu_n$ for all $1 \leq i \leq n-1$ and choose $\alpha^{**} = \min\{\alpha_1, \ldots, \alpha_{n-1}\}$. Then $\hat{\mu}_i(\alpha^{**}) \leq -\sqrt{c\lambda}$ for all i and $E_c(\alpha^{**}) \subset \{\hat{\underline{\mathbf{X}}}_{n-1}(\alpha^{**}) \leq \underline{\mathbf{0}}\}$. Therefore, for all $\alpha < \alpha^{**}$,

$$\begin{aligned}
&\Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \leq \underline{\mathbf{0}}) \geq \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \in E_c(\alpha)) > 1 - \epsilon \\
\Rightarrow &1 - \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) \geq 1 - \epsilon \\
\Rightarrow &\Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) \leq \epsilon \\
\Rightarrow &\lim_{\alpha\to-\infty} \Pr(\hat{\underline{\mathbf{X}}}_{n-1}(\alpha) \geq \underline{\mathbf{0}}) = 0 \\
\Rightarrow &\lim_{\alpha\to-\infty} g(\alpha) = 0.
\end{aligned}$$

□

Now consider the case where at least one of the entries in Σ_{21} is zero. Let $\underline{\mathbf{X}}_n = (\underline{\mathbf{X}}_p, X_n, \underline{\mathbf{X}}_m)^T$ such that $p + m = n - 1$. The vector of mean values of $\underline{\mathbf{X}}_n$ is

$$(\underline{\boldsymbol{\mu}}_p, \mu_n, \underline{\boldsymbol{\mu}}_m)^T$$

and the covariance matrix is

$$\begin{pmatrix} \Sigma_p & \Sigma_{p,n} & \Sigma_{p,m} \\ \Sigma_{n,p} & \sigma_n^2 & \Sigma_{n,m} \\ \Sigma_{m,p} & \Sigma_{m,n} & \Sigma_m \end{pmatrix},$$

where Σ_k is a $k \times k$ matrix, $\Sigma_{i,j}$ is a $i \times j$ matrix, and $\Sigma_{i,j} = \Sigma_{j,i}^T$. Without loss of generality, each of the m random variables in $\underline{\mathbf{X}}_m$ is uncorrelated with X_n and each of the p random variables in $\underline{\mathbf{X}}_p$ is positively correlated with X_n. (I.e., all entries in $\Sigma_{p,n}$ are positive and all entries in $\Sigma_{n,m}$ are zero.) Then,

$$\begin{aligned} \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) &= \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}, \underline{\mathbf{X}}_p \geq \underline{\mathbf{0}} \mid X_n = \alpha) \\ &= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\underline{\mathbf{X}}_p \geq \underline{\mathbf{0}} \mid \underline{\mathbf{X}}_m = \underline{\mathbf{x}}, X_n = \alpha) \mathbf{d}\underline{\mathbf{x}} \\ &= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) h(\underline{\mathbf{x}}, \alpha) \mathbf{d}\underline{\mathbf{x}} \ . \end{aligned}$$

Let

$$\tilde{\underline{\mathbf{X}}}_{p+1} = (\underline{\mathbf{X}}_p, X_n) \mid \underline{\mathbf{X}}_m = \underline{\mathbf{x}} \ .$$

The mean of $\tilde{\underline{\mathbf{X}}}_{p+1}$ is

$$\begin{pmatrix} \tilde{\underline{\boldsymbol{\mu}}}(\underline{\mathbf{x}}) \\ \mu_n \end{pmatrix} = \begin{pmatrix} \underline{\boldsymbol{\mu}}_p + \Sigma_{pm} \Sigma_m^{-1} (\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_m) \\ \mu_n \end{pmatrix}$$

and the covariance matrix can be written as

$$\boldsymbol{B} = \begin{pmatrix} \tilde{\Sigma}_p & \tilde{\Sigma}_{p,1} \\ \tilde{\Sigma}_{1,p} & \tilde{\Sigma}_1 \end{pmatrix},$$

where $\tilde{\Sigma} = \tilde{\Sigma}_p = \mathrm{Cov}(\underline{\mathbf{X}}_\mathrm{p}, \underline{\mathbf{X}}_\mathrm{p} \mid \underline{\mathbf{X}}_\mathrm{m} = \underline{\mathbf{x}}), \tilde{\Sigma}_1 = \sigma_\mathrm{n}^2$ and $\tilde{\Sigma}_{1,p} = \tilde{\Sigma}_{p,1}^T = (\rho_{1,n}\sigma_1\sigma_n \ldots \rho_{p,n}\sigma_p\sigma_n)$, where $\rho_{i,n}$ is the correlation coefficient between X_i and X_n for $1 \leq i \leq p$.

Lemma B.3: $\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha)$ is a strictly increasing function of α if $p > 0$.

Proof. $\underline{\mathbf{X}}_p$ is positively correlated with X_n (i.e., $\rho_{i,n} > 0$ for all $1 \leq i \leq p$). Hence, all entries of $\tilde{\Sigma}_{1,p}$ are positive. This is the same problem considered in the previous section with $n = p + 1$ and $h(\underline{\mathbf{x}}, \alpha) = \Pr(\underline{\mathbf{X}}_p \geq \underline{\mathbf{0}} \mid \underline{\mathbf{X}}_m = \underline{\mathbf{x}}, X_n = \alpha)$. Thus, using Lemma $B.1$, we can conclude that $h(\underline{\mathbf{x}}, \alpha)$ is strictly increasing in α for each $\underline{\mathbf{x}}$.

Let $\alpha_1 < \alpha_2$. Then $h(\underline{\mathbf{x}}, \alpha_1) < h(\underline{\mathbf{x}}, \alpha_2)$ for each $\underline{\mathbf{x}}$. Also, $f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) > 0$ for all $\underline{\mathbf{x}}$, hence,

$$
\begin{aligned}
& f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) h(\underline{\mathbf{x}}, \alpha_1) < f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) h(\underline{\mathbf{x}}, \alpha_2) \\
& \quad \Rightarrow \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) h(\underline{\mathbf{x}}, \alpha_1) \mathbf{d}\underline{\mathbf{x}} < \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) h(\underline{\mathbf{x}}, \alpha_2) \mathbf{d}\underline{\mathbf{x}} \\
& \quad \Rightarrow \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha_1) < \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha_2) \\
& \quad \Rightarrow \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) \text{ is strictly increasing in } \alpha.
\end{aligned}
$$

□

Note that if $p = 0$, $\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) = \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}})$ for all α.

Lemma B.4:

$$\lim_{\alpha \to -\infty} \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) = 0$$

and

$$\lim_{\alpha \to \infty} \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) = \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}) \ .$$

Proof. Let $\underline{\tilde{\mathbf{X}}}_p(\underline{\mathbf{x}}) = \underline{\mathbf{X}}_p \mid (\underline{\mathbf{X}}_m = \underline{\mathbf{x}}) \sim \mathcal{N}(\underline{\tilde{\mu}}(\underline{\mathbf{x}}), \tilde{\Sigma})$ and $\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) = \underline{\tilde{\mathbf{X}}}_p(\underline{\mathbf{x}}) \mid (X_n = \alpha) \sim \mathcal{N}(\underline{\hat{\tilde{\mu}}}(\underline{\mathbf{x}}, \alpha), \hat{\tilde{\Sigma}})$.

$$\begin{aligned}
\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) &= \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}, \underline{\mathbf{X}}_p \geq \underline{\mathbf{0}} \mid X_n = \alpha) \\
&= \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} \\
&= \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} \\
&\quad + \int_{\{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}\} / \{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M\}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} \ .
\end{aligned}$$

For any $\epsilon > 0$, we can choose $\underline{\mathbf{x}}_M$ such that

$$\int_{\{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}\} / \{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M\}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} < \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{x}}_M) < \frac{\epsilon}{2}$$

and let $\Pr(\underline{\mathbf{0}} \leq \underline{\mathbf{X}}_m \leq \underline{\mathbf{x}}_M) = \delta$.

For the joint distribution of $\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha)$, consider the p-dimensional ellipsoid of a given fixed probability density centered at $\underline{\hat{\tilde{\boldsymbol{\mu}}}}(\underline{\mathbf{x}}, \alpha)$ and let the interior of the ellipsoid be denoted by $E_c(\underline{\mathbf{x}}, \alpha)$. Using the method developed in the proof of Theorem 2, for any $\epsilon > 0$, c can be chosen such that

$$1 - \frac{\epsilon}{2\delta} < \Pr(\underline{\hat{\tilde{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \in E_c(\underline{\mathbf{x}}, \alpha)) \ .$$

Let λ denote the largest eigenvalue of $\hat{\tilde{\Sigma}}$. Then the largest (semi-principal) axis of the ellipsoid is $a = \sqrt{c\lambda}$. For $1 \leq i \leq p$, we know that

$$\hat{\tilde{\mu}}_i(\underline{\mathbf{x}}, \alpha) = \tilde{\mu}_i(\underline{\mathbf{x}}) + \rho_{i,n} \frac{\sigma_i}{\sigma_n} (\alpha - \mu_n)$$

and

$$\begin{aligned}\tilde{\mu}(\underline{\mathbf{x}}) &= \underline{\boldsymbol{\mu}}_p + \Sigma_{pm}\Sigma_m^{-1}(\underline{\mathbf{x}} - \underline{\boldsymbol{\mu}}_m) \\ &= \underline{\boldsymbol{\mu}}_p - \Sigma_{pm}\Sigma_m^{-1}\underline{\boldsymbol{\mu}}_m + \Sigma_{pm}\Sigma_m^{-1}\underline{\mathbf{x}} \ .\end{aligned}$$

Let $\Sigma_{pm}\Sigma_m^{-1}\underline{\mathbf{x}}_M = \underline{\mathbf{q}}'$ and for each i, $q_i = \min\{0, q_i'\}$. Then for $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$, $\tilde{\mu}_i(\underline{\mathbf{x}}) \geq g_i + q_i$ for all i, $0 \leq i \leq p$, where $\underline{\mathbf{g}} = \underline{\boldsymbol{\mu}}_p - \Sigma_{pm}\Sigma_m^{-1}\underline{\boldsymbol{\mu}}_m$. Let $\alpha_i(\underline{\mathbf{x}}) = (\sqrt{c\lambda} - \tilde{\mu}_i(\underline{\mathbf{x}}))\frac{\sigma_n}{\sigma_i\rho_{i,n}} + \mu_n$ for all $1 \leq i \leq p$. Then for $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$, $\alpha_i(\underline{\mathbf{x}}) \leq (\sqrt{c\lambda} - g_i - q_i)\frac{\sigma_n}{\sigma_i\rho_{i,n}} + \mu_n = \alpha_i < \infty$. Choose $\alpha^* = \max\{\alpha_1, \ldots, \alpha_p\}$. Then $\hat{\tilde{\mu}}_i(\underline{\mathbf{x}}, \alpha^*) \geq \sqrt{c\lambda}$ for all i and $E_c(\underline{\mathbf{x}}, \alpha^*) \subset \{\hat{\tilde{\underline{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha^*) \geq \underline{\mathbf{0}}\}$ for each $\underline{\mathbf{x}}, \underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$. Therefore, for all $\alpha > \alpha^*$ and $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$,

$$1 - \frac{\epsilon}{2\delta} < \Pr(\hat{\tilde{\underline{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \in E_c(\underline{\mathbf{x}}, \alpha)) \leq \Pr(\hat{\tilde{\underline{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \ .$$

$$\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) = \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\hat{\underline{\hat{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}}$$
$$+ \int_{\{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}\}/\{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M\}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\hat{\underline{\hat{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}}$$

$$> \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\hat{\underline{\hat{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}}$$

$$> \left(1 - \frac{\epsilon}{2\delta}\right) \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \mathbf{d}\underline{\mathbf{x}}$$

$$= \left(1 - \frac{\epsilon}{2\delta}\right) \Pr(\underline{\mathbf{0}} \leq \underline{\mathbf{X}}_m \leq \underline{\mathbf{x}}_M)$$

$$> \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}) - \frac{\epsilon}{2\delta}\delta - \frac{\epsilon}{2}$$

$$= \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}) - \epsilon \ .$$

$$\Rightarrow \left| \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}) - \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) \right| < \epsilon.$$

Similarly, for proving the lower limit on $\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha)$, let $q_i = \max\{0, q_i'\}$ for each i so that for $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$, $\tilde{\mu}_i(\underline{\mathbf{x}}) \leq g_i + q_i$ for all i, $0 \leq i \leq p$, where $\underline{\mathbf{g}}$ and $\underline{\mathbf{q}'}$ are as defined previously.

Let $\alpha_i(\underline{\mathbf{x}}) = (-\sqrt{c\lambda} - \tilde{\mu}_i(\underline{\mathbf{x}}))\frac{\sigma_n}{\sigma_i \rho_{i,n}} + \mu_n$ for all $1 \leq i \leq p$. Then for $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$, $\alpha_i(\underline{\mathbf{x}}) \geq (-\sqrt{c\lambda} - g_i - q_i)\frac{\sigma_n}{\sigma_i \rho_{i,n}} + \mu_n = \alpha_i > -\infty$. Choose $\alpha^{**} = \min\{\alpha_1, \ldots, \alpha_p\}$. Then $\hat{\tilde{\mu}}_i(\underline{\mathbf{x}}, \alpha^{**}) \leq -\sqrt{c\lambda}$ for all i and $E_c(\underline{\mathbf{x}}, \alpha^{**}) \subset \{\hat{\underline{\hat{\mathbf{X}}}}_p(\underline{\mathbf{x}}, \alpha^{**}) \leq \underline{\mathbf{0}}\}$ for each $\underline{\mathbf{x}}, \underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$.

Therefore, for all $\alpha < \alpha^{**}$ and $\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M$,

$$\begin{aligned}
&\Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \leq \underline{\mathbf{0}}) \geq \Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \in E_c(\underline{\mathbf{x}}, \alpha)) > 1 - \frac{\epsilon}{2\delta} \\
\Rightarrow\ & 1 - \Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \geq 1 - \frac{\epsilon}{2\delta} \\
\Rightarrow\ & \Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \leq \frac{\epsilon}{2\delta} \ .
\end{aligned}$$

$$\begin{aligned}
\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) &= \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} \\
&\quad + \int_{\{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}\} / \{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M\}} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \Pr(\hat{\underline{\mathbf{X}}}_p(\underline{\mathbf{x}}, \alpha) \geq \underline{\mathbf{0}}) \mathbf{d}\underline{\mathbf{x}} \\
&< \frac{\epsilon}{2\delta} \int_{\underline{\mathbf{0}} \leq \underline{\mathbf{x}} \leq \underline{\mathbf{x}}_M} f_{\underline{\mathbf{X}}_m}(\underline{\mathbf{x}}) \mathbf{d}\underline{\mathbf{x}} + \frac{\epsilon}{2} \\
&= \epsilon \ .
\end{aligned}$$

$$\Rightarrow \left| \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) - 0 \right| < \epsilon. \qquad \square$$

Lemma B.5:

$$\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n \geq 0) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}})$$

Proof. Consider the unconditional distribution of $\underline{\mathbf{X}}_{n-1}$. Then

$$\begin{aligned}\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}) &= \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}, \underline{\mathbf{X}}_p \geq \underline{\mathbf{0}}) \\ &= \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}})\ \Pr(\underline{\mathbf{X}}_p \geq \underline{\mathbf{0}} \mid \underline{\mathbf{X}}_m \geq \underline{\mathbf{0}})\ .\end{aligned}$$

Since $0 \leq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}) \leq \Pr(\underline{\mathbf{X}}_m \geq \underline{\mathbf{0}}) \Rightarrow$, it follows from Lemmas $B.3$ and $B.4$, there exists some α' such that

$$\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}) = \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha')\ .$$

For all $\alpha < \alpha'$, $\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) < \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}})$,
and for all $\alpha \geq \alpha'$, $\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n = \alpha) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}})$.

Case I: $\alpha' \geq 0$.

Then for all $\alpha < 0$

$$\int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}|X_n=\alpha}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}} < \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}$$

$$\Rightarrow \int_{\alpha<0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}|X_n=\alpha}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha < \int_{\alpha<0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha$$

$$\Rightarrow \int_{\alpha\geq 0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}|X_n=\alpha}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha \geq \int_{\alpha\geq 0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha$$

$$\Rightarrow \Pr(\underline{\mathbf{X}}_n \geq \underline{\mathbf{0}}) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}) \Pr(X_n \geq 0)$$

$$\Rightarrow \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n \geq 0) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}).$$

Case II: $\alpha' < 0$.

Then for all $\alpha \geq 0$

$$\int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}|X_n=\alpha}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}} \geq \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}$$

$$\Rightarrow \int_{\alpha\geq 0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}|X_n=\alpha}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha \geq \int_{\alpha\geq 0} f_{X_n}(\alpha) \int_{\underline{\mathbf{x}} \geq \underline{\mathbf{0}}} f_{X_1,\ldots,X_{n-1}}(\underline{\mathbf{x}})\ \mathbf{d}\underline{\mathbf{x}}\ d\alpha$$

$$\Rightarrow \Pr(\underline{\mathbf{X}}_n \geq \underline{\mathbf{0}}) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}) \Pr(X_n \geq 0)$$

$$\Rightarrow \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n \geq 0) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}).$$

Hence,

$$\Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}} \mid X_n \geq 0) \geq \Pr(\underline{\mathbf{X}}_{n-1} \geq \underline{\mathbf{0}}). \qquad \square$$

The result stated as Theorem 6.3 is given in [3] for hard-decision Viterbi decoding and i.i.d. Gaussian noise channel. The theorem as stated here hold for soft-decision Viterbi decoding and any symmetric i.i.d. channel - in particular, the mixed-distribution channel.

Theorem 6.3: $P_e \leq 1 - (1 - P_u)^L$.

Proof. Consider the mixed-distribution channel with parameter vector $\underline{\tilde{\mathbf{\Theta}}}$. From equation (3.4),

$$P_e \mid (\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) = \Pr(\cup_{l=0}^{L-1} \cup_{i=1}^{\nu_l} \{X_{i,l} < 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \ .$$

If P_c denotes the probability of correct decision,

$$P_c \mid (\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) = \Pr(\cap_{l=0}^{L-1} \cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \ .$$

The pairwise error-event statistics are conditionally jointly Gaussian given $\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}$. Thus, by repeated application of Lemma $B.5$,

$$P_c \mid (\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \geq \prod_{l=0}^{L-1} \prod_{i=1}^{\nu_l} \Pr(X_{i,l} \geq 0 \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \ .$$

All of the probabilities $\Pr(X_{i,l} \geq 0 \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}})$ are non-decreasing functions of $\mathrm{Var}(n_j)$ for each j. (I.e., they are *comonotonic* in $\mathrm{Var}(n_j)$ for each j.) Since $\underline{\tilde{\mathbf{\Theta}}}$ determines the values

of $\mathrm{Var}(n_j)$ for each j,

$$\begin{aligned}
P_c &= E_{\underline{\tilde{\Theta}}}\left[\prod_{l=0}^{L-1}\prod_{i=1}^{\nu_l}\Pr(X_{i,l}\geq 0 \mid \underline{\tilde{\boldsymbol{\Theta}}}=\underline{\tilde{\theta}})\right] \\
&\geq \prod_{l=0}^{L-1}\prod_{i=1}^{\nu_l} E_{\underline{\tilde{\Theta}}}\left[\Pr(X_{i,l}\geq 0 \mid \underline{\tilde{\boldsymbol{\Theta}}}=\underline{\tilde{\theta}})\right] \\
&= \prod_{l=0}^{L-1}\prod_{i=1}^{\nu_l}(1 - E_{\underline{\tilde{\Theta}}}\left[\Pr(X_{i,l}< 0 \mid \underline{\tilde{\boldsymbol{\Theta}}}=\underline{\tilde{\theta}})\right] \\
&\geq \prod_{l=0}^{L-1}(1 - \sum_{i=0}^{\nu_l} E_{\underline{\tilde{\Theta}}}\left[\Pr(X_{i,l}< 0 \mid \underline{\tilde{\boldsymbol{\Theta}}}=\underline{\tilde{\theta}})\right] \\
&\geq \prod_{l=0}^{L-1}(1 - \lim_{L\to\infty}\sum_{i=0}^{\nu_l} E_{\underline{\tilde{\Theta}}}\left[\Pr(X_{i,l}< 0 \mid \underline{\tilde{\boldsymbol{\Theta}}}=\underline{\tilde{\theta}})\right] \\
&= \prod_{l=0}^{L-1}(1-P_u) \\
&= (1-P_u)^L \ .
\end{aligned}$$

Hence,

$$P_e \leq 1-(1-P_u)^L \ .$$

□

Appendix C Proof of Theorem 6.4

Consider the mixed-distribution channel with parameter vector $\underline{\tilde{\mathbf{\Theta}}}$.

Theorem 6.4:

$$P_e \leq 1 - (1-p)^L \ .$$

Proof. As in the proof of Theorem 6.3,

$$P_c \mid (\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) = \Pr(\cap_{l=0}^{L-1} \cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \ .$$

Since the pairwise error-event statistics are conditionally jointly Gaussian given $\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}$, it follows from Corollary 2 of Lemma 2.1.1 in [16] (following a result of [15]) that

$$P_c \mid (\underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \geq \prod_{l=0}^{L-1} \Pr(\cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \ .$$

All of the probabilities $\Pr(\cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}})$ are comonotonic in $\mathrm{Var}(n_j)$ for each j. Thus,

$$\begin{aligned}
P_c &\geq E_{\underline{\tilde{\mathbf{\Theta}}}} \left[\prod_{l=0}^{L-1} \Pr(\cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \right] \\
&\geq \prod_{l=0}^{L-1} E_{\underline{\tilde{\mathbf{\Theta}}}} \left[\Pr(\cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \right] \\
&\geq \prod_{l=0}^{L-1} \lim_{L \to \infty} E_{\underline{\tilde{\mathbf{\Theta}}}} \left[\Pr(\cap_{i=1}^{\nu_l} \{X_{i,l} \geq 0\} \mid \underline{\tilde{\mathbf{\Theta}}} = \underline{\tilde{\theta}}) \right] \\
&= \prod_{l=0}^{L-1} (1-p) = (1-p)^L \ .
\end{aligned}$$

Thus,

$$P_e \leq 1 - (1-p)^L \ . \qquad \square$$

Bibliography

[1] M. B. Pursley, "The role of spread spectrum in packet radio networks," *Proceedings of the IEEE*, vol. 75, pp. 116–134, Jan. 1987.

[2] ——, "Performance evaluation for phase-coded spread-spectrum multiple-access communication - part 1: System analysis," *IEEE Transactions on Communications*, vol. COM-25, no. 8, pp. 795–799, Aug. 1977.

[3] M. B. Pursley and D. J. Taipale, "Error probabilities for spread-spectrum packet radio with convolutional codes and Viterbi decoding," *IEEE Transactions on Communications*, vol. COM-35, no. 1, pp. 1–12, Jan. 1987.

[4] A. J. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Transactions on Communications*, vol. COM-19, pp. 751–772, Oct. 1971.

[5] A. Swaminathan, D. L. Noneaker, and H. B. Russell, "The design of a channel-access protocol for a wireless ad hoc network with sectored directional antennas," *Ad Hoc Networks*, (to appear).

[6] S. Lin and D. J. Costello Jr., *Error Control Coding*. Upper Saddle River, NJ: Pearson-Prentice Hall, 2004.

[7] Consultative Committee for Space Data Systems, "Recommendation for space data systems standard, telemetry channel coding," *Blue Book*, vol. CCSDS, pp. 101.0–B–2, issue 2, Jan. 1987.

[8] M. Schwartz, W. R. Bennett, and S. Stein, *Communication systems and techniques*. New York: McGraw-Hill, 1966.

[9] L. C. Perez, J. Seghers, and D. J. Costello Jr., "A distance spectrum interpretation of turbo codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1698–1709, Nov. 1996.

[10] S. Benedetto and G. Montorsi, "Unveiling turbo-codes: Some results on parallel concatenated coding schemes," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 409–428, Mar. 1996.

[11] N. L. Johnson and S. Kotz, *Urn Models and Their Applications*. New York: John Wiley and Sons, 1977.

[12] I. M. Onyszchuk, "On the performance of convolutional codes," Ph.D. dissertation, California Institute of Technology, 1990.

[13] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding.* New York: McGraw-Hill, 1979.

[14] M. K. Simon, "Some new twists to problems involving the Gaussian probability integral," *IEEE Transactions on Communications*, vol. 46, no. 2, pp. 200–210, Feb. 1998.

[15] D. Slepian, "The one-sided barrier problem for Gaussian noise," *Bell System Technical Journal*, vol. 41, pp. 463–501, 1962.

[16] Y. L. Tong, *Probability Inequalities in Multivariate Distributions.* New York: Academic Press, 1980.