COLLABORATIVE DECODING AND ITS PERFORMANCE ANALYSIS

By

XIN LI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006
To my wife, Li, and my parents.
ACKNOWLEDGMENTS

First of all, I would like to give my sincere gratitude to my advisor, Dr. Tan F. Wong, for his advice during my doctorate endeavors. Without his patience, guidance and encouragement, none of this work would not have been possible. As a mentor, his wisdom, kindness and enthusiasm benefitted me greatly in both of my research work and life. I would also like to give special thanks to the co-chair of my supervisory committee, Dr. John Shea, for many fruitful discussions and great help. His valuable insights and ideas directly and significantly contributed to the work in this dissertation.

Many thanks go to my fellow graduate students in the laboratory. In particular, I appreciate Arun Avudainayagam for a lot of helpful discussions and collaboration, which directly contributed to the partial work of Chapter 4 in this dissertation. I also extend my gratitude to Dr. Jose A. B. Fortes and Dr. Shigang Chen for being my committee members, and I also thank Dr. Sanjay Ranka for ever being my committee member. I appreciate for their constructive suggestions and precious time. Meantime, great appreciation must go to the University of Florida for awarding me the Alumni Graduate Fellowship, which provided me a full four years support during my graduate study.

Last, but not the least, my sincere thanks go to my family for their endless love, continuous support and encouragement during my life. This work dedicates to all of them.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGMENTS</th>
<th>iv</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## CHAPTER

### 1 INTRODUCTION

1.1 Motivation ................................................................. 1
1.2 Multi-antenna Diversity Techniques ............................. 2
1.3 Distributed Array ......................................................... 5
1.4 Collaborative Decoding ................................................. 7
1.5 Scope of This Work ....................................................... 11

### 2 COLLABORATIVE DECODING IN A TWO-NODE DISTRIBUTED ARRAY

2.1 System Model ............................................................... 13
2.2 Collaborative Decoding for Rectangular Parity-Check Code .... 14
2.3 Collaborative Decoding for Convolutional Code ..................... 15
2.4 Summary ........................................................................ 20

### 3 COLLABORATIVE DECODING FOR CODED MODULATION

3.1 System Model ................................................................. 24
3.2 Iterative Demodulation and Decoding for BICM ...................... 26
3.2.1 Iterative Demodulation and Decoding Algorithm .................. 28
3.2.2 Effect of Mapping in BICM-ID ......................................... 29
3.3 Collaborative Decoding for BICM-ID with Rectangular Parity-Check Code .................................................. 33
3.4 Performance Evaluation ..................................................... 36
3.5 Summary ........................................................................ 39

### 4 COLLABORATIVE DECODING FOR DISTRIBUTED ARRAY WITH TWO OR MORE NODES

4.1 System Model for Distributed Array with Two or More Nodes .... 40
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5–1</td>
<td></td>
</tr>
<tr>
<td>Different choices of ( {p_j} ) and corresponding average information exchange amount ( \Theta ) with ( M = 8 ) for rate 1/2 CC(5, 7) code. ( \Theta ) is calculated with respect to the information exchange amount of MRC, ( \Theta_{\text{MRC}} ).</td>
<td>80</td>
</tr>
<tr>
<td>6–1</td>
<td></td>
</tr>
<tr>
<td>Different choices of ( {p_j} ) and the corresponding average information exchange amount ( \Theta ) with ( M = 6 ) for rate-1/2 CC(5, 7) code. ( \Theta ) is calculated with respect to the information exchange amount of MRC, ( \Theta_{\text{MRC}} ).</td>
<td>145</td>
</tr>
</tbody>
</table>
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–1</td>
<td>Linear combiner for a SIMO system</td>
</tr>
<tr>
<td>1–2</td>
<td>Distributed array</td>
</tr>
<tr>
<td>1–3</td>
<td>Iterative decoding</td>
</tr>
<tr>
<td>1–4</td>
<td>Collaborative decoding</td>
</tr>
<tr>
<td>2–1</td>
<td>A cluster of two nodes forming a distributed antenna array.</td>
</tr>
<tr>
<td>2–2</td>
<td>Example of rectangular parity-check code for packet of nine information bits.</td>
</tr>
<tr>
<td>2–3</td>
<td>Conditional probability of the event that an error occurs in the bits whose soft output magnitudes rank in the lowest percentiles.</td>
</tr>
<tr>
<td>2–4</td>
<td>Performance of collaborative decoding for the $32^2$ RPCC with information exchange between two receiving nodes over an AWGN channel.</td>
</tr>
<tr>
<td>2–5</td>
<td>Performance of collaborative decoding for the $32^2$ RPCC with information exchange between two receiving nodes over a Rayleigh fading channel.</td>
</tr>
<tr>
<td>2–6</td>
<td>Performance of collaborative decoding for the CC(5,7) with information exchange between two receiving nodes over an AWGN channel.</td>
</tr>
<tr>
<td>2–7</td>
<td>Performance of collaborative decoding for the CC(5,7) with information exchange between two receiving nodes over a Rayleigh fading channel.</td>
</tr>
<tr>
<td>3–1</td>
<td>System model of BICM-ID with RPCC</td>
</tr>
<tr>
<td>3–2</td>
<td>BER for BICM-ID with $32^2$ RPCC and 8PSK in the two-node distributed array over AWGN channel.</td>
</tr>
<tr>
<td>3–3</td>
<td>BER for BICM-ID with $32^2$ RPCC and 8PSK in the two-node distributed array over Rayleigh fading channel.</td>
</tr>
<tr>
<td>3–4</td>
<td>Average SNR at $10^{-5}$ BER versus spectral efficiency for BICM-ID with $32^2$ RPCC in a two-node distributed array over AWGN channels.</td>
</tr>
<tr>
<td>3–5</td>
<td>Average SNR at $10^{-5}$ BER versus spectral efficiency for BICM-ID with $32^2$ RPCC in a two-node distributed array over Rayleigh fading channels.</td>
</tr>
</tbody>
</table>
4–1 Distributed array with multiple nodes. ................................................. 41
4–2 Typical probability distribution functions of soft-output for convolutional codes .................................................. 47
4–3 BER performance of collaborative decoding with LRB exchange for the cases of $M = 2, 3, 4, 6$ and 8 on AWGN channels, where CC$(5, 7)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used. .......................................................... 52
4–4 BER performance of collaborative decoding with LRB exchange on Rayleigh fading channels, parameter settings are the same as Fig. 4–3. .................. 52
4–5 BER performance of collaborative decoding with MRB exchange on AWGN channels, where $\{p_j\} = \{0.1, 0.2\}$ are used. ........................ 53
4–6 BER performance of collaborative decoding with MRB exchange on Rayleigh fading channels, parameter settings are the same as Fig. 4–5. .............. 54
4–7 Comparison of information exchange amount with respect to MRC for LRB and MRB exchange schemes ................................... 54
5–1 System model for collaborative decoding process. ......................... 58
5–2 Empirical pdfs of extrinsic information generated by the MAP decoders in successive iterations in collaborative decoding with the LRB exchange for $M = 6$ and $E_b/N_0 = 3$dB on AWGN channels, where the maximum free distance 4-state nonrecursive convolutional code is used. .......................... 59
5–3 Comparison of mean and variance of the extrinsic information from the density evolution model and that from the actual collaborative decoding process ......................................................... 63
5–4 Comparison of threshold estimated from the density evolution model that from the actual collaborative decoding process ....................... 64
5–5 Comparison of the proposed bounds, simulation results for the cases of $M = 2$ and 6 on AWGN channels, where CC$(5, 7)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used. ................................................................. 78
5–6 Comparison of the proposed bounds, simulation results in the last iteration for the cases of $M = 2, 3, 4, 6$ and 8 on AWGN channels, where CC$(15, 17)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used. .................................................... 79
5–7 Comparison of performance for $M = 8$ and CC$(5, 7)$ on AWGN channels with different choices of $\{p_j\}$ in Table 5–1 ........................................ 81
5–8 Asymptotic performance for $M = 8$ and CC$(5, 7)$ on AWGN channels with different choices of $\{p_j\}$ in Table 5–1 ........................................ 81
6–1 Empirical pdfs of extrinsic information generated by the MAP decoder at successive iterations in collaborative decoding with MRB exchange for $M = 6$ on an AWGN channel with $E_b/N_0 = 5$dB, for which the maximum free distance 4-state nonrecursive convolutional code is used.

6–2 Empirical pdfs of extrinsic information generated by the MAP decoders in successive iterations in collaborative decoding with the MRB exchange for $M = 8$ and $E_b/N_0 = 8$dB on independent Rayleigh fading channels, where the maximum free distance 4-state nonrecursive convolutional code is used.

6–3 Density evolution model for collaborative decoding process

6–4 Comparison of mean and variance of the extrinsic information obtained from the density evolution model and that from the actual collaborative decoding process.

6–5 Comparison of threshold estimated from the density evolution model and that from the actual collaborative decoding process.

6–6 Comparison of the bounds and simulation results for the cases of $M = 6$ on an AWGN channel, where CC(5, 7) and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.

6–7 Comparison of the bounds and simulation results for the cases of $M = 6$ on an AWGN channel, where CC(15, 17) and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.

6–8 Comparison of the bounds and simulation results for the cases of $M = 6$ on an independent Rayleigh fading channel, where CC(5, 7) and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.

6–9 Comparison of the bounds and simulation results for the cases of $M = 6$ on an independent Rayleigh fading channel, where CC(15, 17) and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.

6–10 Comparison of the proposed bounds, simulation results in the last iteration for the cases of $M = 2, 3, 4$ and 8 on an AWGN channel, where CC(5, 7) and $\{p_j\} = \{0.1, 0.2, 1\}$ are used.

6–11 Comparison of the proposed bounds, simulation results in the last iteration for the cases of $M = 2, 3$ and 4 on an independent Rayleigh fading channels, where CC(5, 7) and $\{p_j\} = \{0.1, 0.2, 1\}$ are used.

6–12 Comparison of performance for $M = 6$ and CC(5, 7) on an AWGN channel with the different choices of $\{p_j\}$ in Table 6–1.

6–13 Comparison of performance for $M = 6$ and CC(5, 7) on Rayleigh fading channels with the different choices of $\{p_j\}$ in Table 6–1.
C–1 Summation order switch procedure for \( \{\nu_i, m_i\}_{i=0}^{j-1} \).
Antenna array processing is a common technique that utilizes spatial diversity to enhance the robustness of a digital communication system against deteriorative wireless transmission effects such as channel noise and fading. In order to obtain spacial diversity effectively, the physically connected antenna array elements in a traditional array are required to be separated far apart so that the received signals at different antennas are independent of each other. This situation usually makes the size of the antenna array too large to be feasible in many practical scenarios.

Unlike traditional arrays, distributed arrays alleviate the array size constraint by utilizing a cluster of independent physically separated receivers in a wireless network as its array elements. All the array elements (receiving nodes) can communicate with each other through a wireless broadcast channel. By exchanging information among these receiving nodes during the reception process, it is possible to obtain spatial diversity gain (or antenna gain) with such a distributed array. Since the information exchanging traffic load among the receiving nodes is an important issue for a wireless network, conventional receive diversity techniques such as maximum ratio combining (MRC) become inefficient in distributed arrays.
Collaborative decoding is an iterative receive diversity approach suitable for distributed array when error correction codes are used in the transmission process. Receive diversity is achieved by exchanging only a portion of decoding information among receiving nodes in collaborative decoding. By carefully selecting the decoding information, collaborative decoding can lower the information exchange amount, while providing performance close to that of MRC. Based on the statistic characteristics of the output of maximum a posteriori decoders, we study two information exchange schemes for collaborative decoding: least-reliable-bit (LRB) and most-reliable-bit (MRB) exchange schemes. Error performance of these two schemes under different transmission environments is investigated and compared with Monte Carlo simulations. Theoretical analysis is also carried out for collaborative decoding with the LRB and MRB exchange schemes when nonrecursive convolutional codes are used.
CHAPTER 1
INTRODUCTION

1.1 Motivation

Wireless communications primarily study the problem of reliable information (e.g., voice, video, images, text, data, etc.) transmission through an atmosphere medium (the wireless channel) by using a radio wave as the information carrier. Wireless communication techniques make it possible for people to communicate freely. Since the birth of the wireless communication era in the 1970’s, the demand for wireless transmission has been growing at a very rapid pace. With the developments of digital communication techniques, radio frequency circuit fabrication, and very-large-scale integrated circuit technologies, continual improvement of wireless communication techniques has been fulfilling the demands largely in the past few decades. However, the demands are still growing exponentially [1].

As is well known, the two underlying resources of wireless channel are radio bandwidth and transmitter power. Unfortunately, these two resources are very limited. Traditional single antenna communication techniques usually try to increase the capacity of a wireless channel by increasing the radio bandwidth and transmission power. However, with the rapid growth of wireless networks, bandwidth in usable spectrum has been highly saturated, while transmission power is limited due to the physical equipment constraints, e.g., limited battery life. Moreover, the transmission power should be restricted below some limitation in order to reduce the mutual interference among wireless communication devices using the same wireless channel. Thus, it becomes more and more difficult to fulfill the continuously and rapidly growing demand of wireless channel capacity by using single antenna techniques.
Another challenge in wireless communication is the hostile nature of the wireless channel. One common problem in signal transmission through any channel is additive noise [2]. The additive noise is usually modeled as statistically independent Gaussian noise with a flat power spectral density. This noise is also called thermal noise. The primary source of thermal noise is the internal components such as resistors and solid-state devices used in the receiver. When the transmitted signal goes through the receiver, the data symbol will be inevitably corrupted by the thermal noise. Interference, as an external performance degradation factor, is another challenge in wireless communication systems. Signals from other transmitters using the same wireless channel are usually the significant sources of interference. Besides noise and interference, fading is also one of the main channel impediments in wireless communication. Due to the nature of radio signals and the propagation characteristics of the wireless channel, signals transmitted through a wireless channel can suffer from attenuation, amplitude, phase and multipath distortion [3].

In order to combat the severe channel impairments due to fading and noise without excessively increasing the transmission power, it is indispensable to adapt some auxiliary wireless communication techniques different from the traditional ones used in single antenna systems. In this scenario, multi-antenna, or space, diversity techniques are particularly attractive because they can be readily combined with other forms of diversity and offer dramatic performance gains when other forms of diversity are unavailable [4, 5].

1.2 Multi-antenna Diversity Techniques

Multi-antenna diversity is widely considered to be the most promising avenue for significantly increasing the bandwidth efficiency of wireless data transmission systems. In multi-antenna diversity techniques, diversity is obtained by employing multiple antennas (also called an antenna array) at the transmitter and/or the receiver. The basic idea behind the multi-antenna diversity techniques is that, if the antennas
are placed sufficiently far apart, the channel fading between different antenna pairs will become more or less independent. Hence independent signal paths are created between the transmitter and the receiver. Reliable communication can be guaranteed only if one of the independent paths is strong.

If multiple antennas are employed at the receiver end but only one antenna is used for the transmitter, then the channel between the transmitter and receiver is called a single-input, multi-output (SIMO) channel. The space diversity obtained in a SIMO channel is called receive diversity. If multiple antennas are employed for the transmitter only, then the channel is called a multi-input, single-output (MISO) channels. Diversity in a MISO channel is called transmit diversity. If multiple antennas are employed for both the transmitter and receiver, then the channel is called a multi-input, multi-output (MIMO) channel. In this case, both transmit and receive diversities are provided by the channel. In this research work, we only consider SIMO channels; i.e., only receive diversity is studied here.

For a SIMO system, there are several ways to obtain receive diversity. Usually, the independent fading paths are combined to obtain a resultant signal that is then passed through a standard demodulator and/or decoder. Most combining techniques are linear: the output of the combiner is just a weighted sum of the received signals at different antenna array elements [6]. Fig. 1.2 shows a linear combiner for a SIMO system. In the figure, we suppose that the receive antenna array contains \( M \) antenna elements and a signal \( s(t) \) is transmitted through a flat fading wireless channel. These antenna elements are far enough apart so that \( M \) independent fading channels between the transmitter and receiver are created. Let \( r_i(t) \) denote the received signal at the \( i \)th antenna, then it can be expressed as

\[
r_i(t) = a_i s(t) + n_i(t), \quad i = 1, \ldots, M,
\]
where $a_i$ is the complex fading gain of the $i$th fading channel and $n_i(t)$ is the additive white Gaussian noise (AWGN) at the $i$th antenna. Then under the assumption that the channel fading gains are perfectly known, the optimal choice of the combining weights $\alpha_i$ is the conjugate of the channel fading gain $a_i$ for all $i$ [2]. The resultant combiner output signal $r_\Sigma(t)$ is given by

$$r_\Sigma(t) = \sum_{i=1}^{M} a_i^* r_i(t).$$

This optimum combining technique is known as *maximum ratio combining* (MRC). In fact, MRC maximizes the signal-to-noise ratio (SNR) of the output signal, which increases linearly with the number of independent fading channels $M$ [6]. The MRC combiner achieves the full receive diversity order of the channel and provides the optimal performance in comparison with other receive diversity techniques.

Figure 1–1: Linear combiner for a SIMO system
1.3 Distributed Array

It has been shown that the potential gain in channel capacity of multi-antenna systems over single-antenna system is rather large under the assumption of independent fading and noises at different receiving antennas [5]. However, fading correlation does exist when the elements are not spaced sufficiently far apart in practice. This can significantly reduce the capacity of the multiple-antenna system [7]. It is well known that increasing antenna spacing can decorrelate the multiple channels created by the antenna array. Thus, in order to make the independence assumption valid for the multi-antenna system, antenna elements in the arrays must be spaced far apart enough. Since conventional antenna arrays are composed of several physically connected antenna elements, this requirement implies a big size for the arrays.

The required antenna separation depends on the local scattering environment as well as on the carrier frequency. For a mobile transmitter which is near the ground with many scatterers around, the channel decorrelates over shorter spatial distances, and typical antenna separation of half to one carrier wavelength is necessary. For base stations on high towers, larger antenna separation of several to 10s of wavelengths may be required [3]. The carrier wavelength of a radio signal at frequency $f$ is given by $\lambda = c/f$, where $c = 3 \times 10^8 \text{m/s}$ is the speed of light. For illustration, consider the concrete example of a uniform linear antenna array, where the antennas are evenly spaced on a straight line. Suppose the multi-antenna system works at the carrier frequency of 2GHz; then carrier wavelength is about 0.15m. Thus, for a uniform linear array of 8 antennas with small scatterers, the length of the array will be larger than 3 feet! Antenna array of such a size usually is too large to be feasible in many practical scenarios. The physical size of the array will limit the applicability of spatial diversity techniques, especially for mobile applications.

To overcome the physical constraint of conventional multi-antenna arrays and take advantage of diversity techniques, we consider a network-based approach to
obtain spatial diversity without the use of physically connected antenna arrays. This approach makes use of the fact that communicating nodes in a local wireless network are inherently physically separated in space. When a remote source transmits a message through the wireless channel to a cluster of nodes, a SIMO channel will be essentially created between the source and the cluster of receiving nodes. Usually, these receiving nodes are far enough apart that independent fading at different nodes can be guaranteed. Meanwhile, the nodes are in close proximity so that simple lower-power, high-rate, reliable signaling techniques are permitted for the communications with the cluster. Hence, these nodes can coordinate their reception processes to effectively form a distributed antenna array. In this way, spatial diversity can be obtained through collaboration and communication among the nodes in the cluster.

Fig. 1–2 illustrates the concept of the distributed antenna array. In contrast to a conventional multi-antenna array in Fig. 1.2, where the received signals at all antenna elements are collected and processed in a centralized manner, each node in the distributed array is an integral receiver and possesses the capability of processing its received signal independently. From the viewpoint of the whole cluster, the reception process can be thought as a distributed process performed at different nodes in the array. Spatial diversity is then achieved by exchanging information among the
nodes through the local wireless network during the distributed processing. It is the combination of the distributed processing and local wireless network that allows us to overcome the physical constraint of conventional multi-antenna arrays.

It is worthwhile to point out that the kind of topology depicted by Fig. 1–2 is applicable to many practical wireless systems. For instance, consider a cellular system in which a mobile unit is within the range of multiple base stations. The base stations, which are linked together by optical or high-speed wired connections, receive independent copies of the transmitted signal from the mobile unit and jointly process the received signals to gain diversity from the independent channels. The same scenario applies to a wireless LAN system in which the base stations are replaced by access points, and the links joining the access points are usually wired Ethernet links. For a military communication example, consider a cluster of local sensors in a sensor network. The close proximity of the sensors allows the wireless links between the nodes to be very high speed, while requiring only low-power and low-complexity processing. A transmitter, either from another cluster in the network or external to the sensor network, sends a signal to this cluster. Each sensor receives a copy of the transmitted signal and processes the signal in a distributed manner using information from other sensors. The same scenario applies to inter-group communications between small groups of soldiers, each carrying a mobile communicator, or to a group of collaborating mobile users communicating with a base-station in a cellular network [8, 9].

1.4 Collaborative Decoding

In distributed processing, the information exchanging traffic load among the array nodes is an important issue for distributed arrays. It is undesirable to exchange an extensive amount of information in the reception process because the wireless network resource is limited. Conventional diversity techniques using linear combining described in Section 1.2 become expensive in terms of the information exchange
amount because all the received symbols at each node need to be forwarded through the network in order to achieve the full spatial diversity. In fact, it can be shown from the information-theoretic viewpoint that it is possible to achieve the full diversity advantage with a much smaller amount of information exchange than used in the common combining techniques such as MRC.

To explore efficient diversity techniques for distributed arrays, information must be exchanged selectively and that information must be used effectively at the receiving nodes. Usually, the received signal before the reception processing (such as detection, demodulation, decoding, etc.) suffers minimum information loss caused by the performance constraint (or capability) of the receiver and the inherent uncertainty in the communication systems. However, the information or signal before processing usually suffers maximum corruption caused by fading and noise compared with the information after processing. Thus, we can directly learn much less about the transmitted message from the signal before than that after processing. Although it may suffer certain information loss due to the receiving procedure, the information after processing usually reflects the true message with high confidence, Thus it is possible to use the information after processing effectively for exploiting spatial diversity. Moreover, information after processing may possess desired properties such that an effective information selection method can be adopted to reduce the information exchange amount.

Error correction coding provides the capability of detecting and correcting bit errors encountered in the transmission process. It is one of the most often used techniques in wireless communication systems. The maximum a posteriori (MAP) decoding is widely used in error correction coding techniques. MAP decoder decodes message bits by finding the possible ones that maximize their a posteriori probabilities and output the maximum a posteriori probabilities for each bit. The output is often
Figure 1–3: Iterative decoding

expressed in log-likelihood ratio (LLR) form as

$$L(\hat{x}) = \log \frac{P(\hat{x} = +1 | y)}{P(\hat{x} = -1 | y)},$$

where $\hat{x}$ is the decision of information bit $x$, $y$ is the channel observation, $P(\hat{x} = +1 | y)$ and $P(\hat{x} = -1 | y)$ are the a posteriori probabilities for $\hat{x}$ to be +1 and −1 given $y$, respectively. The LLR value does not only reflect the sign of a binary bit, but also indicates the reliability of the decision. It turns out that the output of MAP decoders can be the proper information to exchange for obtaining spatial diversity in distributed arrays.

The capacity-approaching turbo codes [10] have aroused great attention and have been extensively studied since their introduction. Turbo codes have become a landmark in the field of error correction coding. The key idea of turbo codes is using iterative, or “turbo”, decoding to exploit the multi-component code structure of the turbo encoder by associating a decoder with each of the component codes. In the decoding procedure, each decoder performs MAP decoding or any soft-in, soft-out (SISO) decoding that approximates the MAP decoding for its corresponding code component. The decoders help each other by using the extrinsic information (output) generated in the decoding process of the other decoders as a priori information for their own decoding process. By repeating the procedure in an iterative fashion,
Figure 1–4: Collaborative decoding

turbo codes can achieve the performance close to the Shannon capacity limit. Fig. 1–3 shows the iterative decoding procedure with two decoding components. The idea of iterative decoding is then generalized to many transmission systems with multiple code components parallel or serially concatenated together, such as coded modulation, iterative detection and equalization systems. This iterative decoding approach is usually very powerful and exhibits near-capacity performance.

With the above considerations, we study a new approach to achieve diversity called collaborative decoding in distributed arrays when error correction codes are used in the transmission process. The basic idea of collaborative decoding is to extend the iterative decoding techniques to the distributed array scenario. Fig. 1–4 depicts how the idea of iterative decoding is extended to collaborative decoding. By viewing receiving nodes in the array as a set of physically separated decoding components and the information exchanging process as the extrinsic information feedback process in Fig. 1–3, the typical iterative decoding procedure can be performed in a distributed array.
For MAP decoders, we notice that for the bits with high decoding reliability in previous iterations, the contribution of their \textit{a priori} information to the average decoding performance is marginal. However for the less reliable bits, the contribution of their \textit{a priori} information can be significant. This fact makes it possible for collaborative decoding to achieve diversity by exchanging only a portion of decoding information among the receiving nodes in a distributed array. It turns out that by carefully selecting the decoding information to exchange, collaborative decoding can lower the amount of information that must be exchanged in the array, while providing performance close to that of MRC. This advantage makes collaborative decoding an attractive diversity technique for distributed array systems.

\section*{1.5 Scope of This Work}

In this research work, we study the new approach of collaborative decoding with distributed arrays to achieve spatial diversity in wireless communications. We first investigate the possibility of using collaborative decoding in a two-node distributed array to obtain receive diversity when different channel coding techniques are adopted for AWGN and flat Rayleigh fading channels. Then the approach is extended to coded modulation systems with high-order signal constellations, which provide higher spectral efficiencies and are desired for bandwidth-constrained wireless channels. By exchanging only a small amount of information among the distributed array in contrast to conventional spatial diversity combining techniques, the collaborative decoding technique is shown to be able to achieve a significant spatial diversity gain and perform close to the optimal MRC.

Taking into account the scalability of the distributed array, we extend collaborative decoding into the more general case of an arbitrary number of nodes. Based on the statistic characteristics of the output of maximum \textit{a posteriori} decoders, we propose two efficient information exchange schemes for collaborative decoding: least-reliable-bit and most-reliable-bit exchange schemes. Error performance of these two
schemes under different transmission environments with different parameter settings is investigated and compared with Monte Carlo simulations.

To further study the proposed approach, theoretical analysis on the collaborative decoding technique is carried out. For analysis tractability, we consider the cases in which nonrecursive convolutional codes are used in the collaborative decoding procedure. The analysis is based on the assumption that the extrinsic information generated in the collaborating decoding process for nonrecursive convolutional codes can be approximately described by a class of Gaussian and generalized asymmetric Laplace distributions for AWGN and independent Rayleigh fading channels, respectively. With this assumption, we reduce the collaborating decoding to a density evolution model with a single MAP decoder, and propose a systematic method to evaluate the error performance of collaborating decoding semi-analytically. The analysis results show that with proper choices of parameters, collaborative decoding can achieve full diversity and approach the theoretical performance bounds asymptotically.
CHAPTER 2
COLLABORATIVE DECODING IN A TWO-NODE DISTRIBUTED ARRAY

In this chapter, we investigate the possibility of collaborative decoding achieving spatial diversity in distributed array. We first focus on the simple case of a two-node network. Consider a pair of nodes that are connected via a communication channel that has relatively benign characteristics that permit simple lower-power, high-rate, reliable signaling techniques to be employed for communications between these two nodes. Typically, these two nodes are in close proximity. A distant transmitter sends a packet of coded data bits to the two nodes. Each of the two nodes receives an independent copy of the transmitted signal.

For the distributed array, we employ iterative decoding to extract important information from the received signal at each node, and only pass this information between the two nodes. More precisely, each node decodes the signal that it receives and generates reliability estimates (soft outputs) for the transmitted data bits. The two nodes then exchange soft outputs of a small portion of the bits that are least reliable. Upon receiving additional information about the least reliable bits from another node, a node will restart the decoding process. This process of information exchange and iterative decoding then continues for a number of iterations. The objective is to obtain the maximum degree of diversity advantage from the signals received at the two nodes, while requiring a minimum amount of information exchange between them.

This chapter is primarily based on the work of Wong et al. [8, 9]. We will present the results of the simulations that we carried out to investigate the viability of the proposed distributed iterative decoding approach. In Section 3.1, we describe the system and channel model assumed in the simulations. In Sections 2.2 and 2.3,
we report decoding designs and simulation results employing a rectangular parity-check code and a convolutional code to encode packets from the distant transmitter, respectively. In Section 2.4, we discuss the potentials of the proposed distributed iterative decoding approach in different application scenarios.

2.1 System Model

We consider a system with the topology shown in Fig. 2.1. A distant transmitter sends a packet of coded data bits to the two receiving nodes. For simplicity, we assume that the two nodes can communicate with each other reliably. We are only interested in the communication link from the distant transmitter to the two nodes. We assume that the channels from the transmitter to the two receiving nodes are independent. We further assume that the coded bits from the transmitter are modulated using binary phase-shift keying (BPSK). After matched-filtering and proper normalization, the decision statistics for the $i$th coded bit obtained at the two receiving nodes are

$$y_i^{(1)} = a_i^{(1)} x_i + n_i^{(1)},$$
$$y_i^{(2)} = a_i^{(2)} x_i + n_i^{(2)},$$

where $x_i$ is the BPSK symbol ($\pm 1$) representing the $i$th bit, and $n_i^{(1)}$ and $n_i^{(2)}$ are independent zero-mean, circular-symmetric complex Gaussian random variables with per-component variance $N_0/2$ representing the thermal noise components at the first and second receiving nodes, respectively. We consider two different channel models. The first model is the additive white Gaussian noise (AWGN) model. For AWGN
channels, both the channel gains $a_{i}^{(1)}$ and $a_{i}^{(2)}$ are 1, i.e., the normalized received energy per coded bit $E_{c} = 1$. The second model we consider is the flat Rayleigh fading model. For Rayleigh fading channels, $a_{i}^{(1)}$ and $a_{i}^{(2)}$, for all $i$, are modeled as independent zero-mean circular-symmetric unit-variance complex Gaussian random variables. This corresponds to the assumption of having a perfect channel interleaver and the normalized average received energy per coded bit $E_{c} = 1$. For both AWGN and Rayleigh fading models, we assume that perfect phase estimation is achieved and hence coherent demodulation is performed at each node. In the case of Rayleigh fading model, we assume that perfect channel state information is available at the nodes.

2.2 Collaborative Decoding for Rectangular Parity-Check Code

In this section, we consider the design of the distributed iterative decoder when a rectangular parity-check code (PRCC) is employed to encode the data bits in the transmitted packet.

The rectangular parity-check code (RPCC) [11] is a punctured version of the product of two single parity-check codes. An example of a $3 \times 3$ RPCC is shown in Fig. 2.2. The code consists of single parity-check codes that operate on rows and columns of a square matrix that contains the information bits. RPCCs with large block sizes are very high-rate systematic codes that can be decoded by a very simple iterative algorithm [11, 12, 13]. Note that for a packet of $N^2$ bits, the number of parity bits is $2N$ ($N$ bits each in the horizontal and vertical directions). Thus, the rate of the RPCC is $N^2/(N^2 + 2N) = N/(N + 2)$. Clearly, as $N$ becomes large, the rate of the RPCC code becomes very high. Maximum a posteriori (MAP) decoding of the RPCC can be approximately performed by an iterative decoding procedure that treats the RPCC as a parallel concatenation of the parity check codes defined along the rows and columns of the data bit matrix. For each component code, the
“soft-in/soft-out” (SISO) decoding module in Hagenauer et al. [11] and Wong et al. [13] amounts to the following simple procedure:

1. Find the two smallest magnitudes among all soft inputs on a row/column.
2. Take hard decisions on the row/column and check the parity.
3. For each data bit (expect the one with the minimum magnitude soft input) on the row/column, the extrinsic information is the minimum magnitude if the parity matches with the hard decision of that bit, otherwise the extrinsic information is the negative of the minimum magnitude. For the data bit with the minimum magnitude soft input, the second smallest magnitude is employed.
4. Pass the extrinsic information as \textit{a priori} information to the component code in the other direction.

The soft inputs for the first iteration are simply scaled channel observations [13] for both the AWGN and Rayleigh fading models. At the end, the extrinsic informations provided by the two component codes are added with the initial channel observation to give the soft output, based on which the bit decision is made. It is clear that
this decoding process is very simple. In [12], RPCCs and their extensions to higher dimensions are shown to be able to achieve performance near the capacity limit for transmission over AWGN and bursty channels for very high code rates. In [13], it was shown that RPCCs can be used to obtain a significant diversity gain on fading channels with virtually no penalty in information rate. For details of the RPCC encoding and decoding algorithms, see Appendix A.

As mentioned before, the key to performing distributed decoding that requires only small amount of information to be passed between the two nodes is to identify the set of bits that are likely to be in error. Since the MAP decoder outputs the a posteriori log-likelihood ratios of the data bits, the soft outputs of the iterative decoder above are good reliability measures for the data bits. For both the AWGN and Rayleigh fading models, a data bit with a small soft output magnitude is more likely to be in error. To illustrate this, we consider a packet that has errors and plot the conditional probability of the event that an error (given that it occurs) occurs in the bits whose soft output magnitudes rank in the lowest percentiles. We plot this probability in Fig. 2.2 for the $32^2$ RPCC. The conditional probability is estimated from Monte Carlo simulations after 5 decoding iterations. We can see from Fig. 2.2 that at a high enough $E_b/N_0$, essentially at the convergence abscissa, most of the errors will occur in the bits whose soft output magnitudes rank in the lowest, say, 5%.

Based on this observation, we can employ the following simple strategy to gain diversity advantage while requiring a small amount of information exchange between the receiving nodes. At first, each node decodes the data bits from the signal that it receives. After the decoding, each node ranks the data bits according to their soft output magnitudes. Then each node requests additional information from the other node for those bits whose soft output magnitudes rank in the lowest $x\%$. Upon receiving a request, a node sends the soft outputs of the requested bits generated
in its own decoding process. Each node will use the soft outputs obtained from the other node as a priori information to continue the iterative decoding process. The whole process then repeats with additional exchange of soft outputs between the two nodes.

To illustrate the advantage of this approach, consider a sample system in which a node requests additional information for 5% of the bits with the smallest soft output magnitudes at each iteration. A total of 3 iterations of information exchange occur between the nodes, i.e., altogether the overall traffic between the nodes is 15% (neglecting the overhead involved in the requesting protocol) of what is required by MRC. In the case of MRC, we assume that each node passes all its channel observations to the other node and maximally combines the channel observations before decoding. Fig. 2–4 shows the bit error rate (BER) performance\(^1\) of the 32\(^2\) BERs are plotted against \(E_b/N_0\) per receiving node in Figs. 2–4, 2–5, 2–6, and 2–7.

\(^1\) BERs are plotted against \(E_b/N_0\) per receiving node in Figs. 2–4, 2–5, 2–6, and 2–7.
RPCC over an AWGN channel. We see that the $32^2$ RPCC, which has a code rate of 0.94, provides a coding gain of about 3dB at $10^{-5}$ BER$^2$. With MRC, an additional 3dB “antenna gain” is obtained as expected. The most interesting observation from Fig. 2–4 is that we can obtain a 2.4dB gain (out of the maximum possible 3dB gain) using the soft output exchange and iterative decoding algorithm described before, exchanging only a total of 15% of all soft outputs.

For the case of Rayleigh fading, the BER curves are shown in Fig. 2–5. The diversity gain provided by the MRC is about 8dB at $10^{-5}$ BER. More interestingly in this fading case, we can get all of the 8dB diversity gain that MRC can provide.

---

$^2$ The BERs presented here are averages of the BERs at the 2 nodes. There is a slight difference between the BERs at the two nodes obtained from simulation. However, the difference is always small as expected because of the symmetry between the nodes. On the other hand, this observation indicates that none of the nodes are disadvantaged against each other in the iterative decoding process.
at 10^{-5} BER by using the soft output exchange and iterative decoding algorithm described before, exchanging only a total of 15% of all soft outputs.

### 2.3 Collaborative Decoding for Convolutional Code

In this section, we consider the design of collaborative decoding when a standard convolutional code (CC) is employed to encode the data bits in the transmitted packet. We employ a rate-1/2, non-systematic, non-recursive, 4-state CC with generator matrix \([1 + D^2, 1 + D + D^2]\). We will refer to this CC as CC(5,7), based on the octal representation of the generator polynomials. It is well-known [14] that this CC has the largest free distance of 5 among all rate-1/2, 4-state CCs. The MAP decoder for this CC is the SISO decoder [15] based on the well-known BCJR algorithm [16]. Here we employ the less complex max-log-MAP decoder [15] as an approximation to the MAP decoder.

We employ the same collaborative decoding strategy described in the previous section. The only difference in this case is that the channel observations do not
directly correspond to the soft inputs for the data bits due to fact that the CC is non-
systematic. The soft output of a bit is generated by summing the extrinsic information
generated at the current decoding iteration and the cumulative a priori information
from previous iterations. After the current decoding iteration, each node ranks the
bits according to their soft output magnitudes. Then each node requests additional
information from the other node for those bits whose soft output magnitudes rank
in the lowest $x\%$. Upon receiving a request, a node sends the soft outputs of the
requested bits generated in its own decoding process. Each node will use the soft
outputs obtained from the other node as a priori information to continue the iterative
decoding process. The whole process then repeats with additional exchange of soft
outputs between the two nodes.

Similar to the previous case, we consider a sample system in which a node re-
quests additional information for 5% of the data bits with the smallest soft output
magnitudes at each iteration. A total of 3 iterations of information exchange occur
between the nodes, i.e., altogether the overall traffic between the nodes is 7.5% (ne-
glecting the overhead involved in the requesting protocol) of what is required by MRC,
in which each node passes all its channel observations to the other node and max-
imally combines the channel observations before decoding. The packet size is 1024
data bits (2048 coded bits). Fig. 2–6 shows the BER performance of the CC(5,7) over
an AWGN channel. Similar to the case of the $32^2$ RPCC, we can obtain a 2.4dB gain
at $10^{-5}$ BER, out of the maximum possible 3dB antenna gain, using the soft output
exchange and iterative decoding algorithm described before, exchanging only a total
of 7.5% of information required by MRC.

For the case of Rayleigh fading, the BER curves are shown in Fig. 2–7. The
diversity gain provided by the MRC is about 6dB at $10^{-5}$ BER. As seen from Fig. 2–
7, we can get 5dB out of the 6dB diversity gain that MRC can provide at $10^{-5}$ BER.
Figure 2–6: Performance of collaborative decoding for the CC(5,7) with information exchange between two receiving nodes over an AWGN channel.

Figure 2–7: Performance of collaborative decoding for the CC(5,7) with information exchange between two receiving nodes over a Rayleigh fading channel.
by using the soft output exchange and iterative decoding algorithm described before, exchanging only a total of 7.5% of information required by MRC.

2.4 Summary

The results in Sections 2.2 and 2.3 clearly indicate the possibility of getting full, or close-to-full, diversity advantage by proper collaborative decoding of the signals received at different nodes with a small amount of information exchange between the nodes. The crucial points appear to be identifying the bits that need additional information from other nodes and employing proper iterative decoding techniques to make the best use of the additional information. We can obtain some very promising results even with the simple, ad-hoc design for the RPCC and CC presented in Sections 2.2 and 2.3. This leads us to believe that collaborative decoding can be a viable technique to improve the performance of wireless communication systems that have topologies similar to the one described in Fig. 2.1.
CHAPTER 3
COLLABORATIVE DECODING FOR CODED MODULATION

The exponential growth in demand of high bit-rate data transmission in wireless systems continuously propels the research of using antenna array to increase the capacity of wireless communication systems. Meanwhile, the use of error correction coding techniques also helps greatly in exploiting the capacity of wireless communication channels. As well known, the powerful channel coding techniques such as turbo codes and low-density parity check codes can attain the rates approaching the Shannon limit, primarily for AWGN channels with binary modulation. However, it is clear that these error-correction coding schemes reduce transmit power at the expense of increased bandwidth or reduced data rate.

Coded modulation, by combining binary error correcting codes and higher level modulation together, provides an effective method to achieve coding gain without using additional bandwidth, thus high bit-rate communication can be achieved. Hence, the spectrally-efficient CM technique is suitable for bandwidth-constrained channels especially.

The first spectrally-efficient coding breakthrough came when Ungerboeck [17] introduced a coded-modulation technique to jointly optimize both channel coding and modulation. Ungerboeck’s trellis-coded modulation (TCM), which uses multi-level/phase signal modulation and simple convolutional coding with mapping by set partitioning (SP), can provide considerable coding gain for AWGN channels. This scheme maximizes the minimum free Euclidean distance of a code, which is the dominating factor to determine the code performance for AWGN channels. However, this scheme usually gives low diversity order, and leads to a performance degradation over
a Rayleigh fading channel. One general solution to this drawback is to apply symbol interleaver to the TCM.

It was latter recognized by Zehavi [18] that the diversity order can be increased to the minimum number of distinct bits instead of channel symbols by using bit-wise interleaving to yield a better coding gain over a Rayleigh channel. Following Zehavi’s idea, Caire et al. [19] proposed the bit-interleaved code modulation (BICM) that increases the diversity order further to the minimum Hamming distance of the code, thus, leads to a performance improvement over fading channels. But the random modulation caused by bit interleaving decreases the minimum free Euclidean distance of the codes. So BICM was thought not suitable for AWGN channels [19].

However, BICM, developed primarily for fading channels, latterly turned out to be able to give very good performance for AWGN channels as well. Li and Ritcey [20, 21] showed that by using a simple iterative decoding (ID) with BICM (BICM-ID), the minimum free Euclidean distance degradation, hence, performance degradation, can be overcome. With soft-decision feedback, BICM-ID significantly outperforms TCM, and the performance is even comparable with Turbo-TCM for AWGN channels.

The authors [20] also concluded that at high SNR, SP mapping clearly outperforms Gray mapping for BICM-ID using soft-feedback. Besides, an advantage of BICM is its flexibility in design. In BICM, encoder is serially connected to modulation by a single bit-by-bit interleaver. This structure treating coding and modulation separately makes it very convenient to employ different structure codes with different code rate in the schemes. By using some powerful codes such as long parallel or serially concatenated turbo codes and iterative decoder, it is possible for BICM to obtain good performance closed to capacity over Gaussian channels [22].

In Chapter 2, we studied the collaborative decoding technique in a two-node distributed array with error correction coding. When BPSK signal is used in the transmission, the collaborative decoding technique described in Chapter 2 can obtain
a diversity gain close to that provided by MRC. In this chapter, we consider employing coded modulation (CM) in the distributed array system to explore the possibility of obtaining spatial diversity with higher spectral efficiency for bandwidth-constrained wireless channels. Similar to Chapter 2, we still consider the simple case of two-node distributed array. But we will only consider the cases when rectangular parity-check code codes are used. This chapter is mainly based on the work of [23].

The remainder of this chapter is organized as follows. In Section 3.1, we present the two-node distributed array system and channel model. In Section 3.2, the BICM iterative demodulation for RPCCs and the design of distributed decoding are described in detail. Following that, Monte Carlo simulation results for different signal constellations are shown in Section 3.4. Finally, summary is given in Section 3.5.

### 3.1 System Model

As in Chapter 2, we consider a distributed array system with two identical receiving nodes. A distant transmitter sends a block of modulated signal to the two receiver nodes, as shown in Fig. 2.1. The two receiver nodes are physically separated far apart enough that fading at each node is i.i.d.. Each individual node receives and decodes its received signal independently. For simplicity, we assume that the two nodes can communicate with each other reliably. We are only interested in the communication link from the distant transmitter to the two nodes.

The transmitter adopts a typical BICM approach [19], as shown in Fig. 3–1. A block of data bits $u$ to be transmitted are encoded with an RPCC encoder with code rate $R_c$. Then the coded bit stream $c$ are fed into a bit-wise random interleaver $\pi$, generating bit stream $v = \pi(c)$. After that, the bit stream $v$ is modulated onto a signal sequence $x$ over a 2-dimension signal set $\chi$ of size $|\chi| = M = 2^m$ by a $M$-ary modulator with a one-to-one binary map $\mu : \{0, 1\}^m \rightarrow \chi$. This signal sequence is then sent through the channel. The overall spectral efficiency of this system is $mR_c$ bits/symbol.
Here we use a memoryless fading channel model that includes AWGN channel as a special case. In this model, the received signal $y$ at the two antenna nodes corresponding to the transmitted signal $x \in \chi$ can be expressed as

$$y^{(1)} = g^{(1)} x + n^{(1)},$$
$$y^{(2)} = g^{(2)} x + n^{(2)},$$

where: i) $g^{(1)}$ and $g^{(2)}$ are channel fading gains. For AWGN channels, $g^{(1)} = g^{(2)} = 1$. For Rayleigh fading channels, $g^{(1)}$ and $g^{(2)}$ are independent circular-symmetric complex Gaussian random variables with $E[g^{(i)}] = 0$ and $E[|g^{(i)}|^2] = 1$ for $i = 1, 2$; ii) $n^{(1)}$ and $n^{(2)}$ are independent zero-mean, circular-symmetric complex additive Gaussian noise with covariance $E[|n^{(i)}|^2] = \sigma^2$ for $i = 1, 2$. We normalize the signal energy $E[|x|^2] = 1$. Thus, the average signal-to-noise ratio (SNR) is $1/\sigma^2$. In this channel model, we assume that perfect channel state information (CSI) $(g^{(1)}, g^{(2)})$ is available at the receiver nodes and hence coherent demodulation is performed at each node. With this model the pdf $p(y^{(i)}|x)$, for $i = 1, 2$, with perfect CSI is given by

$$p(y^{(i)}|x) = \frac{1}{\pi\sigma^2} \exp \left(-|y^{(i)} - g^{(i)} x|^2/\sigma^2\right).$$  \hspace{1cm} (3–1)
At each receiver node, we treat the modulation and code as two components of a concatenated coding system. By employing a maximum a posteriori (MAP) demodulator, we feed the extrinsic information from the RPCC decoder back to the demodulator as the a priori information to carry out the demodulation and decoding in an iterative manner. After some iterations, we exchange information for a portion of symbols between the two nodes and restart the demodulation and decoding processes.

3.2 Iterative Demodulation and Decoding for BICM

3.2.1 Iterative Demodulation and Decoding Algorithm

One important component in our bit-interleaved coded modulation system is the rectangular parity-check code. Another important component in the BICM-ID system is the iterative demodulation module. Based on the idea that performing demodulation and decoding in an iterative manner is a key to improve the performance of BICM [20, 21], we employ the receiver model as illustrated in Fig. 3–1. Since the encoding and decoding for rectangular parity-check codes is addressed in Section 2.2 and Appendix A, we here emphasize on the demodulation component.

To simplify the iterative decoding process, we first modify the demodulator to work in the log-likelihood ratio (LLR) domain. Suppose that each $m$-bit vector $v = (v_1, v_2, \ldots, v_m)$ from the interleaver are mapped into one signal $x$ out of the $2^m$ signals in the set $\mathbf{X}$ by mapping rule $\mu$, i.e., $x = \mu(v) \in \mathbf{X}$, at the modulator, and that the received signal corresponding to $x$ is $y$. Let $\ell^i(x)$ denote the $i$th ($i = 1, 2, \ldots, m$) bit of the label of $x$. For convenience, we assume that $\ell^i(x) = b$ is in the GF(2) with the elements $\{+1, -1\}$. In our soft demodulator, we will consider the MAP rather than maximum-likelihood (ML) bit metric. It is easy to see that the MAP bit metric
of $v_i = b \in \{+1, -1\}$ is given by

$$
\lambda(v_i = b, y) = \log P(v_i = b, y) \\
= \log \sum_{z \in \chi} p(y|z) P(z|v_i = b) P(v_i = b),
$$

(3–2)

where $p(y|z)$ is given in (4–3) according to our channel model. We assume a perfect bit-interleaver $\pi$ such that \{ $v_1, v_2, \cdots, v_m$ \} are independent to each other. With this assumption, we have

$$
P(z) = P(z = \mu(v_1, v_2, \cdots, v_m)) = \prod_{j=1}^{m} P(v_j = \ell_j(z)).
$$

(3–3)

Hence, the MAP bit metric can be simplified to

$$
\lambda(v_i = b, y) \approx \max_{z \in \chi^b_i} \left\{ \log p(y|z) + \sum_{j \neq i} \log P(v_j = \ell_j(z)) + \log P(v_i = b) + C \right\},
$$

(3–4)

where $\chi^b_i$ denotes the subset of all signal $z \in \chi$ with $\ell_i(z) = b$, and $C$ is a constant. Above, the approximation $\log(\sum_i a_i) \approx \max_i (\log a_i)$ is used. For convenience we choose the constant as

$$
C = -\frac{1}{2} \sum_{j=1}^{m} \left( \log P(v_j = +1) + \log P(v_j = -1) \right).
$$

(3–5)

Then the metric becomes

$$
\lambda(v_i = b, y) = \max_{z \in \chi^b_i} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j=1}^{m} \ell_j(z)L(v_j) \right\},
$$

(3–6)
where $L(v_j) = \log \left( P(v_j = +1) / P(v_j = -1) \right)$ is the \textit{a priori} LLR of bit $v_j$. Thus the soft value of bit $v_i$ in LLR form is computed by

$$
L(v_i|y) = L(v_i, y) = \lambda(v_i = +1, y) - \lambda(v_i = -1, y)
$$

$$
= L(v_i) + \max_{z \in \chi_i^+} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell_j(z)L(v_j) \right\}
- \max_{z \in \chi_i^-} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell_j(z)L(v_j) \right\}.
$$

(3–7)

Subtracting the \textit{a priori} LLR of $v_i$, $L(v_i)$, from (3–7) we can obtain the extrinsic information of $v_i$

$$
L_e(v_i) = \max_{z \in \chi_i^{+1}} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell_j(z)L(v_j) \right\}
- \max_{z \in \chi_i^{-1}} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell_j(z)L(v_j) \right\}.
$$

(3–8)

We treat this extrinsic information as the output of the soft demodulator. From (3–8), we can see that in order to obtain the extrinsic LLR of a bit of a signal, we need to use the \textit{a priori} LLRs of the other $m - 1$ bits and the channel observation of the signal as input.

With the modification above, the demodulation and decoding procedure can perform in an iterative way conveniently. In Fig. 3–1 we use $L^{(n)}(\cdot)$ to denote the LLR at the $n$th iteration. First, we initialize all the \textit{a priori} LLRs $L^{(n)}(v)$ and $L^{(n)}(u)$ to zeros for $n = 0$. At the $n$th iteration, when the channel observation $y$ of the transmitted signal sequence is received, we demodulate it using (3–8) to produce $L_e^{(n)}(v)$. After deinterleaver $\pi^{-1}$, $L_e^{(n)}(v) = \pi^{-1}(L_e^{(n)}(v))$ is fed into the RPCC decoder for decoding. Since the RPCC decoding is an SISO iterative algorithm, we shall use the extrinsic information $L_e^{(n-1)}(u)$, produced in the $(n - 1)$th iteration, as the \textit{a priori} information $L^{(n)}(u)$ of decoder in the $n$th iteration. The extrinsic information $L_e^{(n)}(u)$ generated by the RPCC decoder is then passed thought the interleaver $\pi$ and fed back as the \textit{a priori} information $L^{(n+1)}(v)$ for the soft demodulator again. After
a number of iterations the estimate of data bits \( \hat{u} \) is obtained from the hard decision on \( L^{(n)}(\hat{u}) \).

### 3.2.2 Effect of Mapping in BICM-ID

The mapping \( \mu \) has a significant effect on the performance of BICM-ID. For BICM, Gray code mapping outperforms set partitioning (SP) mapping [19]. However, when associated with the iterative demodulation, SP mapping outperforms Gray mapping at high SNR [20, 21]. This can be seen from (3–8) that, due to the property of the Gray mapping that the label of a symbol has only one bit different form its nearest neighbors, the effect of a priori LLRs can be weakened significantly. However, this is not the case for SP mapping. Thus the MAP demodulator can make a more effective use of the a priori information for SP mapping than for Gray mapping.

To illustrate the effect of constellation mapping, consider that the signal \( x = \mu(v_1, v_2, \cdots, v_m) \in \chi \) is transmitted and channel observation is \( y \). The MAP decision for bit \( v_i \) made by the demodulator at the \( n \)th iteration is

\[
\begin{align*}
\alpha_i^{(n)} &= \arg \max_{z \in \chi^{-1}_i} \{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell^i(z)L^{(n)}(v_j) \} \\
\beta_i^{(n)} &= \arg \max_{z \in \chi^1_i} \{ \log p(y|z) + \frac{1}{2} \sum_{j \neq i} \ell^i(z)L^{(n)}(v_j) \},
\end{align*}
\]

(3–9)

where \( \arg \max_{z \in \chi^b_i} \{ \cdot \} \) means finding the signal maximizing the expression in the braces from the constellation subset \( \chi^b_i \). From (3–9), we have \( \ell^i(\alpha_i^{(n)}) \neq \ell^i(\beta_i^{(n)}) \).

With this notation, the extrinsic LLR \( L_e^{(n)}(v_i) \) in (3–8) can be written as

\[
L_e^{(n)}(v_i) = \log \frac{p(y|\alpha_i^{(n)})}{p(y|\beta_i^{(n)})} + \frac{1}{2} \sum_{j \neq i} \left( \ell^i(\alpha_i^{(n)}) - \ell^i(\beta_i^{(n)}) \right)L^{(n)}(v_j).
\]

(3–10)
For $n = 0$, there is no a priori information available, i.e., $L^{(0)}(v) = 0$. Thus, the demodulator gives ML decision result at the 0th iteration,

$$
\alpha_i^{(0)} = \arg \max_{z \in \chi_i^{+1}} \{ \log p(y|z) \},
$$

$$
\beta_i^{(0)} = \arg \max_{z \in \chi_i^{-1}} \{ \log p(y|z) \};
$$

(3–11)

and

$$
L_e^{(0)}(v_i) = \max_{z \in \chi_i^{+1}} \{ \log p(y|z) \} - \max_{z \in \chi_i^{-1}} \{ \log p(y|z) \}.
$$

(3–12)

Usually, the pairwise error probability $P(x \to \hat{x})$ (i.e., the probability that the demodulator chooses $\hat{x}$ when $x$ is transmitted) in (3–12) is dominated by the minimum free Euclidean distance of the constellations. Let us only consider the case that $(\alpha_i^{(0)}, \beta_i^{(0)})$ satisfying (3–11) is a signal pair having the minimum Euclidean distance in the constellation $\chi$. Otherwise, the large Euclidean distance between $\alpha_i^{(0)}$ and $\beta_i^{(0)}$ will make the pairwise error probability so small that it can be neglected.

We assume that the demodulator makes a single error at bit $v_i$ for the symbol $x$ at the 0th iteration. If a Gray mapping is used in this case, then with the property of Gray code that a code has only one bit different from its nearest neighbors, we have

$$
\ell^j(\alpha_i^{(0)}) = \ell^j(\beta_i^{(0)}) \quad \text{for all } j \neq i.
$$

(3–13)

Suppose that in the following iteration the a priori LLRs for other bits are reliable, i.e.,

$$
\ell^j(\alpha_i^{(0)}) L^{(1)}(v_j) = \ell^j(\beta_i^{(0)}) L^{(1)}(v_j) \geq 0 \quad \text{for all } j \neq i.
$$

(3–14)

For any constellation point $z \neq \alpha_i^{(0)}$ or $\beta_i^{(0)}$, there exists at least one $j \neq i$ such that $\ell^j(z) L^{(1)}(v_j) \leq 0$. So the demodulator will make the choice $(\alpha_i^{(1)}, \beta_i^{(1)}) = (\alpha_i^{(0)}, \beta_i^{(0)})$ in (3–9). Thus, using (3–10) and (3–13), the demodulator still gives the wrong ML decision result on the bit $v_i$

$$
L_e^{(1)}(v_i) = \max_{z \in \chi_i^{+1}} \{ \log p(y|z) \} - \max_{z \in \chi_i^{-1}} \{ \log p(y|z) \}.
$$

(3–15)
With above argument, the error can not be corrected no matter how many times
the demodulator iterates in this case. Thus, by using Gray mapping, the iterative
demodulation can not improve the performance of BICM.

However, for a SP mapping, Eq. (3–13) is not true. Consequently, at high SNR,
with large extrinsic LLRs from previous iteration, the second term on the right-
hand side of (3–10) could help correcting the pairwise error. This is the reason
why SP mapping outperforms Gray code mapping for BICM-ID at high SNR. The
performance comparison between Gray and SP mappings in Section 3.4 will verify
this statement.

3.3 Collaborative Decoding for BICM-ID with Rectangular
Parity-Check Code

The presented BICM-ID scheme is readily applicable to a distributed array. As
we pointed out in Chapter 2, a decoded data bit with a small soft output magnitude
from the RPCC decoder is more likely to be in error. However, if the bit-based
strategy in [8] is used here to gain diversity from other receiving node, we will lose
the advantage against MRC in term of saving information exchange traffic when
the modulation order $M$ increases. Hence, we develop a symbol-based strategy for
BICM-ID to reduce the information exchanging traffic. At first, we define the symbol
reliability measure out of the decoder as

$$L(\hat{x}) = \log \frac{P(\hat{x})}{1 - P(\hat{x})} = \log \frac{P(\hat{x})}{\sum_{z \neq \hat{x}} P(z)}, \quad (3–16)$$

where $\hat{x} = \mu(\hat{v}_1, \cdots, \hat{v}_m) \in \chi$ is the estimate of transmitted signal $x$. For convenience,
we define symbol metric for each constellation point $z \in \chi$ as

$$\lambda(z) = \frac{1}{2} \sum_{j=1}^{m} \ell^j(z) L(\hat{v}_j), \quad (3–17)$$

where $L(\hat{v}_j)$ is the soft output of the coded bit $v_j$. This symbol metric reflects the
probability $P(x = z)$ given the LLRs $L(\hat{v}_j)$ for $j = 1, 2, \cdots, m$. In fact, $\hat{x}$ should be
the constellation point that has the largest reliability, i.e., 
\[ \hat{x} = \arg \max_{z \in \chi} \{ \lambda(z) \}. \]
Similar to (3–3)-(3–5), (3–16) can be simplified to
\[ L(\hat{x}) \approx \lambda(\hat{x}) - \max_{z \neq \hat{x}, z \in \chi} \{ \lambda(z) \} = \min_{j=1, \ldots, m} \{|L(\tilde{v}_j)|\}. \quad (3–18) \]
Since the LLR magnitude of a bit can be used as the measure of its reliability, (3–18) indicates that the reliability of a decoded symbol is determined by the soft value of its least reliable bit, which is basically in agreement with the bit-based idea in [8].

With this definition, the collaborative decoding procedure works as follows. After every \( I \) (\( I \geq 1 \)) iterations of demodulation and decoding, each node computes the symbol reliability \( L(\hat{x}) \) and rank the symbols according to their reliability. Then each node requests additional information from the other node for symbol \( x \) that \( L(x) \) ranks in the lowest \( a\% \). We denote the additional information for \( x \) as \( L_a(x) \). Suppose that the estimate corresponding to symbol \( x \) at the other node is \( \tilde{x} = \mu(\tilde{v}_1, \ldots, \tilde{v}_m) \), which may be different from \( \hat{x} \) since the assumption of independent channels. Upon receiving the request, a node sends: i) the reliability of the requested symbols generated in its own decoding process as the additional information, i.e., \( L_a(x) = L(\tilde{x}) \); ii) the hard decision of \( \tilde{x} \), \( \ell(\tilde{x}) \) for \( j = 1, 2, \ldots, m \), which is also the hard decision of \( (\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_m) \).

Herein, we adopt following strategy, a node does not request additional information for the symbol if a request has been made for it in all previous exchanges. In this case, the node will request information for the next symbol in the ranking order to make sure that the request for a total of \( N \cdot a\% \) symbols will be made for the current exchange, where \( N \) is the symbol block size. The advantage of this strategy is that the additional information can cover more symbols for a number of exchanges.

After the exchange, as shown in Fig. 3–1, each node will use \( L_a(x) \) and the hard decision \( \ell(\tilde{x}) \) (\( j = 1, 2, \ldots, m \)) obtained from the other node to reconstruct an additional symbol metric \( \lambda_a(z) \) similar to (3–17) for each possible constellation point.
\( z \in \mathbf{X} \). Since \( \ell_j^{\hat{x}}(\tilde{x}) \) is the hard decision of bit \( \tilde{v}_j \), we have

\[
L(\tilde{v}_j) = \ell_j^{\hat{x}}(\tilde{x}) \mid L(\tilde{v}_j)\mid .
\] (3–19)

From (3–18) we can see that \( \mid L(\tilde{v}_j)\mid \geq L_a(x) \). This means each bit in \( \hat{x} \) has as least a reliability of \( L_a(x) \). Now we replace \( \mid L(\tilde{v}_j)\mid \) with \( L_a(x) \) for \( j = 1, 2, \cdots, m \) in (3–19), which is equivalently to set the reliability of all its bits the same as the reliability of a symbol. Thus, we can construct the additional symbol metric as

\[
\lambda_a(z) = \frac{1}{2} \sum_{j=1}^{m} \ell_j(z) \ell_j(\hat{x}) L_a(x).
\] (3–20)

This additional symbol metric is then used as the \textit{a priori} information for demodulation, and (3–6) becomes

\[
\lambda(v_i = b, y) = \max_{z \in \mathbf{X}^b_i} \left\{ \log p(y|z) + \frac{1}{2} \sum_{j=1}^{m} \ell_j(z)L(v_j) + \delta \lambda_a(z) \right\},
\] (3–21)

where \( \delta < 1 \) is a scaling factor used to reduce the effect of error propagation. Usually, \( \delta \) can be set to \( 0.6 \sim 0.7 \).

In the following \( I \) iterations, the whole process then repeats with additional exchange of symbol reliability and its hard decision between the two nodes. Note that in this strategy we just need to exchange one real number \( L(\hat{x}) \) and \( m \) bits \( \ell_j(\hat{x}) \) \((j = 1, 2, \cdots, m)\) for each symbol. However, for MRC, one needs to exchange a complex number \( y \) (channel observation) and a real number \( |g| \) (magnitude of fading gain, for AWGN channel no need to exchange it since \( |g| = 1 \)) for each symbol. This means we just require less than \( 2/3 \) (for Rayleigh fading channel) of or equal (for AWGN channel) to the exchanging traffic of MRC for each symbol, meanwhile we only need to exchange information for a portion of symbols. Hence with this symbol-based strategy, we can reduce the required information exchange traffic significantly.
3.4 Performance Evaluation

In this section, we examine the performance of the proposed distributed array scheme by Monte Carlo simulations. In the simulations, we set the packet size to 1024 data bits, i.e., the data bits are arranged into a $32 \times 32$ matrix for the RPCC encoding. With this block size, the RPCC gives a code rate of 0.94. In the decoding procedure, a node requests additional information for 15% of the symbols with the smallest reliability at the beginning of every 10 iterations after the first 10. For instance, 3 exchanges cause an overall traffic approximately equal to 30% (for Rayleigh fading channel) or 45% (for AWGN channel) of what required by MRC. In the case of MRC, we assume that each node passes all its channel observations and fading gains to the other node and maximally combines the channel observations before demodulation. Simulations show the bit error rates (BER) at the two nodes are almost the same as each other. So we take the average of them as the performance of the distributed array system.

Fig. 3–2 shows the BER performance of BICM-ID with $32^2$ RPCC in the distributed array over AWGN channels when 8PSK with Gray and SP mapping are used. In the figure, $E_b$ is the received energy per bit per antenna. With MRC, about 3dB spatial diversity gain can be achieved for both mappings. With our distributed array approach, we obtain a 2.4dB and 1.4dB gain for Gray and SP mapping at the traffic cost of 45% (i.e., 3 exchanges in total) of MRC. Fig. 3–3 shows the BER curves for Rayleigh fading channels. The spatial diversity gain provided by MRC is about 8.5dB for both Gray and SP mapping. By exchanging a total of 20% (i.e., 2 exchanges in total) of the information amount required for MRC, our distributed BICM-ID system obtain a 8.3dB and 7.3dB gain at the BER of $10^{-5}$ for Gray and SP mapping, respectively.
Figure 3–2: BER for BICM-ID with $32^2$ RPCC and 8PSK in the two-node distributed array over AWGN channel.

Figure 3–3: BER for BICM-ID with $32^2$ RPCC and 8PSK in the two-node distributed array over Rayleigh fading channel.
Figure 3–4: Average SNR at $10^{-5}$ BER versus spectral efficiency for BICM-ID with $32^2$ RPCC in a two-node distributed array over AWGN channels.

In Figs. 3–4 and 3–5, we show the average SNR ($E_s/N_0$) at BER of $10^{-5}$ versus spectral efficiency for the two-node distributed array system for different constellations with Gray mapping\(^1\) and SP mapping over AWGN channels and Rayleigh fading channels, respectively. The average SNR can be computed approximately by $\text{SNR} = mR_cE_b/N_0$, where $m$ is the number of bits per symbol carrying, and $R_c$ is the code rate of RPCC.

We can see that for both AWGN and Rayleigh fading channels the proposed distributed BICM-ID approach can achieve almost the diversity gain provided by MRC, but with only exchanging 20% (2 exchanges in total) and 45% (3 exchanges in total) of the amount of information required by MRC between the two nodes for AWGN channel and Rayleigh fading channel, respectively. This especially shows the advantage of our approach for fading channels.

\(^1\) For 32QAM, quasi-Gray mapping is used because Gray mapping is impossible in this case.
Figure 3–5: Average SNR at $10^{-5}$ BER versus spectral efficiency for BICM-ID with $32^2$ RPCC in a two-node distributed array over Rayleigh fading channels.

### 3.5 Summary

In this chapter, we have investigated the use of collaborative decoding for a two-node distributed array, in which high-order constellations with iterative demodulation and RPCCs are used. The scheme of bit-interleaved code modulation is used, with an iterative demodulation and decoding approach adopted in the receiver. We develop a SISO demodulation algorithm suitable for iteration. The effect of different choices of bit-to-symbol mapping is analyzed. In order to obtain efficient collaborative decoding schemes, we propose a symbol-based information exchange strategy for the BICM-ID system, which is different from the bit-based strategy used for the BPSK systems in Chapter 2. Monte Carlo simulation results show that, by using our collaborative decoding technique, a significant diversity gain can be obtained with a relatively small amount of information exchange between the independent and physically separated receiving nodes. This results in high spectral efficiencies under both AWGN and Rayleigh fading channels.
In the previous chapters, different collaborative decoding techniques are studied for two-node distributed arrays. It is shown that, with properly designed information exchange schemes, collaborative decoding can achieve close-to-full receive diversity for binary coded and higher order coded modulation systems. Since a distributed array is composed of a cluster of nodes in a wireless network, the number of nodes in the array is usually greater than two. This fact highlights the scalability requirement for the proposed collaborative decoding techniques. Thus, it is necessary to consider general distributed arrays with more than two nodes. To employ collaborative decoding to achieve spatial diversity efficiently in this scenario, the information exchange scheme should be a major concern. It will be shown that with properly designed information exchange schemes, collaborative decoding, compared with MRC, still exhibits the advantage of achieving spatial diversity with significant savings in the information exchange cost.

In this chapter, we study collaborative decoding schemes for distributed arrays with more than two nodes. The discussion will be restricted to systems using BPSK modulation and convolutional codes. We first consider the system model of distributed arrays with more than two nodes in Section 4.1 and extend the two-node collaborative decoding technique developed previously to this case. In Section 4.2 we study the statistical characteristics of the extrinsic information generated by the MAP decoder for the convolutional code. Then a Gaussian approximation for the extrinsic information is introduced. Based on this approximation, a least-reliable-bit and most-reliable-bit information exchange schemes are described. Then, we present
and compare the performance of the two information exchange schemes with different parameter settings in Section 4.3. Finally, we draw a summary in Section 4.4. This chapter is primarily based on the work in [24] and [25].

4.1 System Model for Distributed Array with Two or More Nodes

The general model of a distributed array with more than two nodes is shown in Fig. 4–1. A remote source node transmits message through a single-input/multiple-output forward channel to the destination that contains $M$ ($M \geq 2$) physically separated receiving nodes, denoted by a node set $\mathcal{M} = \{1, 2, \cdots, M\}$. The source encodes and transmits the message with a convolutional code and BPSK modulation. Analogous to the methods discussed Chapters 2 and 3, with proper modifications, the collaborative decoding techniques can be extended to the case of different codes and high-order modulations. Thus, without loss of generality, we will restrict our study here to convolutional codes and BPSK modulation only. In order to be able to apply iterative decoding, each node in $\mathcal{M}$ uses an approximated version of MAP decoding, known as the max-log-MAP decoding algorithm [11], to process the received symbols. All nodes can perform the demodulation and decoding process individually.

We use a memoryless independent fading channel model, that includes the additive white Gaussian noise (AWGN) channel as a special case, to describe the transmission environment between the source and receiving nodes. The received signal
$y_{k,i}$ at the $k$th receiving nodes corresponding to the transmitted BPSK signal $x_i$ (i.e., $x_i \in \{+1,-1\}$) at time instant $i$ can be expressed as

$$y_{k,i} = g_{k,i}x_i + n_{k,i}, \text{ for } k = 1, 2, \ldots, M$$  \hspace{1cm} (4-1)

where $n_{k,i}$ for $k = 1, 2, \ldots, M$ and all $i$ are i.i.d. zero-mean additive Gaussian random variables with variance $E[|n_{k,i}|^2] = \sigma_n^2$, and $g_{k,i}$ is the channel fading gain. For AWGN channels $g_{k,i} = 1$, and for Rayleigh fading channels $g_{k,i}$ for $k = 1, 2, \ldots, M$ and all $i$ are i.i.d. Rayleigh random variables with pdf of

$$p(g_{k,i}) = 2g_{k,i}e^{-g_{k,i}^2}, \quad g_{k,i} \geq 0.$$  \hspace{1cm} (4-2)

We normalize the signal energy $E[|x_i|^2] = 1$. Thus, the average SNR is $1/\sigma_n^2$. In this channel model, we assume that perfect channel state information (CSI) is available at the $k$th receiving nodes and hence coherent detection is performed at each node. With this model the pdf $p(y_{k,i}|x_i, g_{k,i})$, for $k = 1, 2, \ldots, M$ and all $i$, with perfect CSI is given by

$$p(y_{k,i}|x_i, g_{k,i}) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left(-\frac{|y_{k,i} - g_{k,i}x_i|^2}{2\sigma_n^2}\right).$$  \hspace{1cm} (4-3)

At the receiving end, the cluster of nodes form a local network such that they can communicate with one another on an error-free broadcast channel. Broadcast channel is one of the simplest channel models for wireless networks. By using this model, simple wireless LAN protocols such as token ring [26] can be adopted to carry out the communication among the nodes, Thus, the necessity of complicated channel allocation or medium access control protocols can be avoided. The detailed network protocol for collaborative decoding with distributed array is out of the scope of this research work. Here, we simply assume that the error-free communication among the nodes through the broadcast channel has been guaranteed by a certain network protocol. The assumption of error-free communication is also reasonable because the cluster nodes are in close proximity compared with the distance between the
source and the array. Hence the SNR of the broadcast channel among the receiving nodes is significantly higher than that of the forward channel from the source to the receiving nodes. Even if the broadcast channel is noisy in realistic situations, simple error detection or correction coding such as cyclic redundancy check codes [14] can be employed to provide reliable transmission effectively with introducing only a minimum redundancy. In this case, the slight additional process complexity and redundancy due to the coding protection can be ignored. The performance of the forward channel is the main concern here.

4.2 Collaborative Decoding and Information Exchange Schemes

To illustrate collaborative decoding for distributed array with more than two nodes, we consider the turbo decoder with more than two decoder components. It is clear that in the turbo decoding procedure, a decoder component should use the sum of the extrinsic information generated by all other decoding components, except itself, in previous iteration as its a priori information for the current iteration. According to this principle, collaborative decoding is proceeded as follows. Each node collects soft-output (including the extrinsic information and received signal) from all other nodes generated in the previous iteration, then uses the sum of the collected information, called additional information, as its a priori information for the current decoding iteration. To reduce the information exchange amount, the nodes only exchange the soft-output about a portion of information bits in each iteration.

We note that soft-output is used in collaborative decoding while the extrinsic information is used in turbo decoding. The reason for using the soft-output rather than the extrinsic information is as follows. In turbo decoding the received signal for the data bits is the same for all code components. Only the extrinsic information part at the output of each decoding component contains new information. However, for collaborative decoding the received signals at different nodes suffer from independent fading and noise. Thus, both the channel observation and extrinsic information parts
at the decoding output of a node contains new information for other nodes. To exploit as much diversity as possible, the soft-output, rather than the extrinsic information, should be exchanged in collaborative decoding. Note that for non-systematic convolutional codes, since only the coded bits are transmitted, no channel observation is available for data bits at the receiver. Hence, the soft-output and the extrinsic information of the decoder are the same for non-systematic convolutional codes when there is no \textit{a priori} information for the decoding process.

4.2.1 Information Exchange with Memory

Generally, the output (extrinsic information or soft-output) of a SISO decoder is significantly correlated to the input (\emph{a priori} information and channel observation), and the output for adjacent bits are correlated with each other. In iterative decoding, due to the exchange of extrinsic information among all the decoding components, the \emph{a priori} information of a decoder collected from other decoding components will become more and more correlated to its own output in the previous iterations. This means that the decoder can obtain less and less new information from the other decoders. This can severely limit the performance of iterative decoding. In turbo codes, in order to solve this problem, bit interleavers are applied to each code components. The addition of the interleavers is to randomly permute the bits so that the correlation among adjacent bits is broken. Hence, the correlation between the \emph{a priori} information and previous output of a decoder can be reduced.

Due to the iterative soft-output exchange process, correlation problem similar to the one mentioned above can arise in collaborative decoding. Unfortunately, due to the fact that a single message is broadcast from the source to the distributed array, the code structure and bit-sequence order at all nodes must be the same. This means that the interleaving technique in turbo coding can not be applied to attack the correlation problem in this case. On the other hand, another difference between turbo decoding and collaborative decoding is that soft-output, rather than extrinsic
information is exchanged in collaborative decoding. Due to these two facts, the exchange of information for the same bit in successive iterations will cause correlation.

To illustrate the situation, suppose that a data bit $x_i$ in a packet transmitted from the remote source. In the $j$th iteration, for simplicity we assume that only a node $k \in \mathcal{M}$ broadcasts the soft-output of $x_i$, denoted as $\delta_{k,i}^{(j)}$, to other nodes in $\mathcal{M}$. Then in the $(j + 1)$th iteration, the other $M - 1$ nodes will use $\delta_{k,i}^{(j)}$ as a priori information to perform decoding. For convenience, we denote the set of other $M - 1$ nodes as $\mathcal{M}^c = \{m : m \in \mathcal{M}, m \neq k\}$. According to MAP decoding, the soft-output for bit $x_i$ at node $m \in \mathcal{M}^c$ can be expressed as

$$
\delta_{m,i}^{(j+1)} = \delta_{k,i}^{(j)} + \xi_{m,i}^{(j+1)},
$$

where $\xi_{m,i}^{(j+1)}$ is the extrinsic information generated for bit $x_i$ by node $m$ in the $(j + 1)$th iteration. If there are a subset of nodes, denoted as $\mathcal{M}'$, in $\mathcal{M}^c$ continuing to exchange the soft-output of bit $x_i$, then node $k$ will obtain all this information and use it as the a priori information for the next iteration. In this case, the a priori information for bit $x_i$ at node $k$, denoted as $\eta_{k,i}^{(j+2)}$, is given by

$$
\eta_{k,i}^{(j+2)} = \sum_{m \in \mathcal{M}'} \delta_{m,i}^{(j+1)} = |\mathcal{M}'| \delta_{k,i}^{(j)} + \sum_{m \in \mathcal{M}'} \xi_{m,i}^{(j+1)},
$$

(4–4)

where $|\mathcal{M}'|$ the cardinal size of $\mathcal{M}'$. In (4–4), we can see that the soft-output $\delta_{k,i}^{(j)}$ is explicitly included in $\eta_{k,i}^{(j+2)}$. This implies the existence of significant correlation between the a priori information and the previous soft-output for the decoder at node $k$.

Based on the above discussion, we adopt a simple scheme to solve the correlation problem. The method is to assign a memory to each information bit to record whether the soft-output for the bit has been exchanged or not. Once the soft-output of a bit has been exchanged, no further information about that bit will be exchanged in later iterations. Since the bits which obtain information from other nodes are very likely to
have high reliability values (magnitudes of soft-output) after one iteration, repeating
the information exchange for these bit can not improve the decoding performance.
In some case, this may even hurt the performance because some decoding errors in
these bits may have high reliability values. Also, this scheme helps to increase the op-
portunity for other less reliable bits to receive information in the following iterations.
Besides attacking the correlation problem, another advantage of this memory-based
scheme is that the information exchange amount can be significantly reduced.

4.2.2 Least-Reliable-Bit Information Exchange

In order to achieve spatial diversity without the need of extensive information
exchange as in MRC, only a small amount of information can be exchanged in each
iteration for collaborative decoding. This means that proper information must be
chosen and exchanged so that the decoder at each node can improve the error perform-
ance effectively. In this sense, the selection of information becomes very important.

Although only account for a small portion of the whole packet, the bits in error
directly determine the error performance of a decoder. In MAP decoding, a priori
information of a bit can directly contribute in its soft-output. Thus, we consider a
method to collect a priori information for those bits that are likely to be in the error
in previous decoding iteration at each node.

As mentioned in Chapter 2, the soft-output of the MAP decoder in log-likelihood
ratio form provides a good reliability measure for a data bit. It is directly related
to the a posteriori probability of the decision for an data bit. For both AWGN and
Rayleigh fading channels, a data bit with a small soft-output magnitude is more likely
to be in error. In fact, many decoding algorithms, such as the MAP decoding and
belief-propagation decoding algorithm, used in turbo-like decoders are the min-sum
or min-product algorithms [27]. It turns out that the soft-output of these decoding
algorithms in LLR form possess Gaussian-like statistical properties [28, 29]. Fig. 4–
2 shows the typical probability distribution of the soft-output generated by MAP
Figure 4–2: Typical probability distribution functions of soft-output for convolutional codes

decoder for convolutional codes on AWGN channels. In the figure, we assume that data bits to be decoded are all zeros for clarity. With this assumption, if soft-output of a bit is negative, then the decision on the bit will be in error. From the figure, we can see that the Gaussian-like probability distribution function falls mostly on the right-hand-side of $y$-axis, only a small part of its left tail is negative. Since the tail of a Gaussian distribution function decays exponentially, the probability for the soft-output of erroneous bits to have large reliability values is very small. Conversely, the reliability values for correct bits may have a good chance to be large. Thus, a simple way to identify the possible erroneous bits is to measure the reliability values.

With the above argument, we propose an efficient information exchange scheme, called least-reliable-bit (LRB) information exchange. The idea of the LRB exchange scheme is that each node requests information from other nodes for its least reliably decoded data bits [24]. All the additional information collected from other nodes is used as a priori information in the next decoding iteration. Using the scheme described in Section 4.2.1, a memory is assigned for each data bit to record whether
the information of that bit has been exchanged or not. Once information of a bit has been exchanged, no further information about that bit will be exchanged in later iterations. In each iteration, the bits for which information has not been previously exchanged are called candidate bits, and the remaining bits are non-candidate bits. We denote the total number of exchanges by $I$, and the fraction of candidate bits to exchange in the $j$th ($0 \leq j \leq I-1$) iteration by $p_j$ ($0 \leq p_j \leq 1$), respectively. The procedure of the LRB exchange scheme is as follows:

1. Set all data bits to be candidate bits.
2. Decode the received signals at each node.
3. If $I + 1$ decoding iterations (i.e., $I$ exchanges) have been performed, then stop the decoding procedure and go to step 1) to process a new packet.
4. Otherwise, each node ranks the candidate bits according to their soft output magnitude (absolute value of the soft output) and requests soft information for the bottom $p_j$ fraction of the candidate bits (the least reliable candidate bits) from other nodes.
5. Each node broadcasts soft output for those bits that are requested by other nodes.
6. Those bits involved (received and broadcast) in the current exchange are set to be non-candidate bits for later iterations.
7. Each node adds the information from other nodes to its a priori information and returns to step 2).

Here, $\{p_j\}_{j=0}^{I-1}$ are the design parameters, which are usually chosen based on the tradeoff between performance and information exchange amount. Optimization for the design of the parameters $\{p_j\}_{j=0}^{I-1}$ is an interesting topic, but is outside the scope of this work. In this chapter, we focus on the capability of collaborative decoding to achieve receive diversity given the node number $M$ and some proper choices of $\{p_j\}_{j=0}^{I-1}$. 
Since the information being exchanged is the soft-output (in LLR form) for a portion of the data bits, we use the average total number of LLRs transmitted through the broadcast channel in the distributed array for processing each packet as a simple measure of the amount of exchanged information. The cost of overhead due to the protocol is ignored here. If we use $\Theta$ to denote the average information exchange amount, then with the assumption that soft-output at different nodes and for different bits are independent, for a specific choice of $\{p_j\}_{j=0}^{I-1}$ the value of $\Theta$ for the LRB scheme is given by

$$\Theta_{\text{LRB}} = MN \sum_{j=0}^{I-1} \left[ 1 - (1 - p_j)^{M-1} \right] \prod_{l=0}^{j-1} (1 - p_l)^M, \quad (4-5)$$

where $N$ is the block size of data bits. Correspondingly, the information exchange amount of MRC is given by

$$\Theta_{\text{MRC}} = MN/R_c, \quad (4-6)$$

where $R_c$ is the code rate.

### 4.2.3 Most-Reliable-Bit Information Exchange

In the LRB exchange scheme, each node directly requests information from other nodes for its low reliable bits. Thus, the information exchange process has to be carried out in two stages. In the first stage, each node sends out its request information. In the second stage, each node broadcasts its soft-output according to the request received from other nodes. This two stage process may increase the complexity of the required network protocol. In addition, a node requesting information for its least reliable bits does not necessarily mean that other nodes can provide more reliable information for those bits. It is possible that the information collected by a node is not reliable enough to improve its decoding performance in the next iteration.

As an alternative, we propose another scheme called most-reliable-bit (MRB) information exchange. MRB exchange is usually efficient for the distributed array consisting of a large number of nodes, e.g., $M > 6$. The idea of this MRB scheme is
that each node directly broadcasts the soft-output for its most reliable bits after a decoding iteration performed without the information request stage in the LRB scheme. Similar to LRB, the identification of highly reliable bits is based on the ranking and comparison of the reliability values. A small percentage of bits with high reliability values are chosen as the most reliable bits. According to the statistic characteristics of the soft-output shown in Fig. 4–2, these bits are the correctly decoded bits with very high probability. Assume that the soft output for different bits and/or at different nodes are independent of each other, then the most reliable bits are evenly spread in a packet at each node and the positions are uncorrelated for different nodes. Thus, when the number of nodes is large, even if each node only broadcast information of its top 10% reliable bits, the total information collected from all these nodes can cover more than 50% bits in the whole packet. Therefore, the less reliable bits at a node that collects this information can have a good chance to be covered. The collected additional information in MRB is usually much more reliable than that in LRB.

Combined with the memory-based scheme to avoid correlation among the additional information at different iterations, the MRB exchange scheme is given as follows.

1. Clear the flag registers, i.e., set all data bits to be candidate bits.
2. Decode the received signals at each node.
3. If \( I \) exchanges (i.e., \( I + 1 \) decoding iterations) have been performed, then terminate the decoding procedure and go to step 1) to process a new packet.
4. Otherwise each node ranks the candidate bits according to their soft output magnitude (reliability), and broadcasts the soft information for the top \( p_j \) fraction of the candidate bits (the most reliable candidate bits) to other nodes.
5. Each node sets the flags for those bits involved (received and broadcast) in the current exchange so that they become non-candidate bits for later iterations.
6. Each node adds the additional information to its *a priori* information and returns to step 2).

Ignoring the cost of information exchange due to the network protocol and bit indexes, The information exchange amount for MRB can be computed as

$$
\Theta_{MRB} = MN \sum_{j=0}^{I-1} p_j \prod_{l=0}^{j-1} (1 - p_l)^M. \tag{4–7}
$$

### 4.3 Performance Evaluation

In this section, we use Monte Carlo simulations to evaluate the performance of collaborative decoding with the LRB and MRB information exchange schemes when the distributed arrays consist more than two nodes \((M > 2)\). In the simulations, the packet size is set to 1024 data bits. We set the number of iterations \(I\) to 3 (i.e., 3 exchanges and 4 decoding iterations are performed in total).

We first examine the performance of collaborative decoding with the LRB exchange scheme. Fig. 4–3 shows the BER curves of collaborative decoding on the AWGN channel for the cases of \(M = 2, 3, 4, 6\) and 8, respectively. In the system, we employ the non-recursive convolutional code \(CC(5, 7)\), for which the generation polynomial is \([1 + D^2, 1 + D + D^2]\) and code rate \(R_c = 1/2\). The parameter \(\{p_j\}\) are set to \(\{0.1, 0.15, 0.25\}\). For clarity, only the BERs obtained after the last iteration in collaborative decoding are shown. In the figure we also show the BERs for MRC with \(M = 2, 6\) and 8. From the figure we can see that, for \(M = 2\), the performance of collaborative decoding with the LRB exchange scheme is very close that of MRC, while it is about 2dB and 2.6dB within that of MRC for \(M = 6\) and \(M = 8\), respectively. Fig. 4–4 shows the BER performance on an independent Rayleigh fading channel. Similar to the case of AWGN channel, significant spatial diversity gains are obtained using collaborative decoding.

For collaborative decoding with the MRB exchange scheme, in order to gain spatial diversity, we set \(\{p_j\}\) to be \(\{0.1, 0.2, 1\}\). All other system settings are the
Figure 4–3: BER performance of collaborative decoding with LRB exchange for the cases of $M = 2, 3, 4, 6$ and 8 on AWGN channels, where $CC(5, 7)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used.

Figure 4–4: BER performance of collaborative decoding with LRB exchange on Rayleigh fading channels, parameter settings are the same as Fig. 4–3.
same as that for the LRB scheme. The BER performance for MRB on AWGN and independent Rayleigh fading channels are shown in Fig. 4–5 and 4–5, respectively. Similar to collaborative decoding with LRB exchange, collaborative decoding with MRB exchange can also achieve significant receive diversity gains and performance close to that of MRC for both AWGN and Rayleigh fading channels.

Finally, we compare the information exchange amount of LRB exchange and MRB exchange schemes. For the settings \(\{p_j\} = \{0.1, 0.15, 0.25\}\) for LRB and \(\{p_j\} = \{0.1, 0.2, 1\}\) for MRB, the two schemes roughly achieve the same performance for different numbers of nodes, \(M\). We use (4–5), (4–7) and (4–6) to calculate the information exchange amount with respect to MRC for the two schemes. Fig. 4–7 shows the relative information exchange amount \(\Theta_{LRB}/\Theta_{MRC}\) and \(\Theta_{MRB}/\Theta_{MRC}\) for different values of \(M\). From the figure, we can see that for this setting of \(\{p_j\}\), \(\Theta_{LRB}\) grows with the increasing of \(M\) and approaches to half of the information exchange amount for MRC. In contrast, \(\Theta_{MRB}\) decreases with \(M\). From the comparison we
Figure 4–6: BER performance of collaborative decoding with MRB exchange on Rayleigh fading channels, parameter settings are the same as Fig. 4–5.

Figure 4–7: Comparison of information exchange amount with respect to MRC for LRB and MRB exchange schemes
conclude that, when the number of nodes is small, LRB exchange scheme is more efficient than MRB. With more nodes in the distributed array, MRB will become more efficient than LRB.

4.4 Summary

In this chapter, we extend collaborative decoding to distributed arrays with two or more nodes. Two different information exchange schemes are proposed. In the LRB scheme, the nodes request soft information for a certain percentage of their least reliable bits. In the MRB exchange scheme, nodes send out soft information about a small set of their most reliable bits. Collaborative decoding with both of these two scheme can achieve most of the spatial diversity and provide significant savings in terms of information exchange amount compared to MRC on AWGN and independent Rayleigh fading channels. We also compare the information exchange amount for the two schemes. It is shown that for distributed arrays with small number of nodes, LRB is efficient. When the number of nodes increases, MRB becomes more efficient than LRB.
CHAPTER 5
PERFORMANCE ANALYSIS FOR COLLABORATIVE DECODING WITH LEAST-RELIABLE-BIT EXCHANGE ON AWGN CHANNELS

As an efficient diversity technique, collaborative decoding with the LRB information exchange scheme, discussed in Chapter 4, provides significant savings on the cost of information exchange while still achieves performance close to that of MRC. In this chapter, we focus on the theoretical analysis of the error performance of collaborative decoding with the LRB exchange scheme on the AWGN channel. The system model considered in this chapter is the same as that described in Section 4.1. Since we restrict the analysis to the case of AWGN channel, the channel gain $g_{k,i}$ in (4–1) will be always 1 in this chapter. The analysis is based on the LRB exchange scheme described in Section 4.2.2.

The analysis will be based on the statistical characteristics of soft information obtained from the MAP decoders in collaborative decoding. From simulation, we observe that the extrinsic information generated in the decoding process can be well approximated by Gaussian random variables when nonrecursive convolutional codes are employed. Unfortunately, for recursive convolutional codes the extrinsic information generated in the decoding process can not be approximated by a simple Gaussian distribution, which makes the performance analysis difficult. Due to this difficulty, we only consider nonrecursive convolutional codes in this chapter.

By viewing collaborative decoding as an iterative decoding system, we use a typical analysis technique for turbo-like codes, known as density evolution, to analyze the performance of collaborative decoding. As in most of the literatures (e.g., [30] and [31]) on analysis of turbo codes, we use simulation to obtain the statistical
characteristics of the extrinsic information, which is approximated by a Gaussian distribution. To simplify the problem, we model the collaborative decoding process as a density evolution system with only one MAP decoder. Then we can generate the \textit{a priori} information of the density evolution model according to the LRB exchange scheme. By simulating the density evolution model with only one MAP decoder, we obtain the statistical characteristics of the actual extrinsic information with a modest simulation load in comparison to that of the actual collaborative decoding system. With the knowledge of the extrinsic information probability distribution at each iteration, we derive an approximate bit-error rate (BER) upper bound for collaborative decoding with the LRB exchange scheme.

The rest of this chapter is organized as follows. In Section 5.1, we model collaborative decoding as a concatenated structure consisting of a MAP decoder and an information exchanging device, and employ Gaussian approximation to obtain the density evolution of the extrinsic information. In Section 5.2, we derive an upper bound of the BER of the collaborative decoding process. Numerical results obtained from the analysis are shown in Section 5.3. Finally, conclusions are given in Section 5.4. This chapter is based on the work in [25] and [32].

5.1 Gaussian-Approximated Density Evolution For Nonrecursive Convolutional Codes

Due to the exchange of soft information in the process, knowledge of the statistical characteristics of soft information from maximum \textit{a posteriori} (MAP) decoders in collaborative decoding is important to its performance analysis. Thus, we first consider the soft information generated in collaborative decoding. Note that the soft output for non-systematic codes consists only of extrinsic information and \textit{a priori} information, if a candidate bit has not obtained additional information previously, then ranking and exchanging the soft output for candidate bits is equivalent to ranking and exchanging the extrinsic information for those bits. Also, the sets of candidate
bits and non-candidate bits for a packet in each iteration are exactly the same for all nodes. These facts are important to understand the analysis in the following sections.

Because of the symmetry among the nodes in our system model, the statistical characteristics of the extrinsic information at each node is the same in each iteration. This means that the behavior of the LRB exchange process can be determined by knowing the statistical characteristics of the output from the MAP decoder at a single node. Thus, the collaborative decoding process can be modeled by the joint operation of an information exchange unit and the MAP decoder unit as shown in Fig. 5–1. The output of the information exchange is fed back to the MAP decoder as a priori information for use in the next decoding iteration. The following analysis is based on this system model.

Assuming that the all-zero codeword is transmitted, it is well known that the extrinsic information generated by a MAP decoder, in the log-likelihood ratio (LLR) form, is well approximated by Gaussian random variables when the inputs to the decoder are i.i.d. Gaussian [31]. For the collaborative decoding process described in Section 4.1, the additional information obtained from the information exchanging process, which is used as input to the MAP decoder, has a non-Gaussian distribution. Nevertheless, we observe that the probability distribution of the extrinsic information from the MAP decoder in each iteration can still be well approximated as Gaussian

Figure 5–1: System model for collaborative decoding process.
when nonrecursive convolutional codes are employed. Fig. 5–2 shows the typical histograms of the extrinsic information generated by MAP decoders at successive iterations in the collaborative decoding process for nonrecursive convolutional codes. Comparing to the corresponding ideal Gaussian distributions, we can see that the histograms are very close to Gaussian distributions. Based on these observations, we apply the Gaussian-approximated density evolution technique in [30, 31] to predict the behavior of the MAP decoders in collaborative decoding.

As in [31], we assume that at each node the extrinsic information generated by the MAP decoder for all the information bits at that node are i.i.d. Gaussian random variables in each iteration. We further assume that the extrinsic information for information bits generated by different nodes are independent. Thus, the statistical behavior of the extrinsic information is sufficiently specified by its mean and variance. Unfortunately, obtaining an analytic distribution for the extrinsic information generated by MAP decoders is an intractable problem, especially for the case of
non-Gaussian input. Hence, we use simulation, based on the model in Fig. 5–1, to quantiﬁy the evolution of the probability distribution. By inputting actual additional information to the MAP decoder, the mean and variance of the extrinsic information can be obtained with modest simulation complexity in comparison to the actual collaborative decoding process. This knowledge of extrinsic information is used to evaluate the error performance in Section 5.2.

We ﬁrst describe the generation of the additional information. For the $j$th decoding iteration, let $\xi_{k,i}^{(j)}$ denote the extrinsic information generated by the MAP decoder for the $i$th information bit at node $k$, and let $B_i^{(j)}$ denote the event that bit $i$ is a candidate bit. Under the Gaussian assumption, $\xi_{k,i}^{(j)} \sim \mathcal{N}(\mu_j, \sigma_j^2)$, and $\{\xi_{k,i}^{(j)}\}$ are i.i.d. for all $k$ and $i$, where $\mathcal{N}(\mu, \sigma^2)$ means Gaussian distributed with mean $\mu$ and variance $\sigma^2$. When the bit block size is large enough, the information request criterion for $|\xi_{k,i}^{(j)}|$ to rank in the bottom $p_j$ fraction among the candidate bits in the $k$th node is approximately equivalent to $|\xi_{k,i}^{(j)}| \leq T_j$, where $T_j \geq 0$ is a threshold related to the distribution of $\xi_{k,i}^{(j)}$ and $p_j$. Speciﬁcally, we have

$$P(|\xi_{k,i}^{(j)}| \leq T_j | B_i^{(j)}) = p_j. \quad (5–1)$$

Let $\lambda_{k,i}^{(j)}$ denote the additional information for the $i$th bit at the $k$th node generated by the LRB exchange process in the $j$th iteration. This additional information will be added to the a priori information in the $(j+1)$th iteration by node $k$. Below, let us assume that $M \geq 3$. The case of $M = 2$ will be discussed separately later. According to the LRB scheme, if bit $i$ is a non-candidate bit in the $j$th iteration, then $\lambda_{k,i}^{(j)} = 0$. Otherwise, there are three possibilities for the case of a candidate bit:

i) No node requests information for the $i$th bit, i.e., $\bigcap_{t \in \mathcal{M}} |\xi_{t,i}^{(j)}| > T_j$, then $\lambda_{k,i}^{(j)} = 0$;
ii) The $k$th node does not request information for bit $i$, but there is one other node requesting information for that bit. We denote this event by $\hat{R}_{k,i}^{(j)}$, i.e.,

$$\hat{R}_{k,i}^{(j)} = \bigcup_{r \in M, r \neq k} \{ |\xi_{r,i}^{(j)}| \leq T_j, \bigcap_{t \in M, t \neq r} |\xi_{t,i}^{(j)}| > T_j \}. \quad (5-2)$$

Then the $k$th node will obtain information from other nodes except the one sending out request, i.e.,

$$\lambda_{k,i}^{(j)} = \sum_{t \in M, t \neq r, t \neq k} \xi_{t,i}^{(j)} \triangleq \check{\lambda}_{k,i}^{(j)}, \quad (5-3)$$

iii) The $k$th node or more than one nodes in $M$ request information for bit $i$. We denote this event by $\check{R}_{k,i}^{(j)}$, i.e.,

$$\check{R}_{k,i}^{(j)} = \{ |\xi_{k,i}^{(j)}| \leq T_j \} \cup \left\{ \{ |\xi_{k,i}^{(j)}| > T_j \} \cap \bigcup_{r \in M, r \neq k} \{ |\xi_{r,i}^{(j)}| \leq T_j, \bigcap_{t \in M, t \neq r, t \neq k} |\xi_{t,i}^{(j)}| > T_j \} \right\}. \quad (5-4)$$

In this case, the $k$th node will obtain information from all other nodes, and $\lambda_{k,i}^{(j)}$ is given by

$$\lambda_{k,i}^{(j)} = \sum_{t \in M, t \neq k} \xi_{t,i}^{(j)} \triangleq \check{\lambda}_{k,i}^{(j)}. \quad (5-5)$$

Under the Gaussian assumption, we can see that, without the constraint of candidate bits, $\check{\lambda}_{k,i}^{(j)} \sim \mathcal{N}((M-2)\mu_j, (M-2)\sigma^2_j)$ while $\check{\lambda}_{k,i}^{(j)} \sim \mathcal{N}((M-1)\mu_j, (M-1)\sigma^2_j)$. Clearly, $\hat{R}_{k,i}^{(j)}$ and $\check{R}_{k,i}^{(j)}$ are disjoint events. According to the LRB scheme, only under case i) bit $i$ will be a candidate bit again in next iteration. Hence,

$$P(B_{i}^{(j+1)}|B_{i}^{(j)}) = P\left( \bigcap_{k \in M} \{ |\xi_{k,i}^{(j)}| > T_j \} | B_{i}^{(j)} \right) = (1 - p_j)^M. \quad (5-6)$$

From this recursive relation, we immediately obtain

$$P(B_{i}^{(j)}) = P(B_{i}^{(j)}, B_{i}^{(j-1)}, \cdots, B_{i}^{(0)}) = \prod_{l=0}^{j-1} P(B_{i}^{(l+1)}|B_{i}^{(l)}) = \prod_{l=0}^{j-1} (1 - p_l)^M, \quad (5-7)$$
and

$$B_t^{(j)} = \bigcap_{l=0}^{j-1} \bigcap_{k \in M} \{ |\xi_{k,i}^{(l)}| > T_l \}. \quad (5-8)$$

Different from case i), in cases ii) and iii) bit $i$ will become a non-candidate bit in the next iteration. Thus,

$$P(B_t^{(j+1)} | B_t^{(j)}) = P(\hat{R}_{k,i}^{(j)} \cup \tilde{R}_{k,i}^{(j)} | B_t^{(j)}) = P(\hat{R}_{k,i}^{(j)} | B_t^{(j)}) + P(\tilde{R}_{k,i}^{(j)} | B_t^{(j)}). \quad (5-9)$$

With the above arguments, we can easily simulate the additional information generated in the actual LRB process for the density evolution model in Fig. 5–1. Without loss of generality, we assume that the MAP decoder in Fig. 5–1 is in the $M$th node. Also, we assume that the block length of the code is long enough to ensure the Gaussian approximations and thresholding. In the $j$th iteration, the MAP decoder generates $\xi_{M,i}^{(j)}$ for the $i$th bit. Then the values of $\mu_j$ and $\sigma_j^2$ of the extrinsic information for the information bits are estimated. To find $T_j$ using (5–1), we first use nonparametric estimation method to estimate the cumulative distribution function $F_j(x)$ of the extrinsic information for the candidate bits, i.e.,

$$F_j(x) = P(\xi_{k,i}^{(j)} < x | B_t^{(j)}). \quad (5–10)$$

Then, according to (5–1) we have

$$p_j = F_j(T_j) - F_j(-T_j). \quad (5–11)$$

For $p_j < 1$, by solving (5–11) numerically we can obtain $T_j$ approximately. For the case of $p_j = 1$, we set $T_j = \infty$. In the information exchange module, we use $(M - 1)$ i.i.d. random variables following distribution $F_j(x)$ to simulate the extrinsic information $\xi_{k,i}^{(j)}$ for candidate bit $i$ at node $k$, for $k = 1, 2, \cdots, M - 1$, respectively. Then, with $\{\xi_{k,i}^{(j)}\}$, for all $k \in \mathcal{M}$ we check if $\hat{R}_{M,i}^{(j)}$ or $\tilde{R}_{M,i}^{(j)}$ occurs, set $\lambda_{M,i}^{(j)}$ to $\hat{\lambda}_{M,i}^{(j)}$ or $\tilde{\lambda}_{M,i}^{(j)}$ accordingly, and flag this bit as a non-candidate bit for the next iteration. For all other cases, we set $\lambda_{M,i}^{(j)} = 0$. Then based on the LRB scheme we construct the a
Figure 5–3: Comparison of mean and variance of the extrinsic information from the density evolution model and that from the actual collaborative decoding process

priori information, denoted by $\eta^{(j+1)}_{M,i}$, for the $(j+1)$th iteration, as

$$
\eta^{(j+1)}_{M,i} = \sum_{l=0}^{j} \lambda^{(l)}_{M,i} = \begin{cases} 
\lambda^{(l)}_{M,i} & \text{if } \mathcal{R}^{(l)}_{M,i} \neq \emptyset, \forall 0 \leq l \leq j \\
0 & \text{otherwise,}
\end{cases} \quad (5–12)
$$

where

$$
\mathcal{R}^{(l)}_{M,i} \triangleq \mathcal{R}^{(l)}_{M,i} \cup \mathcal{R}^{(l)}_{M,i}. \quad (5–13)
$$

By inputting this a priori information to the MAP decoder and iterating the above procedure, we can obtain the statistical characteristic of the Gaussian-approximated extrinsic information in the whole collaborative decoding process. Fig. 5.1 and 5.1 show the comparison of the mean and variance of the extrinsic information and the threshold $T_j$ estimated in our density evolution model and the actual collaborative decoding process for the case of $M = 6$. In the figure, the maximum free distance 4-state non-recursive convolution code is used, and $\{p_j\}$ is set to $\{0.1, 0.2, 1\}$. From the figure, we see that our density evolution model gives an excellent approximation for the actual collaborative decoding process with only $1/M$th of the simulation load.
Figure 5–4: Comparison of threshold estimated from the density evolution model that from the actual collaborative decoding process

The analysis in the next section also shows that to evaluate the error performance for the total $I$ iterations, we only need the statistical knowledge of the extrinsic information in the first $I - 1$ iterations.

5.2 Error Performance Analysis

With knowledge of the statistical characteristics of the extrinsic information, we evaluate the error performance of collaborative decoding with LRB exchange. We again consider the decoding process and performance at the $M$th node. Let $\mathcal{M}' = \{1, 2, \cdots, M - 1\}$ denote the set of the other $M - 1$ nodes. Since the average BER is considered, we drop the bit index, i.e., the subscript $i$, in the notation of variables and events for the bit of interest. For convenience, we also drop the subscript $M$ for the $M$th node in following derivation. From the definition (5–3), we know that $\dot{\lambda}_{M,i}^{(j)}$ is a Gaussian random variable for $M \geq 3$ but equals zero for $M = 2$. Thus we treat $M = 2$ as a special case and consider the case of $M \geq 3$ first below.
5.2.1 BER Upper Bound for $M \geq 3$

For the case of $M \geq 3$, the BER of the MAP decoders in the $j$th ($j > 0$) iteration is the probability that the soft output of a bit is smaller than zero given that the all-zero sequence is transmitted, i.e.,

$$P_b^{(j)} = P(\xi^{(j)} + \eta^{(j)} < 0), \quad (5–14)$$

where $\xi^{(j)}$ is the extrinsic information, and $\eta^{(j)}$ is the a priori information in the $j$th iteration given in (5–12) at the $M$th node, respectively. Here, we evaluate the error performance by finding an upper bound for (5–14).

According to (5–12), (6–10) and (6–16), we rewrite (5–14) as

$$P_b^{(j)} = P(\xi^{(j)} + \eta^{(j)} < 0, \mathcal{B}^{(j)}) + P(\xi^{(j)} < 0, \mathcal{B}^{(j)})$$

$$= \sum_{l=0}^{j-1} P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} < 0, \mathcal{B}^{(j)}). \quad (5–15)$$

We first consider the first part in (5–15). Using (6–16), (5–3) and (5–5), we have

$$P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)})$$

$$= P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0, \ddot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)})$$

$$\leq P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0). \quad (5–16)$$
With (5–2) and (5–8), when \( p_t < 1 \) (i.e., \( T_t < \infty \)) for \( 0 \leq t \leq l \), we upper bound the first term in (5–16) as follows

\[
P(\xi^{(j)} + \hat{\lambda}^{(l)} < 0, \hat{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) = P\left(\xi^{(j)} + \hat{\lambda}^{(l)} < 0, \bigcup_{r \in \mathcal{M}'} \left\{ |\xi_{t,r}^{(l)}| \leq T_t, \bigcap_{t \in \mathcal{M}, t \neq r} |\xi_{t,t}^{(l)}| > T_t \right\}, \mathcal{B}^{(l)} \right)
\]

\[
= \sum_{r=1}^{M-1} P\left(\xi^{(j)} + \hat{\lambda}^{(l)} < 0, |\xi_{r,r}^{(l)}| \leq T_t, \bigcap_{t \in \mathcal{M}, t \neq r} |\xi_{t,t}^{(l)}| > T_t \bigcap_{t=0}^{l-1} \bigcap_{k \in \mathcal{M}} |\xi_{k,k}^{(l)}| > T_t \right)
\]

\[
\leq (a) \sum_{r=1}^{M-1} P\left(\xi^{(j)} + \hat{\lambda}^{(l)} < 0, |\xi_{r,r}^{(l)}| \leq T_t, \bigcap_{t=0}^{l-1} |\xi_{1,t}^{(l)}| > T_t \right)
\]

\[
= (b) (M - 1) P(\xi^{(j)} + \hat{\lambda}^{(l)} < 0) P\left(|\xi_{1,1}^{(l)}| \leq T_l, \bigcap_{t=0}^{l-1} |\xi_{1,t}^{(l)}| > T_l \right),\tag{5–17}
\]

where (a) is obtained by dropping all the events in \( \{ \hat{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)} \} \) associated with \( \xi_{k,k}^{(l)} \) for all \( t \) and \( k \in \mathcal{M}, k \neq r \), and (b) is due to the fact that the probabilities in the sum in (5–17) are equal for \( 1 \leq r \leq M - 1 \), and that \( \xi^{(j)} \) and \( \hat{\lambda}^{(l)} \) are independent of \( \xi_{r,r}^{(l)} \) for all \( t \).

To evaluate the probability \( P\left(|\xi_{1,1}^{(l)}| \leq T_l, \bigcap_{t=0}^{l-1} |\xi_{1,t}^{(l)}| > T_l \right) \) in (5–18), we use (5–8) to rewrite \( P(\mathcal{B}^{(j)}) \) as

\[
P(\mathcal{B}^{(j)}) = P\left(\bigcap_{k=1}^{M} \left\{ \bigcap_{l=0}^{j-1} |\xi_{k,k}^{(l)}| > T_l \right\} \right) = \prod_{k=1}^{M} P\left(\bigcap_{l=0}^{j-1} |\xi_{k,k}^{(l)}| > T_l \right)
\]

\[
= \left[ P\left(\bigcap_{l=0}^{j-1} |\xi_{k,k}^{(l)}| > T_l \right) \right]^{M}, \quad k \in \mathcal{M}.\tag{5–19}
\]

By comparing (5–19) with (5–7), for all \( k \in \mathcal{M} \) we have

\[
P\left(\bigcap_{l=0}^{j-1} |\xi_{k,k}^{(l)}| > T_l \right) = \prod_{l=0}^{j-1} (1 - p_l).\tag{5–20}
\]

In the similar manner, it is easy to see that for all \( k \in \mathcal{M} \)

\[
P\left(|\xi_{k,k}^{(l)}| \leq T_l, \bigcap_{t=0}^{l-1} |\xi_{k,k}^{(l)}| > T_l \right) = p_l \prod_{t=0}^{l-1} (1 - p_t).\tag{5–21}
\]
Thus, with (5–21) and taking into account the fact $P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \hat{R}^{(l)}, B^{(l)}) \leq P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0)$, we refine the upper bound (5–18) as

$$P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \hat{R}^{(l)}, B^{(l)}) \leq \min \left\{ 1, (M-1) \prod_{t=0}^{l-1} (1-p_t) \right\} \cdot P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0). \quad (5–22)$$

This bound is for the case that all $p_t$ are not equal to 1. If there exists a $0 \leq t \leq l$ such that $p_t = 1$, then $P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \hat{R}^{(l)}, B^{(l)}) = 0$ because of $P(\hat{R}^{(l)}, B^{(l)}) = 0$.

To include this case, we rewrite the upper bound (5–22) as

$$P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \hat{R}^{(l)}, B^{(l)}) \leq a_t \cdot P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0), \quad (5–23)$$

where

$$a_t = \begin{cases} 0 & \text{if } \prod_{t=0}^{l} (1-p_t) = 0 \\ \min \left\{ 1, (M-1) p_t \prod_{t=0}^{l-1} (1-p_t) \right\} & \text{otherwise}. \quad (5–24) \end{cases}$$

In the same way, we consider the probability $P(\xi^{(j)} < 0, B^{(j)})$ in (5–15). With (5–7) and (5–20), this probability can be easily expanded and upper bounded by

$$P(\xi^{(j)} < 0, B^{(j)}) = P\left(\xi^{(j)} < 0, \bigcap_{t=0}^{j-1} \bigcap_{k \in M} |\xi^{(l)}_k| > T_j \right)$$

$$= P\left(\xi^{(j)} < 0, \bigcap_{t=0}^{j-1} |\xi^{(l)}| > T_j \right) P\left(\bigcap_{t=0}^{j-1} \bigcap_{k=1}^{M-1} |\xi^{(l)}_k| > T_j \right)$$

$$\leq P(\xi^{(j)} < 0, |\xi^{(j-1)}| > T_{j-1}) \cdot \prod_{t=0}^{j-1} (1-p_t)^{M-1}$$

$$\leq b_j \left[ P(\xi^{(j-1)} < -T_{j-1}) + P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1}) \right], \quad (5–25)$$

where

$$b_j = \prod_{t=0}^{j-1} (1-p_t)^{M-1}. \quad (5–26)$$
By inserting (5–23), (5–25) and (5–16) into (5–15), we obtain the following upper bound

\[ P_b^{(j)} \leq \sum_{i=0}^{j-1} \left[ a_l P(\xi^{(j)} + \lambda^{(i)} < 0) + P(\xi^{(j)} + \lambda^{(i)} > 0) \right] + b_j \left[ P(\xi^{(j-1)} < -T_{j-1}) + P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1}) \right], \]

where \( a_l \) and \( b_j \) are given by (5–24) and (5–26), respectively. Below, we employ a union bound for the max-log-MAP decoder to further upper bound the probabilities in (5–27).

### 5.2.2 Union Bound for Max-log-MAP Decoding

Let \( \mathbf{u} \) and \( \mathbf{c} \) denote a information bit sequence and the corresponding codeword generated by a nonrecursive convolutional code \( \mathcal{C}: \mathbf{u} \rightarrow \mathbf{c} \), where \( \mathbf{u} = (u_0, u_1, \cdots, u_i, \cdots) \), \( \mathbf{c} = (c_0, c_1, \cdots, c_i, \cdots) \), and \( u_i \) and \( c_i \in \{0, 1\} \) are the information bit and coded bit, respectively. Correspondingly, \( y_i \) is the received BPSK (i.e., \( x_i = 1 - 2c_i \) in (4–1)) signal at the decoder. Under the assumption that the all-zero sequence is transmitted, the extrinsic information generated by the max-log-MAP decoder in the LLR form is given by

\[ \xi_k^{(j)} = \max_{\{\mathbf{u}, \mathbf{c}\} \in C^+} \left\{ -\Gamma_{\mathbf{u}, \mathbf{c}}^{(j)} \right\} + \min_{\{\mathbf{u}, \mathbf{c}\} \in C^-} \left\{ \Gamma_{\mathbf{u}, \mathbf{c}}^{(j)} \right\}, \]

where \( C^+ \) and \( C^- \) are the sets of all codewords pair \( (\mathbf{u}, \mathbf{c}) \) that gives the decision of \( u_0 = 0 \) and \( u_0 = 1 \), respectively, and \( \Gamma_{\mathbf{u}, \mathbf{c}}^{(j)} \) is the error event metric for \( (\mathbf{u}, \mathbf{c}) \) in the \( j \)th iteration, defined as

\[ \Gamma_{\mathbf{u}, \mathbf{c}}^{(j)} = \sum_{i \in \{i : u_i = 1\}} \eta_i^{(j)} + L_c \sum_{i \in \{i : c_i = 1\}} y_i. \]

In (5–29), \( i \in \{i : u_i = 1\} \) and \( i \in \{i : c_i = 1\} \) mean taking the indices of the non-zero bits in \( \mathbf{u} \) and \( \mathbf{c} \), \( \eta_i^{(j)} \) is the a priori information of the \( i \)th information bit, and

\[ L_c = \frac{2}{\sigma_n^2}. \]
is known as the channel reliability measure. A detailed proof of (5–28) can be found in Appendix B. Note that since the all-zero codeword \((\mathbf{0}, \mathbf{0}) \in C^+\) and \(\Gamma^{(j)}_{\mathbf{0}, \mathbf{0}} = 0\), we have

\[
\max_{(u, c) \in C^+} \{-\Gamma^{(j)}_{u, c}\} = \max_{(u, c) \in C^+} \{0, -\Gamma^{(j)}_{u, c}\} \geq 0. 
\]  
(5–31)

With (5–31) we can obtain following union bound from (5–28) for the probability that \(\xi^{(j)}_k\) is smaller than an arbitrary value \(x\),

\[
P(\xi^{(j)}_k < x) = P\left(\max_{(u, c) \in C^+} \{-\Gamma^{(j)}_{u, c}\} + \min_{(u, c) \in C^-} \{\Gamma^{(j)}_{u, c}\} < x\right) 
\leq P\left(\min_{(u, c) \in C^-} \{\Gamma^{(j)}_{u, c}\} < x\right) = \frac{1}{K_c} P\left(\bigcup_{(u, c) \in C^-} \Gamma^{(j)}_{u, c} < x\right) 
\leq \frac{1}{K_c} \sum_{(u, c) \in C^-} P(\Gamma^{(j)}_{u, c} < x), 
\]  
(5–32)

where \(K_c\) is the number of input bits per trellis state transition.

Now, let \(d = w(c)\) and \(w = w(u)\) denote the Hamming weights of the codeword \(c\) and the corresponding information bit sequence \(u\), respectively. Since the error event metric \(\Gamma^{(j)}_{u, c}\) in (5–29) does not depend on the codeword pattern and \(k\), but only on the weights \(w\) and \(d\), we can rewrite the metric as

\[
\Gamma^{(j)}_{w, d} = \sum_{i=1}^{w-1} \eta^{(j)}_i + L_c \sum_{i=0}^{d-1} y_i. 
\]  
(5–33)

Since the statistical property for \(\Gamma_{w, d}\) and the probabilities in (5–33) can be regarded as independent of the error events starting position for large coding block size \([14, 34]\), the union bound is valid for arbitrary data bit. Thus, we have dropped the subscript \(k\) in \(\xi_k\) for clarity. Also, we have indexed the other \(w – 1\) non-zero bits \(u_i\) in \(u\) as \(i = 1, 2, \ldots, w – 1\) in (5–33) without loss of generality.

Thus, by using (5–33) and dropping the subscript \(k\), the union bound in (5–32) can be written as

\[
P(\xi^{(j)} < x) \leq \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w, d} P(\Gamma^{(j)}_{w, d} < x). 
\]  
(5–34)
where \( d_{\text{min}} \) is the minimum Hamming distance of the code \( \mathcal{C} \), and \( A_{w,d} \) is the number of error events with Hamming weight \( d \) and input weight of \( w \). Eq. (5–34) is a generalized union bound of max-log-MAP decoding. The well known union bound for maximum likelihood decoding is a special case of (5–34) with \( x = 0 \) and the \textit{a priori} information equal to 0.

### 5.2.3 Applying Max-log-MAP Decoding Union Bound to Collaborative Decoding

To apply the generalized union bound in (5–34) to collaborative decoding, the crucial step is the evaluation of the probability \( P(\Gamma_{w,d}^{(j)} < x) \). Thus we study the error event metric in (5–33). From (5–12), we know that in the \( j \)th decoding iteration, not all the \( w - 1 \) non-zero information bits in \( \mathbf{u} \) (the first non-zero bit \( u_0 \) itself is excluded) can obtain the \textit{a priori} information. Among those bits that obtain \textit{a priori} information, some of them obtain \( \lambda^{(l)}_i \) while the other obtain \( \tilde{\lambda}^{(l)}_i \) for \( l < j \). For convenience, we use \( A_l \) to denote the bit set in the \( w - 1 \) nonzero bits of \( \mathbf{u} \) that obtain additional information \( \lambda^{(l)}_i \) in the \( l \)th iteration for \( l < j \). This means that in the \( l \)th iteration the event \( \mathcal{R}^{(l)}_i \) defined in (6–10) occurs only for \( i \in A_l \) (we do not distinguish bit and bit index here for convenience) in the \( w - 1 \) non-zero information bits of \( \mathbf{u} \). Further, we define \( \hat{A}_l \) as the subset of \( A_l \) that the event \( \mathcal{R}^{(l)}_i \) occurs for \( i \in \hat{A}_l \), which means the information bits in \( \hat{A}_l \) obtain additional information \( \lambda^{(l)}_i \) in the \( l \)th iteration. Also, we define \( \tilde{A}_l \) as the complementary subset of \( \hat{A}_l \) that the event \( \mathcal{R}^{(l)}_i \) occurs for \( i \in \tilde{A}_l \), i.e., those bits obtain additional information \( \tilde{\lambda}^{(l)}_i \) in the \( l \)th iteration. Note due to the fact that no information can be exchanged for a non-candidate bit, we have \( A_l \cap A_k = \emptyset \) for \( l \neq k \). With these notations, the above event can be expressed as

\[
V_j = \left\{ \bigcap_{l=0}^{j-1} \left\{ \bigcap_{i \in \hat{A}_l} \mathcal{R}^{(l)}_i, \bigcap_{i \in \tilde{A}_l} \mathcal{R}^{(l)}_i, \bigcap_{i \in A_l} \mathcal{B}^{(l)}_i, \bigcap_{i \in B_j} \mathcal{B}^{(j)}_i \right\} \right\}, \quad (5–35)
\]
where \( B_j = \bigcup_{l=0}^{j-1} A_l \) is the bit set for which no information exchange occurs in the previous \( j-1 \) iterations. The set \( B_j \) contains all the non-zero candidate bits left for the \( j \)th decoding iteration. From (5–33), the error event metric associated with event \( V_j \) can be defined as

\[
\Gamma_{w,d}^{(j)} = \sum_{l=0}^{j-1} \sum_{i \in A_l} \lambda_i^{(l)} + Y_d = \sum_{l=0}^{j-1} \left( \sum_{i \in \hat{A}_l} \hat{\lambda}_i^{(l)} + \sum_{i \in \ddot{A}_l} \ddot{\lambda}_i^{(l)} \right) + Y_d, \tag{5–36}
\]

where

\[
Y_d := L_c \sum_{i=0}^{d-1} y_i. \tag{5–37}
\]

From (4–1) we know that \( Y_d \sim \mathcal{N}(dL_c, 2dL_c) \).

Since in iteration \( l \) the extrinsic information \( \{\xi_i^{(l)}\} \) are i.i.d. for all \( i \), the statistical characteristics of \( \Gamma_{w,d}^{(j)} \) in (5–36) and the probability of \( V_j \) only depends on the size of \( \hat{A}_l \) and \( \ddot{A}_l \) for \( l < j \) given the statistical knowledge of the extrinsic information, but not on the particular choices of the bit sets. Let

\[
|A_l| = m_l, \quad |\hat{A}_l| = n_l \tag{5–38}
\]

with \( 0 \leq n_l \leq m_l \). Due to the fact that the events \( \hat{R}_i^{(l)} \) and \( \ddot{R}_i^{(l)} \) are disjoint, we know that \( \hat{A}_l \cap \ddot{A}_l = \emptyset \). Since \( A_l = \hat{A}_l \cup \ddot{A}_l \), we have

\[
|\hat{A}_l| = m_l - n_l, \tag{5–39}
\]

which is sufficiently determined by (5–38). Thus, to determine the statistical characteristics of \( \Gamma_{w,d}^{(j)} \) and \( V_j \), it is sufficient to specify a \( 2j \)-tuple \( V_j \) that

\[
V_j = \{|A_l| = m_l, |\hat{A}_l| = n_l\}_{l=0}^{j-1}. \tag{5–40}
\]
For convenience, we use $\Gamma_{w,d}^{(j)}(V_j)$ to denote the error event metric with a particular $V_j$. Then we have $\Gamma_{w,d}^{(j)}(V_j) \sim \mathcal{N}(\mu(V_j), \sigma^2(V_j))$, where

$$
\mu(V_j) = dL_c + \sum_{l=0}^{j-1} \phi_l \mu_l, \text{ and } \sigma^2(V_j) = 2dL_c + \sum_{l=0}^{j-1} \phi_l \sigma^2_l,
$$

in which

$$
\phi_l = m_l (M - 1) - n_l.
$$

Recall that in the LRB exchange scheme no information can be exchanged for a non-candidate bit, i.e., $A_l \cap A_k = \emptyset$ for $l \neq k$. Thus, the value of $m_l$ in (5–38) must satisfy

$$
0 \leq m_l \leq w_l,
$$

where

$$
w_l = w_{l-1} - m_{l-1}, \text{ and } w_0 = w - 1,
$$

is the number of non-zero candidate bits left in $\mathbf{u}$ given the event $\{|A_l| = m_l\}_{l=0}^{j-1}$ occurs. Based on the above arguments, the probability $P(\Gamma_{w,d}^{(j)} < x)$ can be calculated as

$$
P(\Gamma_{w,d}^{(j)} < x) = \sum_{V_j}^{(2j)} P(\Gamma_{w,d}^{(j)}(V_j) < x, |V_j| = V_j),
$$

where $\sum_{V_j}^{(2j)}$ means the $(2j)$-fold summation over all possible values of $V_j$, i.e.,

$$
\sum_{V_j}^{(2j)} = \sum_{m_0=0}^{w_0} \sum_{m_1=0}^{w_1} \cdots \sum_{m_{j-1}=0}^{m_{j-1}} \sum_{n_0=0}^{m_0} \sum_{n_1=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_{j-1}},
$$

and we use $\{|V_j| = V_j\}$ to denote the occurrence of all possible sets $\mathcal{A}(V_j) = \{A_l, \hat{A}_l, \tilde{A}_l\}_{l=0}^{j-1}$ satisfying that $\{|A_l| = m_l, |\hat{A}_l| = n_l\}_{l=0}^{j-1}$. Since the $(2j)$-tuple $V_j$ only constrains the size of $A_l$ and $\hat{A}_l$, for all $l$, $A_l$ can be an arbitrary bit set in the $w_l$ nonzero bits of $\mathbf{u}$ and the subset $\hat{A}_l$ can be arbitrary subset in $A_l$. Hence, for a given $V_j$, there are

$$
\prod_{l=0}^{j-1} \binom{w_l}{m_l} \binom{m_l}{n_l}
$$
possible choices of \( A(V_j) \). For all these choices, the probabilities of the event \( \{ V_j = A(V_j) \} \), are the same. Thus, we can upper bound (5–45) by upper bounding \( P(\Gamma_{w,d}^{(j)}(V_j) < x, \mathcal{V}_j = \mathcal{A}(V_j)) \) for each particular choice of \( \mathcal{A}(V_j) \). In a manner similar to (5–17) and (5–25), we drop all the events associated with \( \dot{\lambda}_i^{(l)} \) or \( \ddot{\lambda}_i^{(l)} \) in \( V_j \) and use (5–20) and (5–21) to obtain

\[
P(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j) \\
= \prod_{t=0}^{j-1} \left( \frac{w_t}{m_t} \right) \left( \frac{m_t}{n_t} \right) P(\Gamma_{w,d}^{(j)}(V_j) < x, \mathcal{V}_j = \mathcal{A}(V_j)) \\
\leq \prod_{t=0}^{j-1} \left( \frac{w_t}{m_t} \right) \left( \frac{m_t}{n_t} \right) P(\Gamma_{w,d}^{(j)}(V_j) < x) \\
\times P \left( \bigcap_{l=0}^{j-1} \bigcap_{i \in \mathcal{A}_j} \left\{ \bigcup_{r \in \mathcal{M}} \{ |\xi_{r,i}^{(l)}| < T_l, \bigcap_{t=0}^{l-1} |\xi_{r,i}^{(t)}| < T_t, \bigcap_{t=0}^{l} |\xi_{M,i}^{(t)}| > T_t \} \right\} \right) \\
\times P \left( \bigcap_{l=0}^{j-1} \bigcap_{i \in \mathcal{A}_j} \bigcap_{t=0}^{l-1} |\xi_{M,i}^{(l)}| > T_t \right) P \left( \bigcap_{i \in \mathcal{B}_j} \mathcal{B}_i^{(j)} \right) \\
\leq c'(V_j) P(\Gamma_{w,d}^{(j)}(V_j) < x),
\] (5–47)

where \( c'(V_j) \) is calculated as

\[
c'(V_j) = \prod_{t=0}^{j-1} \left( \frac{w_t}{m_t} \right) \left( \frac{m_t}{n_t} \right) \left\{ [(M - 1)p_t]^{m_t} (1 - p_t)^{Mw_j + n_t} \prod_{t=0}^{l-1} (1 - p_t)^{m_t + n_t} \right\}.
\] (5–48)

On the other hand, we know that \( P(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j) \leq P(\Gamma_{w,d}^{(j)}(V_j) < x) \).

Then the upper bound in (5–47) can be refined as

\[
P(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j) \leq c(V_j) P(\Gamma_{w,d}^{(j)}(V_j) < x),
\] (5–49)

where \( c(V_j) = \min\{1, c'(V_j)\} \), and

\[
P(\Gamma_{w,d}^{(j)}(V_j) < x) = Q \left( \frac{\mu(V_j) - x}{\sigma(V_j)} \right)
\] (5–50)
with $Q(\cdot)$ being the Gaussian $Q$-function. Then by inserting (5–49) into (5–45) we have

$$P(\Gamma_{w,d}^{(j)} < x) \leq \sum_{V_j} c(V_j)P(\Gamma_{w,d}^{(j)} < x).$$  \hspace{1cm} (5–51)

Combining (5–50), (5–51) and (5–34), we obtain following upper bound

$$P(\xi^{(j)} < x) \leq \frac{1}{Kc} \sum_{d \geq d_{\text{min}}} \sum_{w \geq 1} wA_{w,d} \sum_{V_j} c(V_j)Q\left(\frac{\mu(V_j) - x}{\sigma(V_j)}\right).$$  \hspace{1cm} (5–52)

This closed-form bound is also readily applied to the probabilities $P(\xi^{(j)} + \hat{\lambda}^{(l)} < 0)$, $P(\xi^{(j)} + \hat{\lambda}^{(l)} < 0)$, and $P(\xi^{(j-1)} < -T_{j-1})$ in (5–27) without any difficulty.

Now, we only have $P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1})$ left in (5–27) to evaluate. The difficulty here is the correlation between $\xi^{(j)}$ and $\xi^{(j-1)}$. To unveil this correlation, we consider the extrinsic information expression given in (5–28). Let

$$\left(\mathbf{u}, \mathbf{c}\right)_{\text{opt}}^+ = \arg \max_{(\mathbf{u}, \mathbf{c}) \in C^+} \left\{ -\Gamma^{(j)}_{\mathbf{u}, \mathbf{c}} \right\}$$

denote the optimal decoding sequence found by the decoder in $C^+$, meanwhile $(\mathbf{u}, \mathbf{c})_{\text{opt}}^-$ denote the optimal decoding sequence in $C^-$. According to max-log-MAP decoding, the final survival sequence $(\mathbf{u}, \mathbf{c})_{\text{opt}}$ is generated between $(\mathbf{u}, \mathbf{c})_{\text{opt}}^+$ and $(\mathbf{u}, \mathbf{c})_{\text{opt}}^-$. If $(\mathbf{u}, \mathbf{c})_{\text{opt}}^+$ is not selected to between the survivor sequence, it becomes the competing sequence. We assume the code is good enough that, when the SNR is not too low, the decoder can at least find the correct codeword as the competing sequence if it is not selected to be the survivor sequence. This assumption is the same as that used in [33]. Thus, under the assumption that the all-zero sequence $(\mathbf{0}, \mathbf{0})$ is transmitted, we have $(\mathbf{u}, \mathbf{c})_{\text{opt}}^+ = (\mathbf{0}, \mathbf{0})$ since $(\mathbf{0}, \mathbf{0}) \in C^+$. That is,

$$\max_{(\mathbf{u}, \mathbf{c}) \in C^+} \left\{ -\Gamma^{(j)}_{\mathbf{u}, \mathbf{c}} \right\} = \Gamma^{(j)}_{\mathbf{0}, \mathbf{0}} = 0.$$
With the above arguments, we can rewrite (5–28) as follows by dropping the first term

\[ \xi^{(j)} \approx \min_{(u, c) \in \mathcal{C}} \{ \Gamma^{(j)}_{u, c} \} \]

when the SNR is high. Thus, for \( j > 0 \) we have

\[
P(\xi^{(j)}_k < 0, \xi^{(j-1)}_k > T_{j-1}) \leq P\left( \min_{(u, c) \in \mathcal{C}} \{ \Gamma^{(j)}_{u, c} \} < 0, \min_{(u, c) \in \mathcal{C}} \{ \Gamma^{(j-1)}_{u, c} \} > T_{j-1} \right)
\]

\[
= \frac{1}{K_c} P\left( \bigcup_{(u, c) \in \mathcal{C}} \Gamma^{(j)}_{u, c} < 0, \bigcap_{(u, c) \in \mathcal{C}} \Gamma^{(j-1)}_{u, c} > T_{j-1} \right)
\]

\[
\leq \frac{1}{K_c} \sum_{(u, c) \in \mathcal{C}} P\left( \Gamma^{(j)}_{u, c} < 0, \bigcap_{(u', c') \in \mathcal{C}} \Gamma^{(j-1)}_{u', c'} > T_{j-1} \right)
\]

\[
\leq \frac{1}{K_c} \sum_{(u, c) \in \mathcal{C}} P(\Gamma^{(j)}_{u, c} < 0, \Gamma^{(j-1)}_{u, c} > T_{j-1}), \tag{5–53}
\]

Following the derivation from (5–33) through (5–52), we then obtain

\[
P(\xi^{(j)}_k < 0, \xi^{(j-1)}_k > T_{j-1}) \leq \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} \left[ wA_{w,d} \right.
\]

\[
\times \sum_{V_j^{(2j)}} c(V_j) P(\Gamma^{(j)}_{w,d}(V_j) < 0, \Gamma^{(j-1)}_{w,d}(V_{j-1}) > T_{j-1}) \bigg]. \tag{5–54}
\]

To evaluate the probability \( P(\Gamma^{(j)}_{w,d}(V_j) < 0, \Gamma^{(j-1)}_{w,d}(V_{j-1}) > T_{j-1}) \) in (5–54), we rewrite (5–36) as

\[
\Gamma^{(j)}_{w,d}(V_j) = \Gamma^{(j-1)}_{w,d}(V_{j-1}) + \Psi, \tag{5–55}
\]

where

\[
\Psi = \sum_{i \in \hat{A}_{j-1}} \hat{\lambda}^{(j-1)}_i + \sum_{i \in \hat{A}_{j-1}} \hat{\lambda}^{(j-1)}_i. \tag{5–56}
\]

Given \( V_j \), we know that \( \Gamma^{(j-1)}_{w,d}(V_{j-1}) \sim \mathcal{N}(\mu(V_{j-1}), \sigma^2(V_{j-1})) \), \( \Psi \sim \mathcal{N}(\phi_{j-1}\mu_{j-1}, \phi_{j-1}\sigma_{j-1}^2) \), and \( \Gamma^{(j-1)}_{w,d} \) and \( \Psi \) are independent of one another, where \( \mu(V_j), \sigma(V_j) \) and \( \phi_j \) are given
in (5–41) and (5–42), respectively. Thus, we have

\[
P(\Gamma_{w,d}^{(j)}(V_j) < 0, \Gamma_{w,d}^{(j-1)}(V_{j-1}) > T_{j-1})
\]

\[
= P(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Gamma_{w,d}^{(j-1)}(V_{j-1}) > T_{j-1})
\]

\[
= P(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Psi < -T_{j-1}) - P(\Gamma_{w,d}^{(j-1)}(V_{j-1}) < T_{j-1}, \Psi < -T_{j-1})
\]

\[
= P(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Psi < -T_{j-1}) - P(\Gamma_{w,d}^{(j-1)}(V_{j-1}) < T_{j-1})P(\Psi < -T_{j-1})
\]

\[
= Q\left(\frac{\mu(V_j) - \phi_{j-1}\mu_{j-1}}{\sigma(V_j)}; \frac{\phi_{j-1}\sigma_{j-1}}{\sigma(V_j)}\right)
\]

\[
- Q\left(\frac{\mu(V_{j-1}) - T_{j-1}}{\sigma(V_{j-1})}\right)Q\left(\frac{\phi_{j-1}\mu_{j-1} + T_{j-1}}{\phi_{j-1}\sigma_{j-1}}\right);
\]

(5–57)

where the relations

\[
\mu(V_j) = \mu(V_{j-1}) + \phi_{j-1}\mu_{j-1}, \quad \text{and} \quad \sigma^2(V_j) = \sigma^2(V_{j-1}) + \phi_{j-1}\sigma^2_{j-1}
\]

are used, and

\[
Q(x, y; \rho) \triangleq \int_x^\infty \int_y^\infty \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp\left(-\frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{2(1 - \rho^2)}\right)dz_1 dz_2
\]

(5–58)

is the bivariate Gaussian Q-function, for which [35] gives a simplified expression for numerical evaluation. To this point, we have upper bounded the BER \(P_b^{(j)}\) for the cases of \(M \geq 3\).

## 5.2.4 BER Upper Bound for \(M = 2\)

For the case of \(M = 2\), we note that the additional information \(\lambda_{k,i}^{(j)}\) defined in (5–3) becomes 0 for all \(k, i\) and \(j\), and the event \(\tilde{R}_{k,i}^{(j)}\) defined in (5–4) reduces to

\[
\tilde{R}_{k,i}^{(j)} = \{|\xi_{k,i}^{(j)}| \leq T_j\}.
\]

(5–59)

Due to these differences, it is necessary to make some modifications to the previous analysis to obtain a tight bound for this case. Again, we consider the error performance at the \(M\)th node. Following the notation in Section 5.2.1, the inequality in
except for (5–60) becomes

\[
P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) = P(\xi^{(j)} < 0, \bar{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \bar{\lambda}^{(l)} < 0, \bar{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)})
\]

\[
\leq P(\xi^{(j)} < 0, \bar{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \bar{\lambda}^{(l)} < 0). \quad (5–60)
\]

Analogous to (5–17) and (5–25), we upper bound the first term in (5–60) as

\[
P(\xi^{(j)} < 0, \bar{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) = P(\xi^{(j)} < 0, \left|{\xi^{(l)}_1}\right| \leq T_i, \left|{\xi^{(l)}_i}\right| > T_i, \mathcal{B}^{(l)})
\]

\[
\leq P\left(\xi^{(j)} < 0, \left|{\xi^{(l)}_i}\right| \leq T_i, \left|{\xi^{(l)}_i}\right| > T_i, \bigcap_{t=0}^{l-1} \left|{\xi^{(l)}_i}\right| > T_i\right)
\]

\[
= P(\xi^{(j)} < 0, \left|{\xi^{(l)}_i}\right| > T_i)P\left(\left|{\xi^{(l)}_i}\right| \leq T_i, \bigcap_{t=0}^{l-1} \left|{\xi^{(l)}_i}\right| > T_i\right)
\]

\[
\leq a_t[ P(\xi^{(l)} < -T_i) + P(\xi^{(j)} < 0, \xi^{(l)} > T_i) ], \quad (5–61)
\]

where \(a_t\) is the same as (5–24). Thus, by substituting (5–16) with (5–60), the upper bound of \(P^{(j)}_b\) in (5–27) becomes

\[
P^{(j)}_b \leq \sum_{l=0}^{j-1} \left\{ a_t [ P(\xi^{(l)} < -T_i) + P(\xi^{(j)} < 0, \xi^{(l)} > T_i) ] + P(\xi^{(j)} + \bar{\lambda}^{(l)} < 0) \right\}
\]

\[
+ b_t [ P(\xi^{(j)} < 0, \xi^{(j)} > T_i) ] - a_t [ P(\xi^{(j)} < 0, \xi^{(l)} > T_i) ], \quad (5–62)
\]

Then, similar to the case of \(M \geq 3\), we apply the union bound for max-log-MAP decoding to further upper bound the probabilities in (5–62). All the derivations are same as Section 5.2.3 except for (5–47) and (5–48). Due to the change in (5–59), \(\bar{\mathcal{R}}^{(l)}\) becomes independent of \(\bar{\lambda}^{(l)}\). Thus, for the case of \(M = 2\), we can keep \(\bar{\mathcal{R}}^{(l)}\) for \(i \in \bar{A}_t\), when we drop all the events associated with \(\bar{\lambda}^{(l)}\) in the derivation of (5–47). With this modification, \(c'(V_j)\) in (5–48) becomes

\[
c'(V_j) = \prod_{l=0}^{j-1} \left( \frac{w_l}{m_l} \right)^{\frac{m_l}{n_l}} \left[ p_l^{m_l} (1 - p_l)^{2w_l + n_l} \prod_{l=0}^{l-1} (1 - p_l)^{m_l + n_l} \right]. \quad (5–63)
\]
5.3 Numerical Results

In this section, we first present numerical results to demonstrate tightness of the BER upper bound developed in Section 5.2. Strictly speaking, this bound is an approximated upper bound due to the Gaussian approximation and the semi-analytical density evolution model. First, we set the number of iterations $I$ to 3 (i.e., 3 exchanges and 4 decoding iterations are performed in total), and set $\{p_j\}$ to $\{0.1, 0.15, 0.25\}$ in the collaborative decoding process. Fig. 5–5 compares the upper bounds in each iteration with the simulation results for the cases of $M = 2$ and $M = 6$, respectively. In the system, a non-recursive convolutional code with the generation polynomial of $[1 + D^2, 1 + D + D^2]$ is used. We denote it by CC(5, 7). From the figure, we see that the bounds in the low BER region are very close to the simulation results in all iterations. At the very low $E_b/N_0$ region (i.e., the high BER region), the bounds become loose. This is due to the nature of the union bound given in (5–32). In the figure, we also show the union bounds for MRC. We can see that,
Figure 5–6: Comparison of the proposed bounds, simulation results in the last iteration for the cases of $M = 2, 3, 4, 6$ and $8$ on AWGN channels, where CC(15,17) and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used.

for $M = 2$, the performance of collaborative decoding with the LRB exchange scheme is very close that of MRC, while it is about 2dB within that of MRC for $M = 6$. This means that most of the spatial gain can be obtained through collaborative decoding.

In Fig. 5–6, we show the results for another non-recursive convolutional code with the generation polynomial of $[1 + D^2 + D^3, 1 + D + D^2 + D^3]$, denoted by CC(15,17). The parameter $\{p_j\}$ is the same as that in Fig. 5–5. We compare the upper bounds with simulation results for $M = 2, 3, 4, 6, 8$, respectively. For clarity, we only show the BER in the last iteration for each $M$. We note that, due to the independency assumption and Gaussian approximation in Section 5.1 are not very accurate in the realistic decoding process for CC(15,17) when $M = 2$, the BER bounds is a little bit below the simulation results. However, when $M \geq 3$ the assumptions become much closer to the actual situation. From the figure, we can see that the bounds are very tight.
Table 5–1: Different choices of \( \{p_j\} \) and corresponding average information exchange amount \( \Theta \) with \( M = 8 \) for rate 1/2 CC(5,7) code. \( \Theta \) is calculated with respect to the information exchange amount of MRC, \( \Theta_{MRC} \).

<table>
<thead>
<tr>
<th>No. of exchanges</th>
<th>Value of ( {p_j}_{i=0}^{I-1} )</th>
<th>( \Theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>( I = 3 ), {0.1, 0.15, 0.25}</td>
<td>( \Theta_1 = 0.458\Theta_{MRC} )</td>
</tr>
<tr>
<td>Case 2</td>
<td>( I = 3 ), {0.055, 0.098, 1}</td>
<td>( \Theta_2 = 0.466\Theta_{MRC} )</td>
</tr>
<tr>
<td>Case 3</td>
<td>( I = 5 ), {0.0405, 0.0564, 0.0897, 0.1902, 1}</td>
<td>( \Theta_3 = 0.456\Theta_{MRC} )</td>
</tr>
<tr>
<td>Case 4</td>
<td>( I = 1 ), {1}</td>
<td>( \Theta_4 = 0.5\Theta_{MRC} )</td>
</tr>
</tbody>
</table>

With the proposed analysis tools, we can illustrate the effect of different choices of \( \{p_j\} \) to the error performance in collaborative decoding by comparing the BER upper bounds. While comparing error performance for different collaborative decoding processes, it is important to consider the necessary amount of information exchange during the process. Here, we use \((4–5)\) to calculate the average information exchange amount \( \Theta \), and the information exchange amount of MRC \( \Theta_{MRC} \) is calculated by \((4–5)\) for the purpose of comparison.

Below, we fix the setting of \( M = 8 \) and rate 1/2 code CC(5,7), and compare the 4 cases listed in Table 5–1. In Table 5–1, \( \{p_j\} \) in case 2 is chosen such that for \( M = 8 \), the amount of information each node sending out is almost the same in different iterations on average. In case 3, \( \{p_j\} \) is chosen to make each node request information from other nodes for almost the same number of information bits in different iterations. In case 4, \( \{p_j\} = \{1\} \) means that the nodes exchange information only once, and each node requests information from other nodes for all the information bits. In Fig. 5.3, we show the BER bounds of each iterations for the all 4 cases. From the figure, we see that cases 2, 3 and 4 achieve the same performance in their last iterations, and outperform case 1. In Fig. 5.3, we compare the BER bounds of the all 4 case in their last iterations with that of MRC in the very low BER region. This approximately shows the asymptotic performance of the systems. In the figure, we see that in case 2, 3 and 4, the receivers finally achieve the same error performance as MRC, but with much less information exchange amount. This shows that with proper
Figure 5–7: Comparison of performance for $M = 8$ and CC$(5, 7)$ on AWGN channels with different choices of $\{p_j\}$ in Table 5–1

Figure 5–8: Asymptotic performance for $M = 8$ and CC$(5, 7)$ on AWGN channels with different choices of $\{p_j\}$ in Table 5–1
choices of \{p_j\}, full spacial diversity can be achieved by the collaborative decoding technique.

5.4 Summary

We have analyzed the bit error performance for collaborative decoding with LRB exchange. A density evolution model is proposed to simplify the analysis. With Gaussian approximation, knowledge of the extrinsic information are obtained by simulating the proposed model over AWGN channels. Then, we derive an upper bound for the BER of the collaborative decoding process via a generalized union bound for the max-log-MAP decoder. Numerical results demonstrate the tightness of the bounds. We also show that with proper parameters design, collaborative decoding with LRB exchange can achieve the same performance of MRC at high SNRs. The analysis provides an efficient way to evaluate the error performance of the collaborative decoding system.

The analysis is based on the observation that the extrinsic information generated in the collaborative decoding process can be well approximated by Gaussian distributions when non-recursive convolutional codes are used in the system. This advantage makes the calculations in the analysis simple. For recursive convolutional codes, if we can find a proper probability distribution model for the extrinsic information, then by replacing the Gaussian approximation with the new model, our analysis can be extended to the recursive convolutional code case. Gaussian mixture model [36] is a possible solution in this case.
In Chapter 5 we have analyzed the error performance of collaborative decoding with LRB information exchange over an AWGN channel. The method is primarily based on the density evolution model and the Gaussian approximation for extrinsic information generated by nonrecursive convolutional codes in collaborative decoding. With the statistic characteristics of additional information, it is possible to study to the error events, hence the pairwise error probability, of any single decoding process in collaborative decoding. Once the pairwise error probabilities can be obtained, the error performance of the decoder are evaluated by applying the union bound of MAP decoding as in Section 5.2.2. In this Chapter, we extend this method to the scenario of collaborative decoding with the MRB information exchange scheme, when nonrecursive convolutional codes are.

Similar to the case of LRB, in AWGN channel we still apply Gaussian approximation the to extrinsic information generated by the nonrecursive convolutional codes in MRB. However, for independent Rayleigh fading channels, the density function of extrinsic information exhibits apparent asymmetric property, especially at the middle to high SNR region. Hence, the Gaussian approximation used in the AWGN channel analysis precisely is not valid any more for the case of independent Rayleigh fading channels. Fortunately, from simulation we find that in this case. The statistical characteristics of extrinsic information generation by the nonrecursive convolutional codes in collaborative decoding can be well approximated by a set of generalized asymmetric Laplace distributions \[37\]. This approximate parametric description makes it possible
to describe the statistical behavior of the extrinsic information by a set of few parameters. Hence, the density evolution model can be extend to the case of independent Rayleigh fading channel for collaborative decoding.

With proper statistical approximations and density evolution model, the major work of the performance analysis for collaborative decoding with MRB exchange becomes the evaluation of the pairwise error probabilities (PEP) in the MAP decoding process. Different from LRB, due to the nature of the MRB information exchange scheme the additional information at each decoder involves sum of truncated extrinsic information from other decoders, and the number of such truncated extrinsic information terms is also random. This makes the analysis of the PEP in MRB much more complicated than that in LRB. In this chapter, we primarily rely on upper-bound techniques and combinatorial theory to derive the error probabilities. Laplace transform and saddle point approximation techniques based on moment generating functions are the major tools used in evaluating the upper bounds.

The system model we study in this chapter is described in Section 4.1. The collaborative decoding with MRB exchange is described in Section 4.2.3. We only consider the performance analysis of nonrecursive convolutional codes in this chapter.

The remainder of this chapter is arranged as follows. In Section 6.1, we describe the Gaussian and the generalized asymmetric Laplace approximations of the extrinsic information in the collaborative decoding for the AWGN and independent Rayleigh fading channel models, respectively. In Section 6.2, a density evolution model is developed to evaluate statistical parameters for the extrinsic information. In Section 6.3, a uniform upper bound of bit error rate (BER) is provided in term of probabilities involving the extrinsic information in the current iteration. In Section 6.4 we further study the error event behaviors in MAP decoding with the effect of MRB information exchange, and develop some upper bounds of the PEP involved in the BER bound. In Section 6.5, we address the numerical evaluation for the BER upper bound with
the statistical knowledge of extrinsic information for the AWGN and independent Rayleigh fading channel models, respectively. Then, numerical results are presented in Section 6.6. Finally, a summary is drawn in Section 6.7.

### 6.1 Statistical Approximation for Extrinsic Information

From the procedure of the MRB exchange scheme described in Section 4.2.3, we know that all nodes in the collaborative decoding process are symmetric. This implies that the statistical characteristics of the extrinsic information at all nodes are the same. Thus, we only need to consider a single node, for example, the $M$th node. In order to study the extrinsic information generated by the MAP decoder in collaborative decoding, we will, without loss of generality, assume that the all-zero codeword is transmitted through the channel in following. Under this assumption, we seek to find the probability distribution of the extrinsic information for each data bit. Our performance analysis will be based on the knowledge of statistical characteristics of the extrinsic information.

As we know, the extrinsic information is generated by finding the minimum (or maximum) in a large (theoretically infinite) set of sequence metrics in the MAP decoding process. The sequence metrics are of non-identical distribution and dependent with each other. From the extreme value theory, the closed-form distribution of the minimum (or maximum) does not exist for dependent non-identical random variables generally. In fact, even the type of the distribution for extrinsic information generated by a MAP decoder is very difficult to find analytically.

In this case, a feasible approach to simplify the study of the MAP decoding process is to employ approximate models to describe the distribution of the extrinsic information. By using a simple distribution that fits the observed histogram of extrinsic information obtained from simulations, statistic behavior of the decoder can be quantified and studied in a semi-analytic way. This approach has been successfully applied to the study of iterative decoding process for many codes such as turbo codes...
and low-density parity check (LDPC) codes in [30, 31, 38, 39]. Analysis with this approximation approach turns out to give considerably accurate results in the study of convergence property for turbo codes and LDPC codes. Based on this approximation, effective convergence analysis techniques such as the density evolution model and extrinsic information transfer (EXIT) chart are also developed. These techniques have been widely used in the analysis of many iterative decoding, detection and equalization algorithms. Following this idea, we also use empirical approximation to avoid the analytically unsolvable problem of finding the distribution of the extrinsic information in MRB.

Generic distribution fitting or estimation is a well studied topic in statistics. There are many techniques available for learning distribution of the extrinsic information in collaborative decoding. Techniques such as bootstrap sampling and mixture model can usually fit any distribution very well. However, these techniques usually belong to nonparametric methods or parametric methods with a great number of parameters. Representing the extrinsic information by such kind of distributions usually provides very little benefit to the analysis of error events associated with the extrinsic information generated in the decoding process. Hence, we consider using analytic distributions with only a small number of parameters to approximate statistic characteristics of the extrinsic information.

6.1.1 AWGN Channel

For AWGN channels, it is a well known observation that extrinsic information of information bits generated by a MAP decoder for convolutional codes and turbo codes can be well approximated by independent Gaussian random variables. Although the error events determining the decoding performance become much more complicated than those in a regular iterative decoding process, the Gaussian-like property of the extrinsic information still persists in collaborative decoding with MRB information exchange for nonrecursive convolutional codes.
Figure 6–1: Empirical pdfs of extrinsic information generated by the MAP decoder at successive iterations in collaborative decoding with MRB exchange for $M = 6$ on an AWGN channel with $E_b/N_0 = 5$dB, for which the maximum free distance 4-state nonrecursive convolutional code is used.

Following the notation in Chapter 5, we use $\xi_{k,i}^{(j)}$ to denote the extrinsic information generated by the MAP decoder for the $i$th data bit at node $k$ in the $j$th decoding iteration, and $y_{k,i}$ to denote the $i$th sample of the channel observation at node $k$. In Fig. 6–1, we show the typical histograms of the extrinsic information generated by the MAP decoder at successive iterations in collaborative decoding with MRB exchange for nonrecursive convolutional codes. True Gaussian density functions are compared with the histograms in the figure. Apparently, the Gaussian density functions can approximate the histograms very closely. In fact, we have verified the accuracy of the Gaussian approximation for MRB by extensive simulations using different choices of system parameters such as nodes number $M$, information exchange percentage $\{p_j\}$ and SNR.

Besides the Gaussian approximation, independence is another fundamental assumption for extrinsic information in this chapter. In [31], Gamal and Hammons
established an independence assumption for turbo decoders that any collection of intrinsic information (channel observations) and extrinsic information generated by all of the constituent MAP decoders in all decoding iterations are pairwise independent. This assumption is justified by using the arguments from the viewpoint of the graph structures of codes. As pointed out in Chapter 5, collaborative decoding actually is closely related to the turbo decoding in such way that the MAP decoders at different nodes can be viewed as the constituent decoders in turbo decoders, and the information exchange process in collaborative decoding is analogous to the extrinsic information passing among all constituent decoders in turbo decoders. Thus, the independence assumption for turbo decoders can be directly applied to our case.

Thus, similar to the case of LRB, we formalize the independent Gaussian assumption for the MAP decoder in collaborative decoding with MRB exchange and nonrecursive convolutional codes as follows:

**Assumption 6.1.1 (Independent Gaussian Assumption)** In collaborative decoding with MRB exchange and nonrecursive convolutional codes over AWGN channels, for arbitrary \( j \geq 0 \), the random sequences

\[
\{y_{k,0}, y_{k,1}, \ldots, y_{k,i}, \ldots\}, \{\xi_{k,0}^{(j)}, \xi_{k,1}^{(j)}, \ldots, \xi_{k,i}^{(j)}, \ldots\} \text{ for all } k \in \mathcal{M}
\]

are jointly Gaussian and statistically independent in the sense that any finite collection of the \( y_{k,i} \)'s and \( \xi_{k,i}^{(j)} \)'s are jointly Gaussian and pairwise independent. Also, for arbitrary \( k, r \in \mathcal{M} \) such that \( k \neq r \), the random sequences

\[
\{\xi_{k,0}^{(j)}, \xi_{k,1}^{(j)}, \ldots, \xi_{k,i}^{(j)}, \ldots\}, \{\xi_{r,0}^{(l)}, \xi_{r,1}^{(l)}, \ldots, \xi_{r,i}^{(l)}, \ldots\} \text{ for all } j, l \geq 0
\]

are jointly Gaussian and statistically independent.

Note that, due to the fact that MRB exchange is a memory based information exchange scheme, i.e., the information exchange process for a data bit \( i \) depends on
its status (candidate or non-candidate) in the previous iteration, the extrinsic information for the bit at the same node in different iterations are not independent generally. That is, $\xi_{k,i}^{(j)}$ and $\xi_{k,i}^{(l)}$ for $l \neq j$ are not independent. This is different from the independent assumption established in [31].

With the above assumption, based on the argument that all the information bits are statistically equivalent\(^1\), for any fixed $k$ and $j$, $\xi_{k,i}^{(j)}$ are identically distributed for all bit indices $i$. Thus, with the symmetry of nodes in collaborative decoding, we have that $\xi_{k,i}^{(j)}$ for all $k$ and $i$ are i.i.d. Gaussian random variables for any given $j$. Specifically, $\xi_{k,i}^{(j)} \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for all $k$ and $i$, where $\mathcal{N}(\mu_j, \sigma_j^2)$ denotes the Gaussian distribution with mean $\mu_j$ and variance $\sigma_j^2$. This means that the statistical behavior of all the extrinsic information generated at each iteration in collaborative decoding can be sufficiently described by two parameters: their mean and variance.

### 6.1.2 Independent Rayleigh Fading Channel

For independent Rayleigh fading channels with perfect channel state information (CSI), the independence assumption for extrinsic information still holds. However, the Gaussian assumption becomes invalid. Histograms of extrinsic information obtained from simulation show asymmetry of the distribution, especially in the middle to high SNR region. To the best of our knowledge, there is no result on the distribution fitting or approximating for extrinsic information generated by MAP decoders over Rayleigh fading channel before. In this chapter, we first propose to use the generalized asymmetric Laplace (GAL) distribution [37] to approximate the extrinsic information for Rayleigh fading channels.

\(^1\) Strictly speaking, the bits at the end of a block may have different statistical behavior from other bits. However, we usually assume very large block sizes in the decoding process so that the effect of a few bits at edges of a block on the average behavior of the block can be neglected.
A GAL probability density function (pdf) is defined as

\[
f(x) = \frac{\sqrt{2} e^{-\frac{\mu^2}{2\sigma^2}}}{\sqrt{\pi\sigma\Gamma(\tau)}} \left( \frac{|x|}{\sqrt{\mu^2 + 2\sigma^2}} \right)^{\frac{\tau - \frac{1}{2}}{2}} K_{\frac{\tau - \frac{1}{2}}{2}} \left( \frac{\sqrt{\mu^2 + 2\sigma^2}}{\sigma^2} |x| \right)
\]

(6–1)

for \( \mu \in \mathbb{R} \) and \( \sigma, \tau \geq 0 \), where \( \Gamma(\cdot) \) is the Gamma function, and \( K_\nu(\cdot) \) is the modified Bessel function of the third kind with index \( \nu \). Any random variable \( X \) that has the pdf defined in (6–1) is a GAL random variable. We denote this as \( X \sim \mathcal{GAL}(\mu, \sigma^2, \tau) \). The moment generating function (mgf) \( \Phi(s) \) of a GAL random variable \( X \) is defined as the double-sided Laplace transform of \( f(x) \), i.e.,

\[
\Phi(s) = E[e^{-sX}] = \frac{1}{(1 + \mu s - \frac{\sigma^2}{2} s^2)^{\tau}},
\]

(6–2)

and the region of convergence (ROC) of \( \Phi(s) \) is

\[
\frac{\mu - \sqrt{\mu^2 + 2\sigma^2}}{\sigma^2} < \Re(s) < \frac{\mu + \sqrt{\mu^2 + 2\sigma^2}}{\sigma^2},
\]

(6–3)

where \( \Re(\cdot) \) means taking the real part.

An important property of GAL random variable is that any random variable \( X \sim \mathcal{GAL}(\mu, \sigma^2, \tau) \) admits a mixture representation as follows [37]:

\[
X = \mu W + \sigma \sqrt{W} Z,
\]

(6–4)

where \( W \) and \( Z \) are two statistically independent random variables with the properties that \( Z \sim \mathcal{N}(0, 1) \) and that \( W \) is Chi-square distributed with degree of freedom 2\( \tau \),

\[\text{Different from the definition in [37], we omit the location parameters } \theta \text{ since only the case of } \theta = 0 \text{ is considered in this work.}\]

\[\text{For convenience, we will treat } \sigma^2 \text{ instead of } \sigma \text{ as the parameter in a GAL distribution. Hence, we use a notation different from the one, i.e., } \mathcal{GAL}(\mu, \sigma, \tau), \text{ used in [37]. However, this difference does not change the definition or properties of the GAL distribution.}\]
i.e., the pdf of $W$ is

$$g(x) = \frac{x^{\tau-1}}{\Gamma(\tau)} e^{-x}, \ x > 0.$$  \hspace{1cm} (6–5)

Another important property of GAL random variables is self-decomposability. That is, given an arbitrary number of pairwise independent GAL random variables $X_1, X_2, \ldots, X_i, \ldots$ with common parameters $\mu$ and $\sigma^2$, i.e., $X_i \sim \mathcal{GAL}(\mu, \sigma^2, \tau_i)$ for all $i$, the sum of these random variables, $S = \sum_i X_i$, still has a GAL distribution. The distribution of $S$ is given as $\mathcal{GAL}(\mu, \sigma^2, \tau)$ with $\tau = \sum_i \tau_i$. This property can be easily proved by using the characteristic function of GAL distribution [37], and will be useful in our analysis.

The GAL distribution is closely related to the MAP decoding process over independent Rayleigh fading channels with perfect CSI. With perfect CSI, the channel observation $y_{k,i}$ defined in (4–1) will be scaled by $L_cg_{k,i}$, where $L_c$ is the channel reliability measure defined in (5–30) and $g_{k,i}$ is the channel fading gain corresponded to the observation $y_{k,i}$. This scaled signal sequence $\{L_cg_{k,i}y_{k,i}\}$ is then input to the decoder as the intrinsic information to perform MAP decoding. Since the channel gain sequence $\{g_{k,i}\}$ are i.i.d. Rayleigh random variables, it is easy to verify that $L_cg_{k,i}y_{k,i}$ takes a representation consistent with (6–4). This means that the intrinsic information of the MAP decoder is a sequence of i.i.d. GAL random variables. Moreover, with the self-decomposable property of GAL random variable, the error event metric due to the intrinsic information in the MAP decoding process also has a GAL distribution. This can be seen in the analysis.

Usually, the statistical behavior of MAP decoders is determined by the statistical distribution of its input variables [31]. With the GAL distributed intrinsic information, it turns out that the statistic characteristics of extrinsic information generated by the MAP decoders is very close to those of GAL random variables. In general, the GAL-like extrinsic information is then used as a priori information for the MAP decoders in the iterative decoding process. This GAL-like input with the intrinsic
information will generate the new GAL-like extrinsic information again. Thus, the extrinsic information generated in the regular iterative decoding process, such as turbo decoding, can be approximated by GAL random variables very closely.

Due to its simplicity and nice properties, it is attractive to use the GAL approximation of extrinsic information for the purpose of performance analysis. Motivated by this reason, we also consider employing the GAL approximation for the analysis of collaborative decoding with MRB exchange. Although the a priori information in collaborative decoding may deviate from the GAL distributions due to the MRB information exchange procedure, we observe empirically that the histogram shape of the extrinsic information is not dragged too far away from GAL distributions by the deviation. This is illustrated in Fig. 6–2. In the figure, the typical histograms of the extrinsic information at successive iterations in collaborative decoding with MRB exchange for nonrecursive convolutional codes are compared with the corresponding GAL distributions. The similarity between the histograms and the GAL distributions justifies the GAL approximation for extrinsic information.

Based on the above discussion, we formalize the independent GAL assumption for collaborative decoding over independent Rayleigh fading channel as follows:

**Assumption 6.1.2 (Independent GAL Assumption)** In collaborative decoding with the MRB exchange and nonrecursive convolutional codes over independent Rayleigh fading channels, for arbitrary \( j \geq 0 \), the random sequences

\[
\{y_{k,0}, y_{k,1}, \ldots, y_{k,i}, \ldots\}, \quad \{\xi_{k,0}^{(j)}, \xi_{k,1}^{(j)}, \ldots, \xi_{k,i}^{(j)}, \ldots\}
\]

for all \( k \in \mathcal{M} \)

are GAL distributed and statistically independent in the sense that any finite collection of the \( y_{k,i} \) and \( \xi_{k,i}^{(j)} \) are GAL and pairwise independent. For arbitrary \( k \neq r \) with \( k, r \in \mathcal{M} \), the random sequences

\[
\{\xi_{k,0}^{(j)}, \xi_{k,1}^{(j)}, \ldots, \xi_{k,i}^{(j)}, \ldots\}, \quad \{\xi_{r,0}^{(l)}, \xi_{r,1}^{(l)}, \ldots, \xi_{r,i}^{(l)}, \ldots\}
\]

for all \( j, l \geq 0 \)

are GAL distributed and statistically independent.
Again, no independency is assumed between the extrinsic information at different iterations for the same bit and at the same node, i.e., $\xi_{k,i}^{(j)}$ and $\xi_{k,i}^{(l)}$ for $j \neq l$ may not be independent. Similar to the argument for AWGN channels, we also have that $\xi_{k,i}^{(j)}$ for all $k$ and $i$ are statistically identical for any given $j$. Specifically, $\xi_{k,i}^{(j)} \sim \mathcal{GAL}(\mu_j, \sigma_j^2, \tau_j)$ for all $k$ and $i$. Thus, the statistical behavior of all the extrinsic information generated at each iteration on a Rayleigh fading channel can be sufficiently described by the three parameters, $\mu_j$, $\sigma_j^2$ and $\tau_j$.

### 6.2 Density Evolution Model

With proper statistical approximations for extrinsic information, it is possible to quantify the statistical behavior of the MAP decoder with only a few sets of parameters in the iterative decoding process. Thus, evaluating the distribution parameters for the extrinsic information becomes the next necessary step for analyzing the decoding process. Since there is no analytic method available for finding the parameters for
collaborative decoding, the only approach we can use is to estimate the parameters based on the observations of extrinsic information from simulations.

This idea is similar to that of the density evolution technique proposed in [30] and [31] for turbo decoding and LDPC decoding. In this section, we will develop a density evolution model for the collaborative decoding process. Nevertheless, we need to point out that our goal here is very different from that in usual density evolution techniques. In the usual density evolution technique, decoder components are treated as some input-output functions of distribution parameters. The trajectories of the input-output functions are obtained by simulation. Then the convergence behavior of the iterative decoding process is studied based on these trajectories. However, due to the reason that usually only a few information exchanges (hence decoding iterations) can be performed in collaborative decoding, convergence properties are not of concern in our analysis. Hence, no input-output function and trajectories for the decoders will be explored here. Our goal is to provide an equivalent but simpler density evolution model for the real collaborative decoding process so that the distribution parameters for the extrinsic information can be obtained with a lower complexity.

Based on the statistical assumptions established in Section 6.1 and the symmetry of nodes in the collaborative deciding process, it is easy to see that the statistical parameters needed to describe of $\xi_{k,i}^{(j)}$, for all $k$ and $i$, can be found from any node in the distributed array, and the behavior of the additional information collected as each node are statistically equivalent. Also, the behavior of the MRB information exchange process can be determined with the knowledge of the statistical characteristics of $\xi_{k,i}^{(j)}$. Hence, we model the MRB information exchange as an additional information generation process with the statistical parameters that describe $\xi_{k,i}^{(j)}$ are the input and a sequence of additional information for a node are the output. Thus, the collaborative decoding process can be modeled by the joint operation of an additional information generating unit and the MAP decoder unit as shown in Fig. 6–3.
Figure 6–3: Density evolution model for collaborative decoding process

The output of the additional information generating unit is fed back to the MAP decoder as a priori information for the next iteration of decoding.

**6.2.1 Additional Information Generation**

We first consider the generation of the additional information in the $j$th decoding iteration. Recall that during the $j$th exchange, only when the reliability (i.e., $|\xi_{k,i}^{(j)}|$) for a candidate bit $i$ is ranked in the top $p_j$ fraction among the candidate bits set at node $k$, the extrinsic information $\xi_{k,i}^{(j)}$ can be broadcast to other nodes, as specified in the MRB exchange process. Let $B_i^{(j)}$ denote the event that bit $i$ is a candidate bit, then based on the i.i.d. assumption for $\xi_{k,i}^{(j)}$ for all $k$ and $i$ in Section 6.1, we can see that the events $B_m^{(j)}$ and $B_n^{(l)}$ for $m \neq n$ and arbitrary $j, l$ are independent.

To formalize the information exchange process, we further introduce the following assumption.

**Assumption 6.2.1** The MRB information exchange criterion for the candidate bit $i$ at node $k$ in the $j$th iteration is equivalent to $|\xi_{k,i}^{(j)}| \geq T_j$, for some $T_j \geq 0$ satisfying

$$
P(|\xi_{k,i}^{(j)}| \geq T_j | B_i^{(j)}) = p_j. \tag{6–6}
$$

Let $\lambda_{k,i}^{(j)}$ denote the additional information for the $i$th bit at the $k$th node generated by the MRB exchange scheme in the $j$th iteration. According to Section 4.2.3, $\lambda_{k,i}^{(j)}$ actually is the sum of extrinsic information collected from other nodes for bit $i$
in the $j$th iteration. This additional information will be added to the a priori information in the $(j + 1)$th iteration by node $k$. According to the MRB scheme, if bit $i$ is a non-candidate bit in the $j$th iteration, then $\lambda_{k,i}^{(j)} = 0$. Otherwise, there are three possibilities for the case of a candidate bit:

i) No node in the distributed array broadcasts information for the $i$th bit. According to Assumption 6.2.1, this corresponds to the event $\bigcap_{t \in \mathcal{M}} |\xi_{t,i}^{(j)}| < T_j$. Thus, we have $\lambda_{k,i}^{(j)} = 0$;

ii) Except node $k$, no any other node broadcasts information of bit $i$ in the MRB information exchange. We denote this event by $\mathcal{S}_{k,i}^{(j)}$. With Assumption 6.2.1, that is,

\[
\mathcal{S}_{k,i}^{(j)} = \left\{ |\xi_{k,i}^{(j)}| \geq T_j, \bigcap_{t \in \mathcal{M}'_k} |\xi_{t,i}^{(j)}| < T_j \right\},
\]

(6–7)

where $\mathcal{M}'_k$ is the the complementary set for node $k$, i.e., $\mathcal{M}'_k = \{r : r \in \mathcal{M}, r \neq k\}$. In this case, no information can be collected at node $k$ for bit $i$. Thus, we still have $\lambda_{k,i}^{(j)} = 0$.

iii) At least one node in $\mathcal{M}$, other than node $k$, broadcasts information for bit $i$. We denote this event by $\mathcal{R}_{k,i}^{(j)}$. To describe this event more precisely, we define $\mathcal{K}_{k,i}^{(j)}$ as a non-empty node set that contains all the nodes in $\mathcal{M}'_k$ broadcasting information for bit $i$ in the $j$th iteration. With Assumption 6.2.1, $\mathcal{K}_{k,i}^{(j)}$ can be formalized as

\[
\mathcal{K}_{k,i}^{(j)} = \left\{ t \in \mathcal{M}'_k : |\xi_{t,i}^{(j)}| \geq T_j \right\}.
\]

(6–8)

We also denote the event corresponding to $\mathcal{K}_{k,i}^{(j)}$ by $\mathcal{R}(\mathcal{K}_{k,i}^{(j)})$, i.e.,

\[
\mathcal{R}(\mathcal{K}_{k,i}^{(j)}) = \left\{ \bigcap_{t \in \mathcal{K}_{k,i}^{(j)}} |\xi_{t,i}^{(j)}| \geq T_j, \bigcap_{t \in \mathcal{K}_{k,i}^{(j)}} |\xi_{t,i}^{(j)}| < T_j \right\},
\]

(6–9)
where \( \overline{K}_{k,i}^{(j)} \) is the complementary set of \( K_{k,i}^{(j)} \) in \( \mathcal{M}_k' \). Hence, the event \( R_{k,i}^{(j)} \) is formed by the union of all possible events \( R(K_{k,i}^{(j)}) \), i.e.,

\[
R_{k,i}^{(j)} = \bigcup_{K_{k,i}^{(j)} \subset \mathcal{M}_k'} R(K_{k,i}^{(j)}). \tag{6–10}
\]

In this case, the additional information \( \lambda_{k,i}^{(j)} \) will depend on the set \( K_{k,i}^{(j)} \). For a particular event \( R(K_{k,i}^{(j)}) \), node \( k \) will obtain information from all nodes in \( K_{k,i}^{(j)} \).

We denote this additional information as \( \lambda(K_{k,i}^{(j)}) \), which is given by

\[
\lambda(K_{k,i}^{(j)}) = \sum_{t \in K_{k,i}^{(j)}} \xi^{(j)}_{t,i} \quad \text{with} \quad \bigcap_{t \in K_{k,i}^{(j)}} |\xi^{(j)}_{t,i}| \geq T_j. \tag{6–11}
\]

Since \( \{R(K_{k,i}^{(j)})\} \) are disjoint events, the additional information \( \lambda_{k,i}^{(j)} \) can be written as

\[
\lambda_{k,i}^{(j)} = \lambda(K_{k,i}^{(j)}). \tag{6–12}
\]

According to the MRB scheme, only in the case i) of all above three cases, the candidate bit \( i \) will still be a candidate bit in next iteration. Hence, we have

\[
P(B_{i}^{(j+1)}|B_{i}^{(j)}) = P\left(\bigcap_{k \in \mathcal{M}} |\xi^{(j)}_{k,i}| < T_j | B_{i}^{(j)} \right) = (1 - p_j)^M, \tag{6–13}
\]

where we have used (6–6) and the independent assumption for \( \xi^{(j)}_{k,i} \) for all \( k \) and \( i \).

From (6–13), we can easily obtain

\[
P(B_{i}^{(j)}) = P(B_{i}^{(j)}, B_{i}^{(j-1)}, \ldots, B_{i}^{(0)}) = \prod_{l=0}^{j-1} P(B_{i}^{(l+1)}|B_{i}^{(l)}) = \prod_{l=0}^{j-1} (1 - p_l)^M, \tag{6–14}
\]

and

\[
B_{i}^{(j)} = \bigcap_{l=0}^{j-1} \bigcap_{k \in \mathcal{M}} \{|\xi^{(l)}_{k,i}| < T_l\}. \tag{6–15}
\]

This corresponds to the fact that if no node broadcasts information for bit \( i \) in all iterations prior to the \( j \)th iteration, then bit \( i \) must be a candidate bit in the \( j \)th iteration. On other hand, only if event \( S_{k,i}^{(j)} \) or \( R_{k,i}^{(j)} \) occurs in the \( j \)th iteration for bit
candidate bit $i$, the bit will become a non-candidate bit in the next iteration. The probability of this event can be written as

$$P(B_i^{j+1} | B_i^{(j)}) = P(S_{k,i}^{(j)} | B_i^{(j)}) + P(R_{k,i}^{(j)} | B_i^{(j)})$$

$$= P(S_{k,i}^{(j)} | B_i^{(j)}) + \sum_{K^{(j)}_{k,i} \subseteq M'_k} P(R(K^{(j)}_{k,i}) | B_i^{(j)}), \quad (6–16)$$

where the fact that $S_{k,i}^{(j)}$ and $R_{k,i}^{(j)}$ are disjoint events is used.

In above, we have analyzed the additional information $\lambda_{k,i}^{(j)}$ for all $k \in M$ in the MRB exchange process. Now we consider the additional information generating in the density evolution model shown in Fig. 6–3. Due to the symmetry of nodes, the density evolution model includes only one MAP decoder. Without loss of generality, we assume that the MAP decoder is at the $M$th node. Thus, we only need to generate $\lambda_{M,i}^{(j)}$ in the model. Accordingly, we describe the additional information generating unit in the density evolution model as follows:

1. In the $j$th iteration, take $\{\xi_{M,i}^{(j)}\}$ for all $i$, and the statistical parameters ($T_j$ and $\{\mu_j, \sigma_j^2\}$ for AWGN channels or $\{\mu_j, \sigma_j^2, \tau_j\}$ for Rayleigh fading channels) as the input.

2. Set $\lambda_{M,i}^{(j)} = 0$ for all non-candidate bits.

3. For each candidate bit $i$, generate $M-1$ i.i.d. random variables $\xi_{k,i}^{(j)}$ for $k \in M'_M$ according to the conditional distribution of $\xi_{M,i}^{(j)}$ given that bit $i$ is a candidate bit. We obtain this conditional distribution empirically as described in Section 6.2.2 below.

4. Check events $S_{M,i}^{(j)}$ and $\{R(K_{M,i}^{(j)})\}$ according to (6–7) and (6–9). If $S_{M,i}^{(j)}$ or $R_{M,i}^{(j)}$ occurs, then set $\lambda_{M,i}^{(j)} = 0$ or apply (6–12), respectively, and flag bit $i$ as a non-candidate bit for the next iteration. Otherwise, set $\lambda_{M,i}^{(j)} = 0.$
6.2.2 Finding Parameters in Density Evolution Model

To simulate the collaborative decoding process with MRB exchange, the MAP decoder in the density evolution model will accumulate the additional information \( \lambda_{M,i}^{(l)} \) in each iteration, for \( l < j + 1 \), to form the a priori information for the \((j + 1)\)th decoding iteration. Let \( \eta_{M,i}^{(j+1)} \) denote the a priori information at node \( M \) in the \((j + 1)\)th iteration, then \( \eta_{M,i}^{(j+1)} = \sum_{l=0}^{j} \lambda_{M,i}^{(l)} \). Since a candidate bit \( i \) at node \( M \) can obtain additional information \( \lambda_{M,i}^{(l)} \) only if the event \( R_{M,i}^{(l)} \) occurs in a certain iteration \( l \), and will become a non-candidate bit after that, a data bit can obtain at most one copy of additional information at each node in the whole collaborative decoding process. Thus, the a priori information \( \eta_{M,i}^{(j+1)} \) can be rewritten as

\[
\eta_{M,i}^{(j+1)} = \begin{cases} 
    \lambda_{M,i}^{(l)} & \text{if } R_{M,i}^{(l)} \text{ occurs} \\
    0 & \text{otherwise,}
\end{cases}
\]  

(6–17)

With this a priori information \( \{ \eta_{M,i}^{(j+1)} \} \) and the intrinsic information from channel observation as input, we run the MAP decoder to generate the extrinsic information \( \{ \xi_{M,i}^{(j+1)} \} \) for the \((j + 1)\)th iteration.

Based on the observation of \( \{ \xi_{M,i}^{(j+1)} \} \), we can estimate their distribution parameters according to the distribution assumptions in Section 6.1. For AWGN channels, we have \( \xi_{M,i}^{(j+1)} \sim \mathcal{N}(\mu_{j+1}, \sigma_{j+1}^2) \). In this case, the parameters \( \mu_{j+1} \) and \( \sigma_{j+1}^2 \) are obtained by the sample mean and variance of \( \{ \xi_{M,i}^{(j+1)} \} \), respectively. For Rayleigh fading channels, we have \( \xi_{M,i}^{(j+1)} \sim \mathcal{GL}(\mu_{j+1}, \sigma_{j+1}^2, \tau_{j+1}) \). In this case, we use the method of moments to estimate \( \mu_{j+1} \), \( \sigma_{j+1}^2 \) and \( \tau_{j+1} \). For a random variable \( X \sim \mathcal{GL}(\mu, \sigma^2, \tau) \), we can use the standard methods, such as the method of moments and maximum likelihood estimation, to estimate the the parameters \( \mu, \sigma^2 \) and \( \tau \).

Besides the distribution parameters, we also need to estimate the threshold \( T_{j+1} \) defined in (6–6). Since the conditional probability in (6–6) is very complicated, analytic solution for \( T_{j+1} \) is not readily available. In order to estimate the value
of $T_{j+1}$, we first use nonparametric method to estimate the cumulative distribution function $F_{j+1}(x)$ of the extrinsic information for the candidate bits, i.e.,

$$F_{j+1}(x) = P(\xi_{M,i}^{(j+1)} \leq x | \mathcal{B}_i^{(j+1)}).$$  \hfill (6–18)

Then, according to (6–6) we have

$$p_{j+1} = 1 - F_{j+1}(T_{j+1}) + F_{j+1}(-T_{j+1}).$$  \hfill (6–19)

For $p_{j+1} < 1$, by solving (6–19) numerically we can obtain $T_{j+1}$ approximately. For the case of $p_{j+1} = 1$, we set $T_{j+1} = 0$.

By iterating the above procedure, we can obtain the necessary parameters of the extrinsic information in successive collaborative decoding iterations. In this density evolution model, we only need to simulate a single MAP decoder for learning the statistical behaviors of the actual collaborative decoding process with $M$ nodes. Hence, the simulation load is reduced to $1/M$ of that in the actual collaborative decoding process. Fig. 6–4 and 6–5 show the comparison of the mean and variance of the extrinsic information and the threshold $T_j$, respectively, estimated in our density evolution model and the actual collaborative decoding process for the case of $M = 6$.

The maximum free distance 4-state non-recursive convolution code is used, and \{\{p_j\}\} is set to \{0.1, 0.2, 0.8\}. From the figures, we see that our density evolution model gives a good approximation for the actual collaborative decoding process. Based on the density evolution model, we will evaluate the error performance for the $j$th iterations with the statistical knowledge of the extrinsic information in the first $j$ iterations.

6.3 A General Upper Bound for BER

In this section, we derive an upper bound for the BER of collaborative decoding with MRB exchange. Without loss of generality, we will consider the decoding process and performance at the $M$th node. According to the statistically identical assumption

\[100\]
Figure 6–4: Comparison of mean and variance of the extrinsic information obtained from the density evolution model and that from the actual collaborative decoding process.

Figure 6–5: Comparison of threshold estimated from the density evolution model and that from the actual collaborative decoding process.
for data bits, the error performance is the same for all data bits. Hence, we drop the bit index, i.e., the subscript $i$, in the notation of variables and events for the bit of interest. For convenience, we also drop the subscript $M$ for the $M$th node in following derivation.

Under the assumption that the all-zero data bit sequence is transmitted, the BER of the MAP decoder in the $j$th ($j \geq 1$) iteration is the probability that the soft output of a bit is negative. Since the soft output of a MAP decoder is the sum of its extrinsic information and a priori information, the BER is given as

$$ P_b^{(j)} = P(\xi^{(j)} + \eta^{(j)} < 0), \quad (6-20) $$

where $\xi^{(j)}$ is the extrinsic information, and $\eta^{(j)}$ is the a priori information in the $j$th iteration given in (6–17) at the $M$th node, respectively. Similar to the case of LRB exchange, we evaluate the error performance by finding an upper bound for (6–20).

According to (6–17) and the analysis in Section 6.2.1, the a priori information for a candidate bit must be zero, i.e.,

$$ P(\eta^{(j)} = 0 | \mathcal{B}^{(j)}) = 1. \quad (6–21) $$

For a non-candidate bit, we first consider its status transition from a candidate bit to a non-candidate bit. According to the MRB scheme, all data bits are initialized as candidate bits in the collaborative decoding process. Once a bit becomes a non-candidate bit in an iteration, its status will not be changed in later iterations, i.e.,

$$ P(\mathcal{B}^{(j+1)} | \mathcal{B}^{(j)}) = 1. $$

Thus, we have

$$ P(\mathcal{B}^{(j)}) = P(\mathcal{B}^{(j)}, \mathcal{B}^{(j-1)}) + P(\mathcal{B}^{(j-1)}). \quad (6–22) $$

With this recursive relation and (6–16), we can obtain

$$ P(\mathcal{B}^{(j)}) = \sum_{l=0}^{j-1} P(\mathcal{B}^{(l+1)}, \mathcal{B}^{(l)}) = \sum_{l=0}^{j-1} \left[ P(\mathcal{S}^{(l)}, \mathcal{B}^{(l)}) + P(\mathcal{R}^{(l)}, \mathcal{B}^{(l)}) \right]. \quad (6–23) $$
From (6–17), we know that only if the event $R^{(l)}$ occurs for certain $l < j$, the non-candidate bit of interest will have non-zero *a priori* information in the $j$th iteration. More specifically, we have

$$P(\eta^{(j)} = 0|S^{(l)}, B^{(l)}) = 1, \quad \text{for } l < j,$$  
(6–24)

and

$$P(\eta^{(j)} = \lambda^{(l)}|R^{(l)}, B^{(j)}) = 1, \quad \text{for } l < j.$$  
(6–25)

By using (6–21) through (6–25), we can expand $P_b^{(j)}$ in (6–20) as

$$P_b^{(j)} = P(\xi^{(j)} < 0, B^{(j)}) + \sum_{l=0}^{j-1} [P(\xi^{(j)} < 0, S^{(l)}, B^{(l)}) + P(\xi^{(j)} + \lambda^{(l)} < 0, R^{(l)}, B^{(l)})].$$  
(6–26)

When $j \geq 2$, we separate the sum of probabilities $P(\xi^{(j)} < 0, S^{(l)}, B^{(l)})$ for $0 \leq l \leq j-1$ in (6–26) into two parts: $P(\xi^{(j)} < 0, S^{(j-1)}, B^{(j-1)})$ the sum for $0 \leq l \leq j - 2$. That is,

$$P_b^{(j)} = P(\xi^{(j)} < 0, B^{(j)}) + P(\xi^{(j)} < 0, S^{(j-1)}, B^{(j-1)}) + \sum_{l=0}^{j-2} P(\xi^{(j)} < 0, S^{(l)}, B^{(l)}) + \sum_{l=0}^{j-1} P(\xi^{(j)} + \lambda^{(l)} < 0, R^{(l)}, B^{(l)}).$$  
(6–27)

In order to include the case of $j < 2$ into (6–27), we define

$$\sum_{l=0}^{j-2} P(\xi^{(j)} < 0, S^{(l)}, B^{(l)}) = 0, \quad \text{for } j < 2.$$  
(6–28)

With this definition, (6–27) will be valid for all $j > 0$. 

Now we consider the first two terms in (6–27). With (6–7) and (6–13), the first two terms in (6–27) can be written as

\[
P(\xi^{(j)} < 0, \mathcal{B}^{(j)}) + P(\xi^{(j)} < 0, \mathcal{S}^{(j-1)}, \mathcal{B}^{(j-1)})
= P\left(\xi^{(j)} < 0, \bigcap_{k \in M} |\xi^{(j-1)}_k| < T_j, \mathcal{B}^{(j-1)}\right)
+ P\left(\xi^{(j)} < 0, |\xi^{(j-1)}| \geq T_j, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \mathcal{B}^{(j-1)}\right)
= P\left(\xi^{(j)} < 0, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \mathcal{B}^{(j-1)}\right)
+ P\left(\xi^{(j)} < 0, |\xi^{(j-1)}| \geq T_j, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \mathcal{B}^{(j-1)}\right)
= P\left(\xi^{(j)} < 0, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \mathcal{B}^{(j-1)}\right). \tag{6–29}
\]

By inserting (6–15) into (6–29), and using Assumption 6.1.1 and Assumption 6.1.2, we have

\[
P(\xi^{(j)} < 0, \mathcal{B}^{(j)}) + P(\xi^{(j)} < 0, \mathcal{S}^{(j-1)}, \mathcal{B}^{(j-1)})
= P\left(\xi^{(j)} < 0, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \bigcap_{l=0}^{j-2} \bigcap_{k \in M} |\xi^{(l)}_k| < T_l\right)
= P\left(\xi^{(j)} < 0, \bigcap_{k \in M'} |\xi^{(j-1)}_k| < T_j, \bigcap_{l=0}^{j-2} |\xi^{(l)}| < T_l, \bigcap_{l=0}^{j-2} \bigcap_{k \in M'} |\xi^{(l)}_k| < T_l\right)
= P\left(\xi^{(j)} < 0, \bigcap_{l=0}^{j-2} |\xi^{(l)}| < T_l\right) \prod_{k \in M'} P\left(\bigcap_{l=0}^{j-1} |\xi^{(l)}_k| < T_l\right). \tag{6–30}
\]

To further evaluate (6–30), we introduce the following lemma

**Lemma 6.3.1** In collaborative decoding with the MRB exchange scheme, given the information exchange parameters \(\{p_l\}_{l=0}^j\) and the corresponding threshold \(\{T_l\}_{l=0}^j\), for all \(k \in M\) we have

\[
P\left(\bigcap_{l=0}^{j} |\xi^{(l)}_k| < T_l\right) = \prod_{l=0}^{j} (1 - p_l), \tag{6–31}
\]

and

\[
P\left(|\xi^{(j)}_k| \geq T_j, \bigcap_{l=0}^{j-1} |\xi^{(l)}_k| < T_l\right) = p_j \prod_{l=0}^{j-1} (1 - p_l). \tag{6–32}
\]
Proof Based on the definition in (6–15), the probability of the data bit being a candidate bit in the \((j + 1)\)th decoding iteration is given by

\[
P(B^{(j+1)}) = \prod_{k=1}^{M} P \left( \bigcap_{l=0}^{j} |\xi_k^{(l)}| < T_l \right) = \prod_{k=1}^{M} \left[ P \left( \bigcap_{l=0}^{j} |\xi_k^{(l)}| < T_l \right) \right]^{M}, \ k \in \mathcal{M},
\]  

(6–33)

where we have used Assumption 6.1.1 and Assumption 6.1.2. With (6–14), we have

\[
\left[ P \left( \bigcap_{l=0}^{j} |\xi_k^{(l)}| < T_l \right) \right]^{M} = \prod_{l=0}^{j} (1 - p_l)^{M}, \ k \in \mathcal{M}.
\]  

(6–34)

Thus, (6–31) can be obtained from (6–34). In the similar manner, (6–32) can be proved easily.

By applying Lemma 6.3.1 and dropping the event \(\bigcap_{l=0}^{j-2} |\xi(l)| < T_l\) in (6–30), we obtain following upper bound

\[
P(\xi^{(j)} < 0, B^{(j)}) + P(\xi^{(j)} < 0, S^{(j-1)}, B^{(j-1)}) \leq P(\xi^{(j)} < 0) \prod_{l=0}^{j-1} (1 - p_l)^{M-1}.
\]  

(6–35)

In the same way, we consider the third term in (6–27). Again, with (6–7), (6–13) and the independency assumption for extrinsic information, we have

\[
P(\xi^{(j)} < 0, S^{(l)}, B^{(l)}) = P\left(\xi^{(j)} < 0, |\xi(l)| \geq T_l, \bigcap_{k \in \mathcal{M}'} |\xi_k^{(l)}| < T_l, \bigcap_{k \in \mathcal{M}} |\xi_k^{(l)}| < T_l \right)
\]

\[
= P\left(\xi^{(j)} < 0, |\xi(l)| \geq T_l, \bigcap_{k \in \mathcal{M}'} |\xi_k^{(l)}| < T_l \right) \prod_{k \in \mathcal{M}'} P\left(\bigcap_{l=0}^{l-1} |\xi_k^{(l)}| < T_l \right).
\]  

(6–36)

In the above, the second term is given by (6–31) in Lemma 6.3.1. For the first term, we upper bound it as follows:

\[
P\left(\xi^{(j)} < 0, |\xi(l)| \geq T_l, \bigcap_{l=0}^{l-1} |\xi_k^{(l)}| < T_l \right) \leq P(\xi^{(j)} < 0, |\xi(l)| \geq T_l)
\]

\[
= P(\xi^{(j)} < 0, \xi(l) < -T_l) + P(\xi^{(j)} < 0, \xi(l) \geq T_l)
\]

\[
\leq P(\xi(l) < -T_l) + P(\xi^{(j)} < 0, \xi(l) \geq T_l).
\]  

(6–37)
Thus, the probability $P(\xi^{(j)} < 0, S^{(l)}, B^{(l)})$ for $0 \leq l \leq j - 2$ can be upper bounded as

$$P(\xi^{(j)} < 0, S^{(l)}, B^{(l)}) \leq \left[ P(\xi^{(l)} < -T_l) + P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \right] \prod_{t=0}^{l} (1 - p_t)^{M-1}. \quad (6–38)$$

Now, we upper bound the probabilities in the last term of (6–27). By using (6–9) through (6–12), we have

$$P(\xi^{(j)} + \lambda^{(l)} < 0, R^{(l)}, B^{(l)}) = \sum_{\mathcal{K}^{(l)} \subset \mathcal{M}'} P(\xi^{(j)} + \lambda(\mathcal{K}^{(l)}) < 0, R(\mathcal{K}^{(l)}), B^{(l)}). \quad (6–39)$$

According to Assumption 6.1.1 and Assumption 6.1.2, extrinsic information generated by different nodes for the same bit in an iteration is i.i.d.. Then, the statistical characteristics of $\lambda(\mathcal{K}^{(l)})$ and probability of event $R(\mathcal{K}^{(l)})$ depend only on the cardinal size of $\mathcal{K}^{(l)}, |\mathcal{K}^{(l)}|$. Hence, for all sets $\mathcal{K}^{(l)}$ with the same cardinal size, the probabilities $P(\xi^{(j)} + \lambda(\mathcal{K}^{(l)}) < 0, R(\mathcal{K}^{(l)}), B^{(l)})$ are equal.

For convenience, we denote the cardinal size of $\mathcal{K}^{(l)}$ by $K^{(l)}$, i.e., $K^{(l)} = |\mathcal{K}^{(l)}|$. Then the total number of possible choices for subset $\mathcal{K}^{(l)}$ in $\mathcal{M}'$ is $\binom{M-1}{K^{(l)}}$. Without loss of generality, we only need to consider the case of $\mathcal{K}^{(l)} = \{1, 2, \cdots, K^{(l)}\}$. In this case, we will use the notations $\lambda(K^{(l)})$ and $R(K^{(l)})$ to replace $\lambda(\mathcal{K}^{(l)})$ and $R(\mathcal{K}^{(l)})$ for $\mathcal{K}^{(l)} = \{1, 2, \cdots, K^{(l)}\}$, respectively. With the above arguments, we have

$$\sum_{\mathcal{K}^{(l)} \subset \mathcal{M}'} P(\xi^{(j)} + \lambda(\mathcal{K}^{(l)}) < 0, R(\mathcal{K}^{(l)}), B^{(l)})$$

$$= \sum_{K^{(l)}=1}^{M-1} \binom{M-1}{K^{(l)}} P(\xi^{(j)} + \lambda(K^{(l)}) < 0, R(K^{(l)}), B^{(l)}). \quad (6–40)$$
By inserting the definition of \( \lambda(K^{(l)}) \), \( \mathcal{R}(K^{(l)}) \) and \( \mathcal{B}(l) \), we expand and upper bound the probability \( P(\xi^{(j)} + \lambda(K^{(l)}) < 0, \mathcal{R}(K^{(l)}), \mathcal{B}(l)) \) in (6–40) as follows:

\[
P(\xi^{(j)} + \lambda(K^{(l)}) < 0, \mathcal{R}(K^{(l)}), \mathcal{B}(l)) \\
= P\left(\xi^{(j)} + \lambda(K^{(l)}) < 0, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| \geq T_l, \bigcap_{k=K^{(l)}+1}^{M-1} |\xi_k^{(l)}| < T_l, \bigcap_{t=0}^{l-1} \bigcap_{k \in \mathcal{M}} |\xi_k^{(t)}| < T_t\right) \\
= P\left(\xi^{(j)} + \lambda(K^{(l)}) < 0, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| < T_t, \bigcap_{k=1}^{K^{(l)}} \left\{ |\xi_k^{(t)}| \geq T_l, \bigcap_{t=0}^{l-1} |\xi_k^{(t)}| < T_t \right\}\right) \quad (6–41) \\
\quad \cdot \prod_{k=K^{(l)}+1}^{M-1} P\left(\bigcap_{t=0}^{l-1} |\xi_k^{(t)}| < T_t\right) \\
\leq P\left(\xi^{(j)} + \sum_{k=1}^{K^{(l)}} \xi_k^{(l)} < 0, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| \geq T_l\right) \prod_{t=0}^{l-1} (1 - p_t)^{M-K^{(l)}-1}, \quad (6–42)
\]

where we obtain the inequality by dropping the events \( \{\bigcap_{t=0}^{l-1} |\xi_k^{(t)}| < T_t\} \) and \( \{\bigcap_{k=1}^{K^{(l)}} \bigcap_{t=0}^{l-1} |\xi_k^{(t)}| < T_t\} \) in (6–41), and calculate the probability \( P\left(\bigcap_{t=0}^{l-1} |\xi_k^{(t)}| < T_t\right) \) using Lemma 6.3.1.

With (6–39) through (6–42), we can obtain an upper bound for the probability \( P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) \) for \( 0 \leq l \leq j - 1 \) in (6–27) as

\[
P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) \\
\leq \sum_{K^{(l)}=1}^{M-1} P\left(\xi^{(j)} + \sum_{k=1}^{K^{(l)}} \xi_k^{(l)} < 0, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| \geq T_l\right) \left[\left(\frac{M-1}{K^{(l)}}\right) \prod_{t=0}^{l-1} (1 - p_t)^{M-K^{(l)}-1}\right] \quad (6–43)
\]

To this point, we have upper bounded all the probabilities in (6–27). By substituting (6–35), (6–38) and (6–42) into (6–27), an general upper bound of the BER \( P_b^{(j)} \) can be obtained. We summarize the above results in the following theorem.

**Theorem 6.3.2** For collaborative decoding with MRB exchange and nonrecursive convolutional codes, given the number of nodes \( M \), the information exchange percentage parameters \( \{p_l\}_{l=0}^{j-1} \) and their corresponding threshold \( \{T_t\}_{l=0}^{j-1} \), the BER at an
arbitrary node in the \(j\)th decoding iteration, for \(j \geq 1\), is upper bounded as

\[
P_{b}^{(j)} \leq a_{j-1}^{M-1} P(\xi^{(j)} < 0) + \sum_{l=0}^{j-2} a_l^{M-1} \left[ P(\xi^{(l)} < -T_l) + P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \right] \\
+ \sum_{l=0}^{j-1} \sum_{K^{(l)}=1}^{M-1} b_{l,K^{(l)}} P\left(\xi^{(j)} + \sum_{k=1}^{K^{(l)}} \xi_k^{(l)} < 0, \bigcap_{k=1}^{K^{(l)}} \abs{\xi_k^{(l)}} \geq T_l\right),
\]

(6–44)

where

\[
a_l = \begin{cases} 
\prod_{t=0}^{l}(1 - p_t) & l \geq 0 \\
1 & l < 0
\end{cases},
\]

(6–45)

\[
b_{l,K} = \left(\frac{M-1}{K}\right) a_{l-1}^{M-K-1},
\]

(6–46)

and \(\{\xi_k^{(l)}\}\) are the extrinsic information generated by all nodes in the \(l\)th iteration, under the assumption that the all-zero codeword is transmitted.

Note that, according to the definition in (6–28), the summation \(\sum_{l=0}^{j-2}(\cdot)\) in (6–44) equals 0 for \(j < 2\). Since the probability distribution assumptions for the extrinsic information are not used in the derivation, (6–44) is a general bound valid for both of AWGN channels and Rayleigh fading channels.

For the special case of only one exchange and \(p_0 = 1\), i.e., all nodes broadcast the extrinsic information for all data bits in the first (0th) decoding iteration, we have the following corollary.

**Corollary 6.3.3** In collaborative decoding with MRB exchange and nonrecursive convolutional codes, if \(p_0 = 1\), then the final BER is upper bounded by

\[
P_b \leq P\left(\xi^{(1)} + \sum_{k=1}^{M-1} \xi_k^{(0)} < 0\right).
\]

(6–47)

**Proof** Since \(p_0 = 1\), all data bits will become non-candidate bits immediately after the first information exchange. Thus, only 1 exchange (i.e., 2 decoding iterations) can be performed in the collaborative decoding process, i.e., \(P_b^{(1)}\) is the final BER. With \(p_0 = 1\), we know \(a_0 = 0\) and \(b_{0,K^{(0)}} = 0\), for \(K^{(0)} \neq M - 1\), and \(b_{0,K^{(0)}} = 1\), for \(K^{(0)} = M - 1\), from (6–45) and (6–46). According to (6–19), \(p_0 = 1\) yields \(T_0 = 0\)
regardless the distribution of $\xi_k^{(0)}$. Hence, $P(\bigcap_{k=1}^{K(0)} |\xi_k^{(0)}| \geq T_0) = 1$. Applying these results to (6–44) for $j = 1$, we obtain the corollary.

In fact, for this special case, we can easily see that the a priori information $\eta^{(1)} = \sum_{k=1}^{M-1} \xi_k^{(0)}$. Inserting $\eta^{(1)}$ into (6–20) directly yields $P_{b}^{(1)}(\xi^{(1)} + \sum_{k=1}^{M-1} \xi_k^{(0)} < 0)$, which is consistent with Corollary 6.3.3. The consistency verifies the validity and tightness of the bounds in Theorem 6.3.2 for this simple case.

### 6.4 Error Events and Probabilities Analysis

Theorem 6.3.2 gives an upper bound of the BER $P_{b}^{(j)}$ in terms of probabilities involving the extrinsic information generated in the $j$th and all previous iterations of the collaborative decoding process. In order to evaluate the upper bound, it is necessary to further evaluate those probabilities using the statistical assumptions in Section 6.1 for the extrinsic information for AWGN channels and Rayleigh fading channels, respectively.

For this purpose, we will study the decoding process with MRB information exchange from the viewpoint of sequential decoding error event behavior due to the following reasons. Firstly, as mentioned previously, our goal is to evaluate the BER performance in the $j$th iteration by only using the statistical knowledge of extrinsic information up to the $(j-1)$th iteration. This allows us to predict the collaborative decoding performance with less knowledge about the extrinsic information. Thus, we need to express the probabilities involving $\xi^{(j)}$ in (6–44) in terms of $\xi^{(l)}$ for $l < j$. To unveil this relation, an effective way is to analyze the error events in the decoding procedure. Secondly, the probabilities $P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l)$ for $l < j - 1$ in (6.3.2) involve the extrinsic information for the same data bit generated by the same node in different decoding iterations. Since collaborative decoding with MRB information exchange is a memory-based approach (the a priori information is accumulated for all decoding iterations and the decoding and information exchange processes depend on all previous iterations), $\xi^{(j)}$ and $\xi^{(l)}$ are correlated. This factor is also reflected in
the statistical assumption for the extrinsic information in Section 6.1. Thus, in order to evaluation of $P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l)$, analysis of the decoding process in the context of MRB information exchange is necessary.

In the following, we study the probabilities involving $\xi^{(j)}$ based on the generalized union bound for max-log-MAP decoding and the error events analysis with MRB information exchange.

### 6.4.1 Union Bound for Collaborative Decoding

Following the notation in Section 5.2.2, we denote a nonrecursive convolutional code as a binary sequence mapping $C: \mathbf{u} \rightarrow \mathbf{c}$, where $\mathbf{u}$ and $\mathbf{c}$ are data bit sequence and the corresponding codeword. Let $\mathbf{u} = (u_0, u_1, \ldots, u_i, \ldots)$ and $\mathbf{c} = (c_0, c_1, \ldots, c_i, \ldots)$, then $u_i$ and $c_i \in \{0, 1\}$ are the data bit and coded bit, respectively. Consistent with the definition in (4–1), we denote $y_i$ the received BPSK (i.e., $x_i = 1 - 2c_i$ in (4–1)) signal at the decoder of interest. Again, throughout the following analysis, we assume that the all-zero sequence is transmitted. Then, we apply the generalized union bound in Section 5.2.2 to the case of collaborative decoding with the MRB exchange scheme. That is,

$$P(\xi^{(j)} < x) \leq \frac{1}{K_c} \sum_{d \geq d_{\text{min}}} \sum_{w \geq 1} w A_{w,d} P(\Gamma_{w,d}^{(j)} < x),$$  \quad (6–48)

where $x$ is an arbitrary constant, $d_{\text{min}}$ is the minimum Hamming distance of the code $C$, $A_{w,d}$ is the number of error events with Hamming weight $d$ and input weight $w$, and $\Gamma_{w,d}^{(j)}$ is the metric for an arbitrary error event $(\mathbf{u}, \mathbf{c})$ with input weight $w$ and output weight $d$ in the $j$th decoding iteration, which is given by

$$\Gamma_{w,d}^{(j)} = \sum_{i=1}^{w-1} \eta_i^{(j)} + L_c \sum_{i=0}^{d-1} g_i y_i,$$  \quad (6–49)

with $L_c$ defined in (5–30). According to Section 5.2.2, the extrinsic information $\xi^{(j)}$ in (6–48) is for the first bit of $\mathbf{u}$, i.e., $u_0$. Since the statistical property for $\Gamma_{w,d}^{(j)}$ and the probabilities in (6–49) can be regarded as independent of the error events starting
position for large coding block size [14, 34], the union bound is valid for arbitrary data bit. Thus, we have dropped the subscript 0 in \( \xi_0^{(j)} \) for clarity. Also, we have indexed the other \( w - 1 \) non-zero bits \( u_i \) in \( \mathbf{u} \) as \( i = 1, 2, \cdots, w - 1 \) in (6–49) without loss of generality.

In order to apply the union bound (6–48) to our analysis, we need to evaluate the PEP \( P(\Gamma^{(j)}_{w,d} < x) \). For this purpose, we consider the error event metric \( \Gamma^{(j)}_{w,d} \) defined in (6–49). We know that for an error event \((\mathbf{u}, \mathbf{c})\) with Hamming weights \( w \) and \( d \), \( \Gamma^{(j)}_{w,d} \) actually is the sum of \textit{a priori} information for the \( w - 1 \) non-zero data bits in sequence \( \mathbf{u} \) (the first non-zero bit \( u_0 \) itself is excluded) and the intrinsic information for codeword \( \mathbf{c} \). Since the \textit{a priori} information \( \eta_i^{(j)} \) is the sum of random copies of truncated independent extrinsic information from other nodes according to (6–17), the distribution of \( \Gamma^{(j)}_{w,d} \) is generally intractable. Thus, direct evaluation of \( P(\Gamma_{w,d}^{(j)} < x) \) is impossible. In order to make the evaluation of PEP tractable, we will decompose the error event corresponding to \( \Gamma^{(j)}_{w,d} \) into union of simpler events based on the analysis of MRB information exchange process on error events.

### 6.4.2 Analysis for MRB Information Exchange on Error Events

We consider the \( w - 1 \) non-zeros data bit sequence \( \{u_i\}_{i=1}^{w-1} \) in \( \mathbf{u} \). According to the analysis in Section 6.1, not every non-zeros data bit \( u_i \) can obtain \textit{a priori} information in the \( j \)th decoding iteration. This depends on the information exchange process in the previous iterations. The information exchange in each iteration does not only determine whether a bit obtains additional information, but also its status in later iterations. Specifically, we use \( A_l \) to denote the bit set in \( \{u_i\}_{i=1}^{w-1} \) for which the information exchange process occurs in the \( l \)th iteration. According to Section 6.2.1, information exchange for bit \( i \) in the \( l \)th iteration means that either \( \mathcal{S}_i^{(l)} \) or \( \mathcal{R}(\mathcal{K}_i^{(l)}) \) for certain \( \mathcal{K}_i^{(l)} \) occurs, and bit \( i \) must be a candidate bit in the \( l \)th iteration. Thus, \( A_l \) can be written as

\[
A_l = \{ i : 1 \leq i \leq w - 1, \mathcal{S}_i^{(l)} \cup \mathcal{R}(\mathcal{K}_i^{(l)}), \mathcal{B}_i^{(l)} \}, \tag{6–50}
\]
where we do not distinguish the difference between bit and bit index for convenience.

Further, we define \( \dot{A}_l \) as a subset of \( A_l \) that the event \( S^{(l)}_i \) occurs for \( i \in \dot{A}_l \), and \( \ddot{A}_l \) as the complementary subset of \( \dot{A}_l \) in \( A_l \) that the event \( \mathcal{R}(\mathcal{K}^{(l)}_i) \) occurs for \( i \in \ddot{A}_l \), i.e.,

\[
\dot{A}_l = \{ i : 1 \leq i \leq w - 1, S^{(l)}_i, B^{(l)}_i \}, \tag{6–51}
\]
and

\[
\ddot{A}_l = \{ i : 1 \leq i \leq w - 1, \mathcal{R}(\mathcal{K}^{(l)}_i), B^{(l)}_i \}. \tag{6–52}
\]

With these definitions, we can see that in \( \{ u_i \}_{i=1}^{w-1} \), only the bits for \( i \in \ddot{A}_l \) can obtain the additional information \( \lambda(\mathcal{K}^{(l)}_i) \) in the \( l \)th iteration. Since all bits in \( A_l \) become non-candidate in later iterations, we have \( A_l \cap A_k = \emptyset \) for \( l \neq k \). Thus, in the \( j \)th iteration, \( \bigcup_{l=0}^{j-1} A_l \) forms the complete non-candidate bit set in \( \{ u_i \}_{i=1}^{w-1} \). The remaining bits form the candidate bit set in that iteration. We denote this set by \( B_j \), which is given by

\[
B_j = \bigcup_{l=0}^{j-1} A_l. \tag{6–53}
\]

Based on the above definitions, we formalize the corresponding information exchange event for the sequence \( \{ u_i \}_{i=1}^{w-1} \) during the first \( j \) exchanges as

\[
\mathcal{W}_j = \left\{ \bigcap_{l=0}^{j-1} \left\{ \bigcap_{i \in \dot{A}_l} S^{(l)}_i, \bigcap_{i \in \ddot{A}_l} \mathcal{R}(\mathcal{K}^{(l)}_i), \bigcap_{i \in A_l} B^{(l)}_i \right\}, \bigcap_{i \in B_j} B^{(j)}_i \right\}. \tag{6–54}
\]

With (6–12), (6–17) and (6–49), the error event metric associated with the information exchange event \( \mathcal{W}_j \) can be written as

\[
\Gamma_{w,d}(\mathcal{W}_j) = \sum_{l=0}^{j-1} \sum_{i \in \dot{A}_l} \lambda(\mathcal{K}^{(l)}_i) + Y_d = \sum_{l=0}^{j-1} \sum_{i \in \ddot{A}_l} \sum_{k \in \mathcal{K}^{(l)}_i} \xi^{(l)}_{k,i} + Y_d, \tag{6–55}
\]

where

\[
Y_d = L_c \sum_{i=0}^{d-1} g_i y_i. \tag{6–56}
\]
Different choices of sets \( \{A_t, \hat{A}_t, \tilde{A}_t\} \) and \( \{K_i^{(l)}\}_{i \in \hat{A}_t} \) for \( l < j \) give different disjoint event \( W_j \). Moreover, with all possible choices of these sets, the union of these events completely describes the information exchange process for \( \{u_{i}^{w-1}\}_{i=1} \) in the first \( j \) iterations. Thus, we can expand the PEP \( P(\Gamma_{w,d}^{(j)} < x) \) as

\[
P(\Gamma_{w,d}^{(j)} < x) = \sum_{W_j} P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j),
\]

(6–57)

where \( P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j) \) is the joint PEP associated with the information exchange event \( W_j \). For this joint PEP with \( W_j \) we have the following lemma.

**Lemma 6.4.1** In collaborative decoding with MRB exchange and parameters \( \{p_l, T_i\}_{i=0}^{j-1} \), for the information exchange event \( W_j \) defined in (6–54) the joint PEP \( P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j) \) is upper bounded by

\[
P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j) \leq c_j P(\Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_t} \bigcap_{k \in K_i^{(l)}} |\xi_{k,i}^{(l)}| \geq T_i),
\]

(6–58)

where

\[
c_j = \prod_{l=0}^{j-1} \left[ p_{l}^{\hat{A}_t} (1 - p_l)^{|A_l| (M-1) + |B_j| M - \sum_{i \in \hat{A}_t} |K_i^{(l)}|} \prod_{l=0}^{j-1} (1 - p_l)^{|A_l| M - |\hat{A}_t| - \sum_{i \in \hat{A}_t} |K_i^{(l)}|} \right].
\]

(6–59)

**Proof** By inserting (6–54) into \( P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j) \), we have

\[
P(\Gamma_{w,d}^{(j)}(W_j) < x, W_j) = P(\Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_t} \bigcap_{i \in \hat{A}_l} \bigcap_{i \in A_l} \bigcap_{i \in B_j} \{ S_i^{(l)}, B_i^{(l)}, B_i^{(j)} \})
\]

\[
= P(\Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \{ R(K_i^{(l)}, B_i^{(l)}) \}) P(\bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \{ S_i^{(l)}, B_i^{(l)} \}) P(\bigcap_{i \in B_j} B_i^{(j)}),
\]

(6–60)

where we have used the fact that events associated with \( \hat{A}_t, \tilde{A}_t \) and \( B_j \) are independent. With the definition of events \( S_i^{(l)} \) and \( B_i^{(l)} \), we first calculate the probabilities

\[
\]
\[ P \left( \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \{ S^{(l)}_i, B^{(l)}_i \} \right) \quad \text{and} \quad P \left( \bigcap_{i \in B_j} B^{(j)}_i \right) \] in (6–60) as follows.

\[
P \left( \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \{ S^{(l)}_i, B^{(l)}_i \} \right) = \prod_{l=0}^{j-1} \prod_{i \in \hat{A}_l} P \left( S^{(l)}_i, B^{(l)}_i \right) \]

\[
= \prod_{l=0}^{j-1} \prod_{i \in \hat{A}_l} P \left( \lvert \xi_{k,M,i}^{(l)} \rvert \geq T_l, \bigcap_{k \in \mathcal{M}} \bigcap_{t=0}^{l-1} \lvert \xi_{k,i}^{(l)} \rvert < T_l, \bigcap_{k \in \mathcal{M}} \bigcap_{t=0}^{l-1} \lvert \xi_{k,i}^{(l)} \rvert < T_l \right) \]

\[
= \prod_{l=0}^{j-1} \prod_{i \in \hat{A}_l} \left[ p_l (1 - p_l)^{M-1} \prod_{t=0}^{l-1} (1 - p_t)^{M} \right]^{\lvert \hat{A}_l \rvert}, \quad (6–61) \]

where we have used Lemma 6.3.1, and \( \lvert \hat{A}_l \rvert \) means the cardinal size of \( \hat{A}_l \). Similarly,

\[
P \left( \bigcap_{i \in B_j} B^{(j)}_i \right) = \prod_{i \in B_j} P \left( B^{(j)}_i \right) = \prod_{i \in B_j} (1 - p_t)^{M} = \left[ \prod_{l=0}^{j-1} (1 - p_t)^{M} \right]^{\lvert B_j \rvert}, \quad (6–62) \]

For the first probability in (6–60), by inserting (6–9), (6–15) and (6–55) we have

\[
P \left( \Gamma^{(j)}_{w,d}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \{ R(K^{(l)}_i), B^{(l)}_i \} \right) \]

\[
= P \left( \sum_{l=0}^{j-1} \sum_{i \in \hat{A}_l} \sum_{k \in K^{(l)}_i} \xi_{k,i}^{(l)} + Y_d < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \bigcap_{k \in K^{(l)}_i} \bigcap_{t=0}^{l-1} \lvert \xi_{k,i}^{(l,t)} \rvert < T_l \right) \]

\[
= P \left( \sum_{l=0}^{j-1} \sum_{i \in \hat{A}_l} \sum_{k \in K^{(l)}_i} \xi_{k,i}^{(l)} + Y_d < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \bigcap_{k \in K^{(l)}_i} \bigcap_{t=0}^{l-1} \lvert \xi_{k,i}^{(l,t)} \rvert < T_l \right) \cdot P \left( \bigcap_{l=0}^{j-1} \bigcap_{i \in \hat{A}_l} \bigcap_{k \in K^{(l)}_i} \bigcap_{t=0}^{l-1} \lvert \xi_{k,i}^{(l,t)} \rvert < T_l \right) \]
\[ \leq P\left(\sum_{l=0}^{j-1} \sum_{i \in \tilde{A}_l, k \in K_i} \xi_{k,i}^{(l)} + Y_d < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \tilde{A}_l, k \in K_i} |\xi_{k,i}^{(l)}| \geq T_l\right) \]

\[ \cdot \prod_{l=0}^{j-1} \left[ \prod_{t=0}^{l} (1 - p_t) \right]^{|\tilde{A}_l|(M - 1) - \sum_{i \in \tilde{A}_l} |K_i^{(t)}|}, \]

(6–63)

where the calculation of probability \( P\left(\bigcap_{l=0}^{j-1} \bigcap_{i \in \tilde{A}_l, k \in K_i} \sum_{l=0}^{j-1} \sum_{i \in \tilde{A}_l, k \in K_i} |\xi_{k,i}^{(l)}| < T_l\right) \) is similar to (6–62), and the fact \(|K_i^{(t)}| = M - |K_i^{(t)}| - 1\) is used.

Finally, by combining (6–60) through (6–63) and using the fact \(|\tilde{A}_l| + |\tilde{A}_l| = |A_l|\),

we can obtain upper bound in (6–58) and (6–59).

**Corollary 6.4.2** For a given sequence \( \{u_i\}_{i=1}^{w-1} \), the upper bound for the joint PEP with event \( W_j \) in Lemma 6.4.1 does not depend on the particular choices of sets \( \{\tilde{A}_l\}, \{\tilde{A}_l\} \) and \( \{K_i^{(t)}\}_{i \in \tilde{A}_l} \), for \( l < j \), that associate with \( W_j \), but only on their cardinal sizes \(|\tilde{A}_l|, |\tilde{A}_l|\) and \( \sum_{i \in \tilde{A}_l} |K_i^{(t)}| \) for \( l < j \).

**Proof** With the fact that \(|A_l| = |\tilde{A}_l| + |\tilde{A}_l| \) and \(|B_j| = w - 1 - \sum_{i=0}^{j-1} |A_l|\), it is obvious that the coefficient \( c_j \) in (6–59) depends only on the cardinal sizes \(|\tilde{A}_l|, |\tilde{A}_l|\) and \( \sum_{i \in \tilde{A}_l} |K_i^{(t)}| \) associated with \( W_j \), given \( \{p_t\}_{t=0}^{j-1} \). Thus, we only need to consider the probability \( P\left(\Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \tilde{A}_l, k \in K_i^{(t)}} |\xi_{k,i}^{(l)}| \geq T_l\right) \) in the bound.

Due to the fact that \( \tilde{A}_l \cap \tilde{A}_k = \emptyset \) for \( l \neq k \), with (6–55) we can see that the intrinsic information \( Y_d \) and all the extrinsic information at different iterations involved in the event \( \{\Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i \in \tilde{A}_l, k \in K_i^{(t)}} |\xi_{k,i}^{(l)}| \geq T_l\} \) are independent according to Assumption 6.1.1 and Assumption 6.1.2. Moreover, we know that in each iteration \( l \), the extrinsic information \( \{\xi_{k,i}^{(l)}\} \) for all \( k \) and \( i \) are i.i.d. random variables. Thus, the statistical characteristics of the event only depends on the number of random variables, i.e., \( \sum_{i \in \tilde{A}_l} |K_i^{(t)}| \), for each \( l \). Hence, Corollary 6.4.2 is proved.

Based on Corollary 6.4.2, it is possible to characterize the upper bound in Lemma 6.4.1 with the cardinal sizes of sets \( \{\tilde{A}_l\}, \{\tilde{A}_l\} \) and \( \{K_i^{(t)}\}_{i \in \tilde{A}_l} \), for \( l < j \,
that associate with $\mathcal{W}_j$. For this purpose, let
\[ m_l = |A_l|, \quad \text{and} \quad n_l = |\tilde{A}_l| \quad (6-64) \]
with $0 \leq n_l \leq m_l$. Since $S_i^{(l)}$ and $R_i^{(l)}$ are disjoint events, we know that $\hat{A}_l \cap \tilde{A}_l = \emptyset$, and $A_l = \hat{A}_l \cup \tilde{A}_l$. Hence,
\[ |\hat{A}_l| = m_l - n_l, \quad (6-65) \]
which is sufficiently determined by (6-64). For convenience, we denote the number of candidate bits in $\{u_i\}_{i=1}^{w-1}$ in the $l$th decoding iteration by $w_l$. Since the number of candidate bits removed after the $l$th iteration is $m_l$, we have
\[ w_{l+1} = w_l - m_l, \quad \text{with} \quad w_0 = w - 1, \quad (6-66) \]
and
\[ 0 \leq m_l \leq w_l, \quad \text{for all} \quad l \geq 0. \quad (6-67) \]
With (6-66), the cardinal size of $B_j$ defined in (6-53) is given by
\[ |B_j| = w_j. \quad (6-68) \]
We also define $\nu_l$ as the total number of copies of extrinsic information collected for sequence $\{u_i\}_{i=1}^{w-1}$ in the $l$th information exchange, i.e.,
\[ \nu_l = \sum_{i \in \tilde{A}_l} |K_i^{(l)}|. \quad (6-69) \]
Since $1 \leq |K_i^{(l)}| \leq M - 1$, we have
\[ n_l \leq \nu_l \leq n_l(M - 1). \quad (6-70) \]
Applying the above notation to the upper bound in Lemma 6.4.1, the coefficient $c_j$ in (6–59) can be written as

$$c_j = \prod_{l=0}^{j-1} \left[ p_l^{m_l-n_l} (1 - p_l)^{m_l(M-1)+w_jM-\nu_l} \prod_{l=0}^{l-1} (1 - p_l)^{m_lM - \nu_l} \right].$$

(6–71)

According to Corollary 6.4.2, given the distribution of \( \{\xi^{(l)}_{k,i}\} \) and threshold \( T_l \) for \( l < j \), the probability \( P\left( \Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i\in\tilde{A}_l} \bigcap_{k\in K_l^{(i)}} |\xi^{(l)}_{k,i}| \geq T_l \right) \) in (6–58) is determined by the parameters \( \{\nu_l\}_{l=0}^{j-1} \). For convenience, we will use \( V_j \) to denote this \( j \)-tuple hereafter, i.e.,

$$V_j = \{\nu_l\}_{l=0}^{j-1}.$$  

(6–72)

To emphasize the above independence on the particular choices for sets \( \{\tilde{A}_l, \hat{A}_l\} \) and \( \{K_l^{(i)}\}_{i\in\hat{A}_l} \) in \( W_j \), we introduce the following new notations to represent the probability.

**Definition** For each \( l \in \{0, 1, \cdots, j - 1\} \), define a class of i.i.d. random variables \( \tilde{\xi}_i^{(l)} \) for \( i = 1, 2, \cdots \) such that \( \{\tilde{\xi}_i^{(l)}\} \) and \( \{\xi^{(l)}_{k,i}\} \) have the same distribution, denoted as \( \tilde{\xi}_i^{(l)} \sim \xi^{(l)}_{k,i} \). Meanwhile, we force the random sequence \( \{\tilde{\xi}_i^{(0)}\}, \{\tilde{\xi}_i^{(1)}\}, \cdots, \{\tilde{\xi}_i^{(j-1)}\} \) to be pairwise independent. Also, these sequence are forced to be independent of \( Y_d \) and \( \xi^{(l)}_{k,i} \) for arbitrary \( i, k, l \). For these new sequences we define a random variable \( \tilde{\Gamma}_d(V_j) \) and a truncation event \( T_{\tilde{\xi}}(V_j) \) as

$$\tilde{\Gamma}_d(V_j) = \sum_{l=0}^{j-1} \sum_{i=1}^{\nu_l} \tilde{\xi}_i^{(l)} + Y_d,$$

(6–73)

and

$$T_{\tilde{\xi}}(V_j) = \left\{ \bigcap_{l=0}^{j-1} \bigcap_{i=1}^{\nu_l} |\tilde{\xi}_i^{(l)}| \geq T_l \right\},$$

(6–74)

respectively.

With the above definitions, it is easy to see that

$$P\left( \Gamma_{w,d}^{(j)}(W_j) < x, \bigcap_{l=0}^{j-1} \bigcap_{i\in\tilde{A}_l} \bigcap_{k\in K_l^{(i)}} |\xi^{(l)}_{k,i}| \geq T_l \right) = P\left( \tilde{\Gamma}_d(V_j) < x, T_{\tilde{\xi}}(V_j) \right).$$

(6–75)
Thus, the upper bound (6–58) can be rewritten as

\[ P(\Gamma_{w,d}(W_j) < x, W_j) \leq c_j P(\tilde{\Gamma}_d(V_j) < x, T_\xi(V_j)). \]  

(6–76)

For convenience, we collect all the parameters \( m_l, n_l \) and \( \nu_l \) for \( 0 \leq l \leq j - 1 \) to form a \( 3j \)-tuple \( W_j \) as

\[ W_j = \left\{ m_l = |A_l|, n_l = |\bar{A}_l|, \nu_l = \sum_{i \in A_l} |K^{(l)}_i| \right\}_{l=0}^{j-1}. \]  

(6–77)

Then the upper bound \( c_j P(\tilde{\Gamma}_d(V_j) < x, T_\xi(V_j)) \) in (6–76) can be characterized by \( W_j \) instead of the particular information exchange event \( W_j \). That is, the probability \( P(\Gamma_{w,d}(W_j) < x, W_j) \) has the same upper bound for every event \( W_j \) subject to the constraint \( W_j \). If we use \( N(W_j) \) to denote the number of events satisfying constraint \( W_j \), then with (6–57) and (6–76) we have the following PEP upper bound,

\[ P(\Gamma_{w,d}^{(j)} < x) \leq \sum_{W_j} c_j P(\tilde{\Gamma}_d(V_j) < x, T_\xi(V_j)) = \sum_{W_j} N(W_j)c_j P(\tilde{\Gamma}_d(V_j) < x, T_\xi(V_j)), \]  

(6–78)

where \( \sum_{W_j} \) is a \( 3j \)-fold summation over all possible \( 3j \)-tuple \( W_j \). According to (6–64), (6–67) and (6–70), it is easy to see that

\[ \sum_{W_j} = \sum_{w_0}^{w_1} \sum_{w_1}^{w_2} \cdots \sum_{w_1}^{w_{j-1}} \sum_{m_0=0}^{m_1} \sum_{m_1=0}^{m_{j-1}} \cdots \sum_{m_1=0}^{m_{j-1}} \sum_{n_0=0}^{n_1} \sum_{n_1=0}^{n_{j-1}} \cdots \sum_{n_1=0}^{n_{j-1}} \sum_{\nu_0=0}^{\nu_1} \sum_{\nu_1=0}^{\nu_{j-1}} \cdots \sum_{\nu_{j-1}=0}^{\nu_{j-1}M'}, \]  

(6–79)

where \( M' = M - 1 \). For the calculation of \( N(W_j) \), we have the following theorem.

**Theorem 6.4.3** In collaborative decoding with \( M \) nodes, the number of all possible MRB information exchange events subject to the constraint \( W_j \) defined in (6–77) for an error sequence of size \( w - 1 \) is given by

\[ N(W_j) = \prod_{l=0}^{j-1} \binom{w_l}{m_l} \binom{m_l}{n_l} \theta_{M'}(n_l, \nu_l), \]  

(6–80)
With \( w_l \) calculated by (6–66), and

\[
\theta_{M'}(n_l, \nu_l) = \binom{n_l M'}{\nu_l} - \sum_{i=1}^{n_l - \lceil \frac{\nu_l}{M'} \rceil} \binom{n_l}{i} \theta_{M'}(n_l - i, \nu_l), \tag{6–81}
\]

where \( \lceil \cdot \rceil \) means taking the greatest integer no larger than, and \( \theta_{M'}(n, \nu) = 0 \) for \( n < \lceil \frac{\nu}{M'} \rceil \).

**Proof** For a given 3\( j \)-tuple \( W_j \), the number of all possible events equals the number of all possible choices for set \( \{ A_l, \bar{A}_l, \{ K_i^{(l)} \}_{i \in A_l} \}_{l=0}^{j-1} \). According to definition, the choice of \( \{ \bar{A}_l \}_{l=0}^{j-1} \) depends on \( \{ A_l \}_{l=0}^{j-1} \), and the choice of \( \{ \{ K_i^{(l)} \}_{i \in A_l} \}_{l=0}^{j-1} \) depends on \( \{ \bar{A}_l \}_{l=0}^{j-1} \). For convenience, we denote the numbers of possible choices for \( \{ A_l \}_{l=0}^{j-1} \), \( \{ \bar{A}_l \}_{l=0}^{j-1} \) conditioning on \( \{ A_l \}_{l=0}^{j-1} \), and \( \{ \{ K_i^{(l)} \}_{i \in A_l} \}_{l=0}^{j-1} \) conditioning on \( \{ \bar{A}_l \}_{l=0}^{j-1} \) by \( N_1 \), \( N_2 \) and \( N_3 \), respectively. Then we have \( N(W_j) = N_1 N_2 N_3 \).

We first consider the choices of \( \{ A_l \}_{l=0}^{j-1} \) and \( \{ \bar{A}_l \}_{l=0}^{j-1} \). In each iteration \( l \), \( A_l \) is arbitrarily chosen from \( w_l \) candidate bits with the constraint \( |A_l| = m_l \). Thus, there are \( N_1 = \prod_{l=0}^{j-1} \binom{w_l}{m_l} \) possible choices for \( \{ A_l \}_{l=0}^{j-1} \). Given a particular choice of \( \{ A_l \}_{l=0}^{j-1} \), \( \bar{A}_l \) is arbitrarily chosen from \( A_l \) with the constraint \( |ar{A}_l| = n_l \) for each \( l \), respectively. This gives \( N_2 = \prod_{l=0}^{j-1} \binom{m_l}{n_l} \) possible choices for \( \{ \bar{A}_l \}_{l=0}^{j-1} \).

For a given \( \{ \bar{A}_l \}_{l=0}^{j-1} \), the set \( K_i^{(l)} \) is randomly chosen from the node set \( M' \). This trial is then repeated for \( i \in \bar{A}_l \) and \( l < j \). If we denote the size of \( K_i^{(l)} \) by \( K_i^{(l)} \), then the choice of \( K_i^{(l)} \) is subject to the constraints \( 1 \leq K_i^{(l)} \leq M' \) and \( \sum_{i \in \bar{A}_l} K_i^{(l)} = \nu_l \). For convenience, we denote \( \theta_{M'}(n_l, \nu_l) \) the number of possible choices for \( \{ K_i^{(l)} \}_{i \in \bar{A}_l} \) given a particular \( \bar{A}_l \) with \( |ar{A}_l| = n_l \). Then, \( \theta_{M'}(n_l, \nu_l) \) can be expressed as

\[
\theta_{M'}(n_l, \nu_l) = \sum_{\{ K_i^{(l)} \}_{i \in \bar{A}_l} \, | \, 1 \leq K_i^{(l)} \leq M', \sum_{i \in \bar{A}_l} K_i^{(l)} = \nu_l} \prod_{i=1}^{n_l} \binom{M'}{K_i^{(l)}}. \tag{6–82}
\]

With this notation, the number of possible choices for \( \{ K_i^{(l)} \}_{i \in \bar{A}_l} \) given \( \{ \bar{A}_l \}_{l=0}^{j-1} \) is given by \( N_3 = \prod_{l=0}^{j-1} \theta_{M'}(n_l, \nu_l) \).
Now we evaluate $\theta_{M'}(n_l, \nu_l)$ in (6–82). Direct calculation of $\theta_{M'}(n_l, \nu_l)$ is usually difficult due to the constraint $\{1 \leq K_{i}^{(l)} \leq M'\}_{i=1}^{n_l}$. To avoid this difficulty, we first consider the choice of set $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$ from $\mathcal{M}'$ subject to the looser constraints that $\{0 \leq \tilde{K}_{i}^{(l)} \leq M'\}_{i=1}^{n_l}$ and $\nu_l = \sum_{i=1}^{n_l} \tilde{K}_{i}^{(l)}$, where $\tilde{K}_{i}^{(l)} = |\tilde{K}_{i}^{(l)}|$. For convenience, we denote the total number of possible choices for $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$ as $\tilde{\theta}_{M'}(n_l, \nu_l)$. Since choosing $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$ is equivalent to randomly choosing $\nu_l$ nodes from a set of $n_l M'$ nodes without replacement, we have

$$\tilde{\theta}_M(n_l, \nu_l) = \sum_{\nu_l} \prod_{i=1}^{n_l} \binom{M'}{K_i^{(l)}} = \binom{n_l M'}{\nu_l}.$$  \hspace{1cm} (6–83)

Meanwhile, we note that by removing all sets with $\tilde{K}_{i}^{(l)} = 0$ (i.e., $\tilde{K}_{i}^{(l)} = \emptyset$) for any $i \in \tilde{A}_l$ from the ensemble of all $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$ we can obtain the ensemble of all possible $\{K_i^{(l)}\}_{i \in \tilde{A}_l}$. Namely, the ensemble for $\{K_i^{(l)}\}_{i \in \tilde{A}_l}$ is a subset of that for $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$. To specify the ensemble of $\{\tilde{K}_{i}^{(l)}\}_{i \in \tilde{A}_l}$ with at least one $\tilde{K}_{i}^{(l)} = 0$, let $\mathcal{E}$ denote a subset of $\tilde{A}_l$ such that $\tilde{K}_{i}^{(l)} = 0$ for $i \in \mathcal{E}$, and let $\bar{\mathcal{E}}$ denote its complementary set in $\tilde{A}_l$. Thus, we have $1 \leq \tilde{K}_{i}^{(l)} \leq M'$ for $i \in \bar{\mathcal{E}}$. According to the constraint $\nu_l = \sum_{i=1}^{n_l} \tilde{K}_{i}^{(l)}$, it can be seen that $\mathcal{E}$ must satisfy $1 \leq |\mathcal{E}| \leq n_l - \lceil \nu_l / M' \rceil$. With the above arguments, $\tilde{\theta}_{M'}(n_l, \nu_l)$ can be written as

$$\tilde{\theta}_{M'}(n_l, \nu_l) = \sum_{\{1 \leq \tilde{K}_{i}^{(l)} \leq M'\}_{i=1}^{n_l}} \prod_{i=1}^{n_l} \binom{M'}{\tilde{K}_i^{(l)}} + \sum_{\nu_l = \sum_{i=1}^{n_l} \tilde{K}_i^{(l)}} \sum_{1 \leq |\mathcal{E}| \leq n_l - \lceil \nu_l / M' \rceil} \prod_{1 \leq \tilde{K}_{i}^{(l)} \leq M'} \prod_{i=1}^{n_l} \binom{M'}{\tilde{K}_i^{(l)}}$$

$$= \sum_{\{1 \leq \tilde{K}_{i}^{(l)} \leq M'\}_{i=1}^{n_l}} \prod_{i=1}^{n_l} \binom{M'}{\tilde{K}_i^{(l)}} + \sum_{i=1}^{n_l - \lceil \nu_l / M' \rceil} \binom{n_l}{i} \sum_{\nu_l = \sum_{j=1}^{n_l-i} \tilde{K}_j^{(l)}} \prod_{j=1}^{n_l-i} \binom{M'}{\tilde{K}_j^{(l)}}$$

$$= \theta_{M'}(n_l, \nu_l) + \sum_{i=1}^{n_l - \lceil \nu_l / M' \rceil} \binom{n_l}{i} \theta_{M'}(n_l - i, \nu_l).$$  \hspace{1cm} (6–84)
By inserting \((6\text{--}83)\) into \((6\text{--}84)\), we can establish a recursion for \(\theta_{M'}(n_t, \nu_t)\) as in \((6\text{--}81)\). With the initial condition \(\theta_{M'}(n, \nu) = 0\) for \(n < \lceil \frac{\nu}{M'} \rceil\) (this is obtained directly from definition), \(\theta_{M'}(n_t, \nu_t)\) can be calculated recursively.

We have obtained \(N_1\), \(N_2\) and \(N_3\). By inserting them into \(N(W_j) = N_1N_2N_3\), Theorem 6.4.3 is proved. \(\square\)

In the following, we refine the result in \((6\text{--}78)\). We note that the summation \(\sum_{W_j}\) indexes over \(\{m_t, n_t, \nu_t\}_{t=0}^{j-1}\), while the probability \(P(\tilde{\Gamma}_d(V_j) < x, \mathcal{T}_\xi(V_j))\) depends only on \(\{\nu_t\}_{t=0}^{j-1}\) but not on \(\{m_t, n_t\}_{t=0}^{j-1}\). Since the evaluation of \(P(\tilde{\Gamma}_d(V_j) < x, \mathcal{T}_\xi(V_j))\) involves some complicated numerical computations (see Section 6.5), the complexity of directly evaluating the upper bound in \((6\text{--}78)\) can be prohibitive. Thus, we hope to switch the summation order for \(\{m_t, n_t\}_{t=0}^{j-1}\) and \(\{\nu_t\}_{t=0}^{j-1}\) so as to avoid the multiple unnecessary evaluation of \(P(\tilde{\Gamma}_d(V_j) < x, \mathcal{T}_\xi(V_j))\). For its importance to our analysis, we state the switching result in the following theorem.

**Theorem 6.4.4** Suppose that \(f(W_j)\) is an arbitrary function of a \(3j\)-tuple \(W_j = \{m_t, n_t, \nu_t\}_{t=0}^{j-1}\) for \(j \geq 1\), and the summation \(\sum_{W_j}^{(3)}\) defined in \((6\text{--}79)\) is operated on \(f(W_j)\) with arbitrary integers \(w_0\) and \(M' > 0\). Then, by switching the summation order in \(\sum_{W_j}\) without any change on \(f(W_j)\), we have

\[
\sum_{W_j} f(W_j) = \sum_{\nu_0=0}^{w_0} \sum_{\nu_1=0}^{\nu_0'} \sum_{\nu_{j-1}=0}^{\nu_j-1} \sum_{m_0=[\frac{\nu_0}{M'}]}^{\frac{\nu_0}{M'}} \sum_{m_1=[\frac{\nu_1}{M'}]}^{\frac{\nu_1}{M'}} \sum_{m_{j-1}=[\frac{\nu_{j-1}}{M'}]}^{\frac{\nu_{j-1}}{M'}} \sum_{n_0=[\frac{\nu_0}{M'}]}^{\frac{\nu_0}{M'}} \sum_{n_1=[\frac{\nu_1}{M'}]}^{\frac{\nu_1}{M'}} \sum_{n_{j-1}=[\frac{\nu_{j-1}}{M'}]}^{\frac{\nu_{j-1}}{M'}} f(W_j), \quad (6\text{--}85)
\]

where \(w_{l+1}' = w_l' - \lceil \frac{\nu_l}{M'} \rceil\), \(w_l'' = w_l - \sum_{l=1}^{j-1} \lceil \frac{\nu_l}{M'} \rceil\), \(w_l''' = \min\{m_t, \nu_t\}\), \(w_{l+1} = w_l - m_t\) for all \(l\), and \(w_0' = w_0\).

**Proof** See Appendix C. \(\square\)

Applying Theorem 6.4.4 to \((6\text{--}78)\), we can obtain a new form of the upper bound for probability \(P(\Gamma_{w,d}^{(j)} < x)\). As one of the major results in this section, we present the new upper bound as the following theorem.
Theorem 6.4.5 For the collaborative decoding with MRB information exchange and nonrecursive convolutional codes, given the nodes number $M$ and the information exchange parameters $\{p_i, T_l\}_{l=0}^{j-1}$, we have the following upper bound for the error event metric $\Gamma_{w,d}^{(j)}$ defined in (6–49) for any arbitrary constant $x$,

$$P(\Gamma_{w,d}^{(j)} < x) \leq \sum_{V_j} \varphi(V_j) P(\hat{\Gamma}_d(V_j) < x, T_{\xi}(V_j)), \quad (6–86)$$

where $\sum V_j$ is a $j$-fold summation defined as

$$\sum_{V_j} = \sum_{v_0=0}^{w'_l} \sum_{v_1=0}^{w''_l} \cdots \sum_{v_{j-1}=0}^{w''''_l}, \quad (6–87)$$

and

$$\varphi(V_j) = \sum_{m_0=\lceil \frac{v_0}{M} \rceil}^{w'_l} \sum_{m_1=\lceil \frac{v_1}{M} \rceil}^{w''_l} \cdots \sum_{m_{j-1}=\lceil \frac{v_{j-1}}{M} \rceil}^{w''''_l} \sum_{n_0=\lceil \frac{v_0}{M} \rceil}^{w'_l} \sum_{n_1=\lceil \frac{v_1}{M} \rceil}^{w''_l} \cdots \sum_{n_{j-1}=\lceil \frac{v_{j-1}}{M} \rceil}^{w''''_l} c_j N(W_j), \quad (6–88)$$

with $w'_l$, $w''_l$, and $w''''_l$ for $l < j$ specified in Theorem 6.4.4. In the above, $\hat{\Gamma}_d(V_j)$, $T_{\xi}(V_j)$, $c_j$, and $N(W_j)$ are given in (6–73), (6–74), (6–71) and (6–80), respectively.

6.4.3 Upper Bounds for Probabilities Involving $\xi^{(j)}$

Through the analysis of MRB information exchange for error events in collaborative decoding, we have obtained an upper bound for the probability $P(\Gamma_{w,d}^{(j)} < x)$ as described in Theorem 6.4.5. This result characterizes the error behaviors of the decoder. With theorem 6.4.5 and the union bound (6–48) for collaborative decoding, it is possible to study some probabilities involving extrinsic information $\xi^{(j)}$. In particular, we will consider obtaining upper bounds for the probabilities $P(\xi^{(j)} < 0)$, $P(\xi^{(j)} + \sum_{k=1}^{K^{(j)}} \xi_k^{(l)} < 0, \bigcap_{k=1}^{K^{(j)}} |\xi_k^{(l)}| \geq T_l)$ and $P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l)$ in the BER bound (6–44).

The first result, also the main result, is for the probability $P(\xi^{(j)} < x)$, which can be obtained by directly applying Theorem 6.4.5 to (6–48).
Theorem 6.4.6  For collaborative decoding with MRB information exchange and non-recursive convolutional codes, given the nodes number $M$ and the information exchange parameters $\{p_l, T_l\}_{l=0}^{j-1}$, we have the following upper bound for the extrinsic information $\xi^{(j)}$ and any arbitrary constant $x$,

$$P(\xi^{(j)} < x) \leq \frac{1}{K_c} \sum_{d \geq d_{\text{min}}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \varphi(V_j) P(\tilde{\Gamma}_d(V_j) < x, T_{\xi}(V_j)).$$  \hfill (6-89)

Based on the analysis of MRB information exchange for error events, we can extend the above result to the probability $P(\xi^{(j)} + \sum_{k=1}^{K(\ell)} \xi^{(l)}_k < 0, \bigcap_{k=1}^{K(\ell)} |\xi^{(l)}_k| \geq T_l)$ in the BER upper bound (6-44). We present this result in the following corollary.

Corollary 6.4.7  For the collaborative decoding system in Theorem 6.4.6, let $\xi^{(j)}$ be the extrinsic information for a data bit at node $M$, and $\xi^{(j)}_k$ for $k = 1, 2, \ldots, M-1$ be the extrinsic information generated at the other $M-1$ nodes for the same bit. Then we have the following upper bound,

$$P(\xi^{(j)} + \sum_{k=1}^{K(\ell)} \xi^{(l)}_k < 0, \bigcap_{k=1}^{K(\ell)} |\xi^{(l)}_k| \geq T_l) \leq \frac{1}{K_c} \sum_{d \geq d_{\text{min}}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \varphi(V_j) P(\tilde{\Gamma}_d(V_j') < x, T_{\xi}(V_j'))$$  \hfill (6-90)

with $V_j' = \{\nu_t'\}_{t=0}^{j-1}$, where $\nu_t' = \nu_t$ for $t \neq l$ and $\nu_l' = \nu_l + K(\ell)$.

Proof  For convenience, let $X = \sum_{k=1}^{K(\ell)} \xi^{(l)}_k$, and $T_{\xi}(K(\ell)) = \left\{ \bigcap_{k=1}^{K(\ell)} |\xi^{(l)}_k| \geq T_l \right\}$. Also, let $X^*$ be the sum of $\xi^{(l)}_k$ conditioned on the truncation $T_{\xi}(K(\ell))$, i.e., $X^*$ equals $X$ conditioned on $T_{\xi}(K(\ell))$. Recall that in the union bound and error events analysis, the extrinsic information $\xi^{(j)}$ is assumed to associate with data bit $u_0$ in error sequence $u$, and all other extrinsic information $\xi^{(l)}_{k,i}$ involved in the event metric are associated with bit $u_i$ for $i \geq 1$. With this independence, $\{\xi^{(j)}_k\}_{k=1}^{M-1}$, and hence $T_{\xi}(K(\ell))$, $X$ and $X^*$ are independent of the error event metric $\Gamma^{(j)}_{w,d}$ and all information exchange.
events for sequence \( u \). Thus, we have

\[
P\left( \xi^{(j)} + \sum_{k=1}^{K^{(i)}} \xi_k^{(i)} < 0, \bigcap_{k=1}^{K^{(i)}} |\xi_k^{(i)}| \geq T_i \right) = P\left( \xi^{(j)} + X < 0, T_{\xi}^{(i)}(K^{(i)}) \right)
\]

\[
= P\left( \xi^{(j)} + X < 0 \big| T_{\xi}^{(i)}(K^{(i)}) \right) P\left( T_{\xi}^{(i)}(K^{(i)}) \right)
\]

\[
= P\left( \xi^{(j)} + X^* < 0 \right) P\left( T_{\xi}^{(i)}(K^{(i)}) \right).
\]  

(6–91)

Suppose that the random variable \( X^* \) has a pdf of \( f(x) \). Since \( X^* \) is independent of \( \xi^{(j)} \) and the associated error events, by employing Theorem 6.4.6, we can obtain

\[
P(\xi^{(j)} + X^* < 0) = \int P(\xi^{(j)} + X^* < 0 | X^* = x) f(x) dx = \int P(\xi^{(j)} + x < 0) f(x) dx
\]

\[
\leq \int \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) P(\tilde{\Gamma}_d(V_j) + x < 0, T_{\xi}(V_j)) f(x) dx
\]

\[
= \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) \int P(\tilde{\Gamma}_d(V_j) + x < 0, T_{\xi}(V_j)) f(x) dx
\]

\[
= \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) P(\tilde{\Gamma}_d(V_j) + X^* < 0, T_{\xi}(V_j)).
\]  

(6–92)

Inserting (6–92) into (6–91), we have

\[
P\left( \xi^{(j)} + \sum_{k=1}^{K^{(i)}} \xi_k^{(i)} < 0, \bigcap_{k=1}^{K^{(i)}} |\xi_k^{(i)}| \geq T_i \right)
\]

\[
\leq \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) P(\tilde{\Gamma}_d(V_j) + X^* < 0, T_{\xi}(V_j)) P\left( T_{\xi}^{(i)}(K^{(i)}) \right)
\]

\[
= \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) P(\tilde{\Gamma}_d(V_j) + X < 0, T_{\xi}(V_j), T_{\xi}^{(i)}(K^{(i)})).
\]  

(6–93)
Now we consider the probability \( P(\tilde{\Gamma}_d(V_j) + X < 0, T_{\xi}(V_j), T_{\xi}^{(l)}(K^{(l)}) \) in (6–93). By substituting the definitions of \( \tilde{\Gamma}_d(V_j), X, T_{\xi}(V_j) \) and \( T_{\xi}^{(l)}(K^{(l)}) \) into it, we obtain

\[
P(\tilde{\Gamma}_d(V_j) + X < 0, T_{\xi}(V_j), T_{\xi}^{(l)}(K^{(l)}))
= P \left( \sum_{t=0}^{j-1} \sum_{i=1}^{\nu} \tilde{\xi}_i + Y_d < 0, \bigcap_{t=0}^{j-1} \bigcap_{i=1}^{\nu} |\tilde{\xi}_i^{(t)}| \geq T_t, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| \geq T_t \right)
= \left( \begin{array}{c}
(a) \quad P \left( \sum_{t=0}^{j-1} \sum_{i=1}^{\nu} \tilde{\xi}_i^{(t)} + Y_d < 0, \bigcap_{t=0}^{j-1} \bigcap_{i=1}^{\nu} |\tilde{\xi}_i^{(t)}| \geq T_t, \bigcap_{k=1}^{K^{(l)}} |\xi_k^{(l)}| \geq T_t \right)
(b) \quad P(\tilde{\Gamma}_d(V_j') + X < 0, T_{\xi}(V_j')) \end{array} \right)
\] (6–94)

In the above, (a) is obtained by substituting \( \{\tilde{\xi}_i^{(t)}\}_{i=\nu+1}^{\nu+K^{(l)}} \) for \( \{\xi_i^{(l)}\}_{i=1}^{K^{(l)}} \) based on the fact that \( \tilde{\xi}_i^{(t)} \) and \( \xi_i^{(l)} \) are independent and have the same distribution. With the definition \( \nu'_t = \nu_t \) for \( t \neq l \) and \( \nu'_l = \nu_l + K^{(l)} \), this substitution will not change the probability. In addition, (b) is obtained based on the definitions of \( \Gamma_d(V_j) \) and \( T_{\xi}(V_j) \) with \( V_j' = \{\nu'_t\}_{t=0}^{j-1} \). Finally, by inserting (6–94) into (6–93), corollary 6.4.7 is proved. □

To this point, we only have probability \( P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \) left to upper bound. Different from the previous results, this probability involves the correlation between the extrinsic information for the same data bits in different iterations, i.e., \( \xi^{(j)} \) and \( \xi^{(l)} \) for \( l < j \). As in the analysis for LRB information exchange, with the approximation of \( \xi^{(l)} = \min_{(u,c) \in C^-} \{\Gamma^{(l)}_{u,c}\} \) (where \( C^- \) is the set of codeword pairs \((u,c)\) that give the decision \( u_0 = 1 \)), it turns out that we still can obtain an upper bound for \( P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \) by following the derivation for Theorem 6.4.6. We state this result in the following corollary.

**Corollary 6.4.8** For the collaborative decoding system in Theorem 6.4.6, let \( \xi^{(l)} \) and \( \xi^{(j)} \) for \( l < j \) be the extrinsic information for the same data bit at node \( M \) in different
decoding iterations. Then we have the following upper bound,

\[
P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \leq \frac{1}{K_c \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \phi(V_j) P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, T_\xi(V_j)), \quad (6-95)
\]

where \( V_l = \{v_t\}_{t=0}^{l-1} \).

Proof With the approximation of \( \xi^{(l)} = \min_{(u,c) \in C^-} \{ \Gamma^{(l)}(u,c) \} \) for \( l < j \), we have

\[
P(\xi^{(j)} < 0, \xi^{(l)} \geq T_l) \leq P \left( \min_{(u,c) \in C^-} \{ \Gamma^{(j)}(u,c) \} < 0, \min_{(u,c) \in C^-} \{ \Gamma^{(l)}(u,c) \} \geq T_l \right)
= \frac{1}{K_c} P \left( \bigcup_{(u,c) \in C^-} \Gamma^{(j)}(u,c) < 0, \bigcap_{(u,c) \in C^-} \Gamma^{(l)}(u,c) \geq T_l \right)
\leq \frac{1}{K_c} \sum_{(u,c) \in C^-} P \left( \Gamma^{(j)}(u,c) < 0, \bigcap_{(u',c') \in C^-} \Gamma^{(l)}(u',c') \geq T_l \right)
\leq \frac{1}{K_c} \sum_{(u,c) \in C^-} P(\Gamma^{(j)}(u,c) < 0, \Gamma^{(l)}(u,c) \geq T_l)
= \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} P(\Gamma^{(j)}(u,c) < 0, \Gamma^{(l)}(u,c) \geq T_l). \quad (6-96)
\]

Note that, in (6–96) \( \Gamma^{(j)}_{w,d} \) and \( \Gamma^{(l)}_{w,d} \) are the metrics for the same error sequence in different decoding iteration. Since \( l < j \), \( \Gamma^{(l)}_{w,d} \) consists of the part of \( \Gamma^{(j)}_{w,d} \) that is generated by the information exchange events during first \( l \) iterations. Thus, given a MRB information exchange event \( \mathcal{W}_j \) defined by (6–54), for any \( l < j \), we can represent \( \Gamma^{(j)}_{w,d}(\mathcal{W}_j) \) as

\[
\Gamma^{(j)}_{w,d}(\mathcal{W}_j) = \Gamma^{(l)}_{w,d}(\mathcal{W}_l) + \sum_{t=l}^{j-1} \sum_{i \in A_t} \sum_{k \in \Gamma^{(l)}_{i,k}} \xi^{(l)}_{k,i}. \quad (6-97)
\]

where \( \mathcal{W}_l \) specifies the information exchange process for the first \( l \) iterations in \( \mathcal{W}_j \). With this fact, similar to (6–57), we can expend the probability \( P(\Gamma^{(j)}_{w,d} < 0, \Gamma^{(l)}_{w,d} \geq T_l) \) as

\[
P(\Gamma^{(j)}_{w,d} < 0, \Gamma^{(l)}_{w,d} \geq T_l) = \sum_{\mathcal{W}_j} P(\Gamma^{(j)}_{w,d}(\mathcal{W}_j) < 0, \Gamma^{(l)}_{w,d}(\mathcal{W}_l) \geq T_l, \mathcal{W}_j). \quad (6-98)
\]
Then, replacing \( P(\Gamma^{(j)}_{w,d}(W_j) < x, W_j) \) by \( P(\Gamma^{(j)}_{w,d}(W_j) < 0, \Gamma^{(l)}_{w,d}(W_l) \geq T_l, W_j) \) in the derivations in Section 6.4.2 yields the following upper bound

\[
P(\Gamma^{(j)}_{w,d} < 0, \Gamma^{(l)}_{w,d} \geq T_l) \leq \sum_{V_j} \varphi(V_j) P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, \tilde{T}_\xi(V_j)), \quad (6–99)
\]

where \( V_l = \{\nu_t\}_{t=0}^{l-1} \). Thus, by combining (6–96) and (6–99), the corollary results. □

### 6.5 Evaluation of BER Upper Bound

As mentioned previously, our goal is to predict the BER performance of collaborative decoding in the \( j \)th iteration with only the statistical knowledge of the extrinsic information and the information exchange parameters up to the \( (j - 1) \)th iteration. For this goal, we substitute the results in Section 6.4.3 into (6–44), with simple arrangements, to obtain the following BER upper bound

\[
P_b^{(j)} \leq \sum_{l=0}^{j-2} a_l^{M-1} P(\xi^{(l)} < -T_l) \quad + \quad \frac{1}{K_d} \sum_{d \geq d_{\text{min}}} \sum_{w \geq 1} w A_{w,d} \sum_{V_j} \varphi(V_j) \left\{ a_{j-1}^{M-1} P(\tilde{\Gamma}_d(V_j) < 0, \tilde{T}_\xi(V_j)) \right\}
\]

\[
+ \quad \sum_{l=0}^{j-2} a_l^{M-1} P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, \tilde{T}_\xi(V_j)) \quad + \quad \sum_{l=0}^{j-1} \sum_{K^{(l)}} b_{l,K} P(\tilde{\Gamma}_d(V_j') < 0, \tilde{T}_\xi(V_j')) \right\}, \quad (6–100)
\]

where the coefficients \( a_l, b_{l,K} \) and \( \varphi(V_j) \) are calculated in (6–45), (6–46) and (6–88), respectively, and the relation between \( V_j \) and \( V_j' \) is defined as in Corollary 6.4.7.

According to the definition of \( \tilde{\Gamma}_d(V_j) \) and \( \tilde{T}_\xi(V_j) \), we can see that the upper bound in (6–100) involves \( \xi^{(l)} \) and \( \tilde{\xi}_i^{(l)} \) only for \( l \leq j - 1 \). This achieves the above mentioned goal. Based on the statistical assumption of the extrinsic information in Section 6.1, we know that, for all \( l \), \( \xi^{(l)} \sim \mathcal{N}(\mu_l, \sigma_l^2) \) for AWGN channels and \( \xi^{(l)} \sim \mathcal{GAL}(\mu_l, \sigma_l^2, \tau_l) \) for independent Rayleigh fading channels, respectively. Thus the probability \( P(\xi^{(l)} < -T_l) \) in (6–100) is given by

\[
P(\xi^{(l)} < -T_l) = Q\left(\frac{T_l + \mu_l}{\sigma_l}\right) \quad (6–101)
\]
for AWGN channels, and by

$$P(\xi^{(l)} < -T_l) = G(-T_l; \mu_l, \sigma^2_l, \tau_l)$$

(6–102)

for independent Rayleigh fading channel, respectively. In above, $Q(\cdot)$ is the Gaussian $Q$-function and $G(\cdot; \mu, \sigma^2, \tau)$ is the cdf of the distribution $\mathcal{GAL}(\mu, \sigma^2, \tau)$. An efficient method for the numerical evaluation of $G(\cdot; \mu, \sigma^2, \tau)$ is given in Appendix D.

Thus the problem of evaluating the BER bound becomes evaluating the probabilities $P(\tilde{\Gamma}_d(V_j) < 0, T_{\tilde{\xi}}(V_j))$, $P(\tilde{\Gamma}_d(V_j') < 0, T_{\tilde{\xi}}(V_j'))$ and $P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, T_{\tilde{\xi}}(V_j))$ for all possible $V_j$, given the statistical knowledge of the sequence $\{\tilde{\xi}^{(l)}_i\}$. With the definitions of $\tilde{\Gamma}_d(V_j)$ and $T_{\tilde{\xi}}(V_j)$, we know that these probabilities are essentially about sums of independent truncated random variables, for which the exact calculation are usually difficult. In this case, we again rely on upper-bounding techniques to approximate the probabilities. In the literature, the well studied cases concern with sums of inner truncated random variables, i.e., $|X_i| < B_i$, which results in a set of bounded random variables. Due to the truncation $|\tilde{\xi}^{(l)}_i| \geq T_l$ in $T_{\tilde{\xi}}(V_j)$, the well established Hoeffding’s inequalities [40] and other related inequalities such as those in [41]–[45] are not applicable to our situation. While no generalized bound is available for our problem, we consider upper bounding the probabilities of concern based on Laplace transform and saddlepoint approximation for AWGN channels and independent Rayleigh fading channels respectively.

### 6.5.1 AWGN Channel

For AWGN channels, it is easy to check that $Y_d \sim \mathcal{N}(dL_c, 2dL_c)$ from (6–56). According to Assumption 6.1.1 and the definition of the pairwise independent sequence $\{\tilde{\xi}^{(l)}_i\}$, we know that for each $l$, $\tilde{\xi}^{(l)}_i \sim \mathcal{N}(\mu_l, \sigma^2_l)$ for all $i$. With this statistical knowledge, we first consider the probability $P(\tilde{\Gamma}_d(V_j) < 0, T_{\tilde{\xi}}(V_j))$. Recall that the truncation threshold $T_l$ is related to the information exchange parameter $p_l$ by (6–6). If $p_l = 1$ in certain iteration $l < j$, then we have $T_l = 0$. In this case, since
\(P(|\xi_{i}^{(l)}| \geq 0) = 1\), the truncation \(\{|\xi_{i}^{(l)}| \geq T_{i}\}\) for all \(i\) can be removed from the event \(T_{\xi}(V_{j})\) and the random variables \(\{\xi_{i}^{(l)}\}\) can be merged into \(Y_{d}\) for simplicity.

In order to include this special case into the following derivation, we define an iteration index \(L \leq j\) such that \(T_{l} > 0\) for \(l < L\) and \(T_{l} = 0\) for \(l = L, L + 1, \cdots, j - 1\). If \(L > j - 1\), then \(T_{l} = 0\) does not occur for all \(l\). With this definition, we have

\[
\tilde{\Gamma}_{d}(V_{j}) = \sum_{l=0}^{j-1} \sum_{i=1}^{v_{l}} \tilde{\xi}_{i}^{(l)} + Y_{d} = S(V_{L}) + Y_{d}',
\]  

(6–103)

where \(V_{L} = \{v_{l}\}_{l=0}^{L-1}\),

\[
S(V_{L}) = \sum_{l=0}^{L-1} \sum_{i=1}^{v_{l}} \tilde{\xi}_{i}^{(l)},
\]  

(6–104)

and

\[
Y_{d}' = \sum_{l=L}^{j-1} \sum_{i=1}^{v_{l}} \tilde{\xi}_{i}^{(l)} + Y_{d}.
\]  

(6–105)

From the above, let

\[
\mu_{Y} = dL_{c} + \sum_{i=L}^{j-1} v_{i} \mu_{l}, \quad \text{and} \quad \sigma_{Y}^{2} = 2dL_{c} + \sum_{i=L}^{j-1} v_{i} \sigma_{l}^{2},
\]  

(6–106)

then \(Y_{d}' \sim \mathcal{N}(\mu_{Y}, \sigma_{Y}^{2})\).

With the above notation, we have

\[
P(\tilde{\Gamma}_{d}(V_{j}) < 0, T_{\xi}(V_{j}) = P(S(V_{L}) + Y_{d}' < 0, T_{\xi}(V_{L})).
\]  

(6–107)

We evaluate the probability \(P(S(V_{L}) + Y_{d}' < 0, T_{\xi}(V_{L}))\) for different values of \(V_{L}\) as follows. When \(V_{L} = \{0\}\), i.e., \(v_{l} = 0\) for \(l < L\), we have \(S(V_{L}) = 0\) and

\[
P(S(V_{L}) + Y_{d}' < 0, T_{\xi}(V_{L})) = P(Y_{d}' < 0) = Q\left(\frac{\mu_{Y}}{\sigma_{Y}}\right)
\]  

(6–108)
with $\mu_Y$ and $\sigma_Y$ defined in (6–106). When $S(V_L)$ contains only one random variable, i.e., $\nu_l = 1$ and $\nu_t = 0$ with $t \neq l$ in $V_L$ for arbitrary $l < L$, we have

\[
P(S(V_L) + Y'_d < 0, T_\xi(V_L)) = P(\xi_i^{\prime(l)} + Y'_d < 0, |\xi_i^{\prime(l)}| \geq T_i)
\]

\[
= P(\xi_i^{\prime(l)} + Y'_d < 0, \xi_i^{\prime(l)} \leq -T_i) + P(\xi_i^{\prime(l)} + Y'_d < 0, \xi_i^{\prime(l)} \geq T_i)
\]

\[
= Q\left(\frac{\mu_l + \mu_Y}{\sqrt{\sigma_i^2 + \sigma_Y^2}}; \frac{\mu_l + T_i}{\sigma_i} \right) + Q\left(\frac{\mu_l + \mu_Y}{\sqrt{\sigma_i^2 + \sigma_Y^2}}; \frac{T_l - \mu_l}{\sigma_l} \right),
\]

(6–109)

where $Q(\cdot, \cdot; \cdot)$ is the bivariate Gaussian $Q$-function specified in (5–58).

When $S(V_L)$ contains more than one random variables, i.e., $\sum_{i=0}^{L-1} \nu_i \geq 2$, we upper bound the probability $P(S(V_L) + Y'_d < 0, T_\xi(V_L))$ based on Laplace transform. For convenience, we rewrite this probability as

\[
P(S(V_L) + Y'_d < 0, T_\xi(V_L)) = P(S(V_L) + Y'_d < 0 | T_\xi(V_L)) P(T_\xi(V_L))
\]

\[
= P(S(V_L) + Y'_d < 0) \prod_{l=0}^{L-1} \prod_{i=1}^{\nu_l} P(|\xi_i^{\prime(l)}| \geq T_i), \quad (6–110)
\]

where $\hat{S}(V_L) = \sum_{i=0}^{L-1} \sum_{i=1}^{\nu_i} \xi_i^{\prime(l)}$ and $\xi_i^{\prime(l)}$ is the truncated version of random variable $\hat{\xi}_i^{\prime(l)}$, i.e.,

\[
\hat{\xi}_i^{\prime(l)} = \tilde{\xi}_i^{\prime(l)} |\tilde{\xi}_i^{\prime(l)}| \geq T_l. \quad (6–111)
\]

Now, denote $\Phi_Y(s)$ the mgf, i.e., the double-sided Laplace transform of $Y'_d$. Since $Y'_d \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, we have

\[
\Phi_Y(s) = E[e^{-sY'_d}] = e^{-s\mu_Y + \frac{s^2\sigma_Y^2}{2}}, \quad (6–112)
\]

where $s = \alpha + j\omega$, and the ROC is $-\infty < \alpha < +\infty$. Also, denote $\hat{\Phi}_l(s)$ the mgf of truncated random variable $\hat{\xi}_i^{\prime(l)}$ for all $i$, then we have the following proposition about $\hat{\Phi}_l(s)$. 
**Proposition 6.5.1** The mgf of the truncated Gaussian random variable \( \hat{\xi}_i^{(l)} \) defined in (6–111) with \( \hat{\xi}_i^{(l)} \sim \mathcal{N}(\mu_l, \sigma^2_l) \) is given by

\[
\hat{\Phi}_l(s) = E[e^{-s\hat{\xi}_i^{(l)}}] = \frac{\Phi_l(s)h_l(s)}{P(|\tilde{\xi}_i^{(l)}| \geq T_i)},
\]

where \( \Phi_l(s) \) is mgf of the Gaussian random variable \( \tilde{\xi}_i^{(l)} \), i.e.,

\[
\Phi_l(s) = e^{-s\mu_l + \frac{s^2\sigma^2_l}{2}},
\]

\( h_l(s) \) is defined as

\[
h_l(s) = Q\left(\frac{T_l + \mu_l + s\sigma^2_l}{\sigma_l}\right) + Q\left(\frac{T_l - \mu_l - s\sigma^2_l}{\sigma_l}\right),
\]

and the ROC is \(-\infty < \alpha < +\infty\).

**Proof** Since \( \tilde{\xi}_i^{(l)} \sim \mathcal{N}(\mu_l, \sigma^2_l) \), with (6–111) we can obtain the pdf of \( \hat{\xi}_i^{(l)} \) as

\[
\hat{f}_l(x) = \begin{cases} 
\frac{1}{P(|\tilde{\xi}_i^{(l)}| \geq T_i)}\phi(x; \mu_l, \sigma^2_l) & |x| \geq T_i \\
0 & -T_i < x < T_i
\end{cases},
\]

where \( \phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \) is the pdf of \( \mathcal{N}(\mu, \sigma^2) \). Thus, the mgf of \( \hat{\xi}_i^{(l)} \) is calculated as

\[
\hat{\Phi}_l(s) = E[e^{-s\hat{\xi}_i^{(l)}}] = \int_{-\infty}^{+\infty} e^{-sx}\hat{f}_l(x)dx
\]

\[
= \frac{1}{P(|\tilde{\xi}_i^{(l)}| \geq T_i)} \left( \int_{-\infty}^{-T_i} e^{-sx}\phi(x; \mu_l, \sigma^2_l)dx + \int_{T_i}^{+\infty} e^{-sx}\phi(x; \mu_l, \sigma^2_l)dx \right)
\]

\[
= \frac{1}{P(|\tilde{\xi}_i^{(l)}| \geq T_i)} \left( \int_{-\infty}^{-T_i} \frac{1}{\sqrt{2\pi}\sigma_l}e^{-\frac{(x-\mu_l-s\sigma^2_l)^2}{2\sigma^2_l}}dx + \int_{T_i}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_l}e^{-\frac{(x-\mu_l+s\sigma^2_l)^2}{2\sigma^2_l}}dx \right)
\]

\[
= \frac{e^{-s\mu_l + \frac{s^2\sigma^2_l}{2}}}{P(|\tilde{\xi}_i^{(l)}| \geq T_i)} \left( Q\left(\frac{T_l + \mu_l + s\sigma^2_l}{\sigma_l}\right) + Q\left(\frac{T_l - \mu_l - s\sigma^2_l}{\sigma_l}\right) \right),
\]

and it is easy to check that the ROC is \(-\infty < \alpha < +\infty\). 

\( \square \)
For convenience, denote $\hat{\Phi}(s)$ the mgf of $\hat{S}(V_L) + Y'_d$, then with the fact that 
$\{Y'_d, \{\hat{\xi}_i^{(0)}\}, \ldots, \{\hat{\xi}_i^{(L-1)}\}\}$ are pairwise independent, $\hat{\Phi}(s)$ is given by

$$
\hat{\Phi}(s) = \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_{\hat{\xi}_l}^{(l)}(s) = \frac{1}{\prod_{l=0}^{L-1} \prod_{i=1}^{\nu_l} P(|\hat{\xi}_l^{(l)}| \geq T_l)} \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_{\hat{\xi}_l}^{(l)}(s) h_{\hat{\xi}_l}^{(l)}(s),
$$

(6–118)

where $\Phi(s)$ is defined as

$$
\Phi(s) = \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_{\hat{\xi}_l}^{(l)}(s) h_{\hat{\xi}_l}^{(l)}(s).
$$

(6–119)

The ROC of $\hat{\Phi}(s)$ is the intersection of the ROCs of $\Phi_Y(s)$ and $\{\hat{\Phi}_l(s)\}_{l=0}^{L-1}$, and hence is still $-\infty < \alpha < +\infty$. By applying the well-known Chernoff bound, we have

$$
P(\hat{S}(V_L) + Y'_d < 0) < \min_{\alpha \in \text{ROC}, \alpha > 0} \Phi(\alpha).
$$

(6–120)

Thus, with (6–110), (6–118) and (6–120), we obtain a Chernoff bound for the joint probability $P(S(V_L) + Y'_d < 0, T_{\hat{\xi}(V_L)})$ as

$$
P(S(V_L) + Y'_d < 0, T_{\hat{\xi}(V_L)}) < \Phi(\alpha_C), \ \text{with} \ \alpha_C = \arg\min_{\alpha > 0} \Phi(\alpha).
$$

(6–121)

From the definition of $\Phi(s)$ in (6–119), we note that due to the function $h_l(s)$, we need to use numerical root finding algorithms such as the Newton-Raphson method to find the minimizer $\alpha_C$ in (6–124). Recall that due to the $j$-fold summation $\sum_{V_j}$ in (6–100), the total complexity of evaluating the BER bound is polynomial with the complexity of evaluating the probabilities inside $\sum_{V_j}$ and is exponential with $j$. In order to simplify the computation, we loosen the bound in (6–124) by substituting some nearby and easy found point $\tilde{\alpha}_C$ for $\alpha_C$. One possible choice of $\tilde{\alpha}_C$ is the minimizer of $\tilde{\Phi}(s)$ defined as

$$
\tilde{\Phi}(s) = \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_{\hat{\xi}_l}^{(l)}(s) = \exp\left\{ -s\left(\mu_Y + \sum_{l=0}^{L-1} \nu_l \mu_l\right) + \frac{s^2}{2} \left(\sigma_Y^2 + \sum_{l=0}^{L-1} \nu_l \sigma_l^2\right)\right\},
$$

(6–122)
which is obtained from (6–119) by dropping all \( h_i(s) \). Then, it is easy to obtain

\[
\tilde{\alpha}_C = \arg\min_{\alpha > 0} \Phi(\alpha) = \frac{\mu_Y + \sum_{t=0}^{L-1} \nu_t \mu_t}{\sigma_Y^2 + \sum_{t=0}^{L-1} \nu_t \sigma_t^2} \tag{6–123}
\]

Thus, with \( \Phi(\alpha_C) \leq \Phi(\alpha) \) for \( \alpha \neq \alpha_C \) and \( \alpha > 0 \), the joint probability \( P(S(V_L), Y_d' < 0, T_\xi(V_L)) \) can be upper bounded as

\[
P(S(V_L), Y_d' < 0, T_\xi(V_L)) < \Phi(\tilde{\alpha}_C), \tag{6–124}
\]

where \( \Phi(s) \) is defined in (6–119).

The above results for \( P(\tilde{\Gamma}_d(V_j) < 0, T_\xi(V_j)) \) can be easily applied to the evaluation for \( P(\tilde{\Gamma}_d(V_j') < 0, T_\xi(V_j')) \) by substituting \( V_j' \) for \( V_j \) properly.

Now, we only have the probability \( P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq l_t, T_\xi(V_j)) \) for \( l \leq j - 2 \) left to evaluate. Since \( V_j = \{\nu_j\}_{t=0}^{j-1} \) and \( V_l = \{\nu_t\}_{t=0}^{l-1} \), we have \( V_j = \{V_l, \{\nu_t\}_{t=1}^{j-1}\} \) and,

\[
\tilde{\Gamma}_d(V_j) = \tilde{\Gamma}_d(V_l) + \sum_{t=0}^{j-1} \nu_t \tilde{\xi}_t^{(j)}. \tag{6–125}
\]

Thus, for any \( V_j \) with \( \{\nu_t\}_{t=1}^{j-1} = \{0\} \), we have

\[
P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq l_t, T_\xi(V_j)) = P(\tilde{\Gamma}_d(V_l) < 0, \tilde{\Gamma}_d(V_l) \geq l_t, T_\xi(V_j)) = 0, \tag{6–126}
\]

since \( l_t \geq 0 \). When all \( \nu_t = 0 \) in \( V_j \) except only one \( \nu_r = 1 \) with \( r \geq l \), i.e., \( V_l = \{0\} \) and \( \sum_{t=0}^{j-1} \nu_t = \nu_r = 1 \), we have \( \tilde{\Gamma}_d(V_l) = Y_d \) and \( \tilde{\Gamma}_d(V_j) = Y_d + \tilde{\xi}_1^{(r)} \). Hence, the probability can be upper bounded as

\[
P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq l_t, T_\xi(V_j)) = P(Y_d + \tilde{\xi}_1^{(r)} < 0, Y_d \geq l_t, |\tilde{\xi}_1^{(r)}| \geq T_r)
\]

\[
= P(Y_d + \tilde{\xi}_1^{(r)} < 0, Y_d \geq l_t, \tilde{\xi}_1^{(r)} \leq -T_r) + P(Y_d + \tilde{\xi}_1^{(r)} < 0, Y_d \geq l_t, \tilde{\xi}_1^{(r)} \geq T_r)
\]

\[
= P(Y_d + \tilde{\xi}_1^{(r)} < 0, Y_d \geq l_t, \tilde{\xi}_1^{(r)} \leq -T_r)
\]

\[
< P(Y_d + \tilde{\xi}_1^{(r)} < 0, \tilde{\xi}_1^{(r)} \leq -T_r)
\]

\[
= \frac{\mu_r + dL_c}{\sqrt{\sigma_r^2 + 2dL_c}} \frac{\mu_r + T_r}{\sigma_r} \frac{\sigma_r}{\sqrt{\sigma_r^2 + 2dL_c}}, \tag{6–127}
\]
where we have used $Y_d \sim \mathcal{N}(dL_c, 2dL_c)$ and $\tilde{z}_1^{(r)} \sim \mathcal{N}(\mu_r, \sigma^2_r)$.

For the other case of $V_j$ with $\{\nu_l\}_{l=1}^{j-1} \neq \{0\}$, i.e., $\sum_{i=1}^{j-1} \nu_i \geq 2$ and $\{\nu_l\}_{l=1}^{j-1} \neq \{0\}$, we upper bound $P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_i) \geq T_i, T_{\tilde{\xi}}(V_j))$ by dropping the event $\{\tilde{\Gamma}_d(V_i) \geq T_i\}$, i.e.,

$$P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_i) \geq T_i, T_{\tilde{\xi}}(V_j)) < P(\tilde{\Gamma}_d(V_j) < 0, T_{\tilde{\xi}}(V_j)), \quad (6-128)$$

which has been discussed above.

### 6.5.2 Independent Rayleigh Fading Channel

For independent Rayleigh fading channels, from (6–56) it is easy to show that $Y_d \sim \mathcal{GAL}(L_c, 2L_c, d)$ by using the GAL characteristic function. According to Assumption 6.1.1 and the definition of the sequence $\{\xi_i(l)\}$, we know that for each $l$, $\tilde{\xi}_i(l) \sim \mathcal{GAL}(\mu_l, \sigma^2_l, \tau_l)$ for all $i$. Following the definition and notation in (6–103) through (6–105), (6–107) still holds for Rayleigh fading channels, but $Y'_d$ in (6–105) is not Gaussian any more. In this case, according to (6–2) and independence among $Y_d$ and $\{\tilde{\xi}_i(l)\}$, the mgf of $Y'_d$ is given by

$$\Phi_{Y}(s) = E[e^{-sY'_d}] = \frac{1}{(1 + L_c s - L_c s^2)^d} \prod_{l=L}^{j-1} \frac{1}{(1 + \mu_l s - \frac{\sigma^2_l}{2}s^2)\nu_l \tau_l}, \quad (6-129)$$

for which the ROC is $\alpha_{Y,1} < \alpha < \alpha_{Y,2}$ with

$$\alpha_{Y,1} = \max_{L \leq i < j} \left\{ \frac{1 - \sqrt{1 + 4/L_c}}{2}, \frac{\mu_l - \sqrt{\mu^2_l + 2\sigma^2_l}}{\sigma^2_l} \right\}, \quad (6-130)$$

and

$$\alpha_{Y,2} = \min_{L \leq i < j} \left\{ \frac{1 + \sqrt{1 + 4/L_c}}{2}, \frac{\mu_l + \sqrt{\mu^2_l + 2\sigma^2_l}}{\sigma^2_l} \right\}. \quad (6-131)$$

Similar to the case of AWGN channels, we consider $P(S(V_L) + Y'_d < 0, T_{\tilde{\xi}}(V_L))$ for different values of $V_L$. When $V_L = \{0\}$, i.e., $\nu_l = 0$ for $l < L$, we have $S(V_L) = 0$ and

$$P(S(V_L) + Y'_d < 0, T_{\tilde{\xi}}(V_L)) = P(Y'_d < 0). \quad (6-132)$$
Since no closed-form result exists for it, we evaluate the probability $P(Y'_d < 0)$ numerically based on the saddlepoint approximation [46, 47], which is known as an accurate approximation technique for probability distributions. As well known, given the mgf $\Phi_Y(s)$, the probability $P(Y'_d < 0)$ can be obtained as

$$P(Y'_d < 0) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{\Phi_X(s)}{s} \, dx,$$

(6–133)

where $c > 0$ lies in the ROC of $\Phi_X(s)$. In this case, based on the saddlepoint approximation theory, (6–133) can be approximated as [48]

$$P(Y'_d < 0) = \frac{1}{2\pi j} \int_{\alpha_S-j\infty}^{\alpha_S+j\infty} \frac{\Phi_Y(s)}{s} \, dx \approx \sqrt{\frac{\Phi_Y(\alpha_S)}{2\pi \Phi_Y''(\alpha_S)}} \frac{\Phi_Y(\alpha_S)}{\alpha_S},$$

(6–134)

where $\alpha_{Y,1} < \alpha_S < \alpha_{Y,2}$ is the saddlepoint (i.e., the solution lying in the ROC to the equation $\Phi_Y'(\alpha) = 0$ for this special case), and $\Phi_Y'(\cdot)$ and $\Phi_Y''(\cdot)$ are the first and second derivatives of the mgf $\Phi_Y(\cdot)$ given in (6–129). Here, we use the Newton-Raphson method to find the saddle point $\alpha_S$ numerically, then (6–134) can be evaluated easily.

When $V_L \neq \{0\}$, i.e., $\sum_{l=0}^{L-1} \nu_l \geq 1$, similar to the case of AWGN channels, we define $\tilde{\xi}_i^{(l)}$ as the truncated version of random variable $\tilde{\xi}_i^{(l)}$ according to (6–111), and define $\tilde{S}(V_L) = \sum_{l=0}^{L-1} \sum_{i=1}^{\nu_l} \tilde{\xi}_i^{(l)}$. Different from the AWGN channel case, the mgf of $\tilde{\xi}_i^{(l)}$ for Rayleigh fading channels is given as follows.

**Proposition 6.5.2** The mgf of the truncated random variable $\tilde{\xi}_i^{(l)}$ as in (6–111) with $\tilde{\xi}_i^{(l)} \sim \mathcal{GAL}(\mu_l, \sigma_i^2, \tau_l)$ is given by

$$\hat{\Phi}_l(s) = E[e^{-s\tilde{\xi}_i^{(l)}}] = \frac{\Phi_l(s)h_l(s)}{P(|\tilde{\xi}_i^{(l)}| \geq T_l)},$$

(6–135)

where $\Phi_l(s)$ is mgf of the GAL random variable $\tilde{\xi}_i^{(l)}$, i.e.,

$$\Phi_l(s) = \frac{1}{(1 + \mu_l s - \frac{\sigma_i^2}{2} s^2)^{\tau_l}},$$

(6–136)
$h_l(s)$ is given as

$$h_l(s) = G(-T_l; \hat{\mu}_l, \hat{\sigma}_l^2, \tau_l) + G(-T_l; -\hat{\mu}_l, \hat{\sigma}_l^2, \tau_l), \quad (6–137)$$

with $G(\cdot; \mu, \sigma^2, \tau)$ the cdf of $\mathcal{GAL}(\mu, \sigma^2, \tau)$,

$$\hat{\mu}_l = \frac{\mu_l - \sigma_l^2 s}{1 + \mu_l s - \frac{\sigma_l^2}{2} s^2}, \quad \text{and} \quad \hat{\sigma}_l^2 = \frac{\sigma_l^2}{1 + \mu_l s - \frac{\sigma_l^2}{2} s^2}, \quad (6–138)$$

and the ROC is $\frac{\mu_l - \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2} < \alpha < \frac{\mu_l + \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2}$.

**Proof** Let $f_i(x)$ and $\hat{f}_i(x)$ denote the pdfs of $\tilde{\xi}_i^{(l)}$ and $\hat{\xi}_i^{(l)}$, respectively. Then, according to (6–111), we have

$$\hat{f}_i(x) = \begin{cases} \frac{f_i(x)}{P(|\tilde{\xi}_i^{(l)}| \geq T_l)} & |x| \geq T_l \\ 0 & -T_l < x < T_l \end{cases}, \quad (6–139)$$

Thus, the mgf of $\hat{\xi}_i^{(l)}$ is calculated as

$$\hat{\Phi}_l(s) = \int_{-\infty}^{+\infty} e^{-sx} \hat{f}_i(x) dx = \frac{1}{P(|\tilde{\xi}_i^{(l)}| \geq T_l)} \left( \int_{-\infty}^{-T_l} e^{-sx} f_i(x) dx + \int_{T_l}^{+\infty} e^{-sx} f_i(x) dx \right). \quad (6–140)$$

Since $\tilde{\xi}_i \sim \mathcal{GAL}(\mu_l, \sigma_l^2, \tau_l)$, from (6–2) and (6–3) we can see that its mgf $\Phi_l(s)$ takes the form of (6–135) with the ROC of $\frac{\mu_l - \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2} < \alpha < \frac{\mu_l + \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2}$. With the mgf $\Phi_l(s)$, the pdf and cdf of $\hat{\xi}_i^{(l)}$ can be represented as

$$f_i(x) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_l(s) e^{sx} ds, \quad (6–141)$$

and

$$G(x; \mu_l, \sigma_l^2, \tau_l) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_l(s) \frac{e^{sx}}{s} ds,$$

respectively, with $c > 0$ in the ROC.
Now we consider the first integral \( \int_{-\infty}^{-T_l} e^{-sx} f_l(x) dx \). By inserting (6–141), we can rewrite this integral as

\[
\int_{-\infty}^{-T_l} e^{-sx} f_l(x) dx = \int_{-\infty}^{-T_l} e^{-sx} \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_l(z) e^{zx} dz dx
\]

\[
= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_l(z) \int_{-\infty}^{-T_l} e^{(z-s)x} dx dz = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \Phi_l(z) \frac{e^{-T_l(z-s)}}{z-s} dz
\]

\[
= \frac{1}{2\pi j} \int_{c'-j\infty}^{c'+j\infty} \Phi_l(z+s) \frac{e^{-T_l z}}{z} dz, \tag{6–142}
\]

where \( c' = c - \Re(s) \). By substituting \( z + s \) for \( s \) in (6–136) with some manipulations, it turns out that \( \Phi_l(z+s) \) in (6–142) can be written as

\[
\Phi_l(z+s) = \Phi_l(s) \Psi_l(z, s), \tag{6–143}
\]

where \( \Psi_l(z) \) is given by

\[
\Psi_l(z, s) = \frac{1}{(1 + \hat{\mu}_l z - \frac{\hat{\sigma}_l^2}{2} z^2)^{\tau_l}}, \tag{6–144}
\]

and \( \hat{\mu}_l \) and \( \hat{\sigma}_l^2 \) are defined in (6–138). We note that for fixed \( s \), \( \Psi_l(z, s) \) is a GAL mgf, which gives

\[
\frac{1}{2\pi j} \int_{c'-j\infty}^{c'+j\infty} \Psi_l(z, s) \frac{e^{-T_l z}}{z} dz = G(-T_l; \hat{\mu}_l, \hat{\sigma}_l^2, \tau_l).
\]

Thus, (6–142) becomes

\[
\int_{-\infty}^{-T_l} e^{-sx} f_l(x) dx = \frac{1}{2\pi j} \int_{c'-j\infty}^{c'+j\infty} \Phi_l(z+s) \frac{e^{-T_l z}}{z} dz
\]

\[
= \frac{\Phi_l(s)}{2\pi j} \int_{c'-j\infty}^{c'+j\infty} \Psi_l(z, s) \frac{e^{-T_l z}}{z} dz = \Phi_l(s) G(-T_l; \hat{\mu}_l, \hat{\sigma}_l^2, \tau_l). \tag{6–145}
\]

In a similar way, we can show that the integral \( \int_{-\infty}^{T_l} e^{-sx} f_l(x) dx \) in (6–140) is given by

\[
\int_{-\infty}^{-T_l} e^{-sx} f_l(x) dx = \Phi_l(s) G(-T_l; -\hat{\mu}_l, \hat{\sigma}_l^2, \tau_l). \tag{6–146}
\]
Thus, by inserting (6–145) and (6–146) into (6–140), we obtain (6–135). Since the truncation does not change the convergence property of mgf [49], the ROC of \( \hat{\Phi}_t(s) \) is the same as that of \( \Phi_t(s) \).

With \( \Phi_Y(s) \) and \( \hat{\Phi}_t(s) \) given in (6–129) and (6–135) for Rayleigh fading channels, following the derivation in AWGN channel case, we obtain the mgf \( \hat{\Phi}(s) \) of the random variable \( \hat{S}(V_L) + Y'_d \) as

\[
\hat{\Phi}(s) = \frac{\Phi(s)}{P(T_{\xi}(V_L))}, \tag{6–147}
\]

where \( \Phi(s) \) is given by

\[
\Phi(s) = \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_l^{\nu_l}(s) h_l^{\nu_l}(s), \tag{6–148}
\]

and the ROC is \( \alpha_1 < \alpha < \alpha_2 \) with

\[
\alpha_1 = \max_{0 \leq l < j} \left\{ \frac{1 - \sqrt{1 + 4/L_c}}{2}, \frac{\mu_l - \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2} \right\}, \tag{6–149}
\]

and

\[
\alpha_2 = \min_{0 \leq l < j} \left\{ \frac{1 + \sqrt{1 + 4/L_c}}{2}, \frac{\mu_l + \sqrt{\mu_l^2 + 2\sigma_l^2}}{\sigma_l^2} \right\}. \tag{6–150}
\]

Then, by using (6–110) and (6–147), we can obtain the saddlepoint approximation and Chernoff bound for the probability \( P(S(V_L) + Y'_d < 0, T_{\xi}(V_L)) \), respectively. For saddlepoint approximation, similar to (6–134) we have

\[
P(S(V_L) + Y'_d < 0, T_{\xi}(V_L)) \approx \sqrt{\frac{\Phi(\alpha_S)}{2\pi \Phi''(\alpha_S)} \Phi(\alpha_S)} \Phi(\alpha_S), \tag{6–151}
\]

where the saddlepoint \( \alpha_1 < \alpha_S < \alpha_2 \) is the solution to \( \Phi'(\alpha) = 0 \), and \( \Phi'(\cdot) \) and \( \Phi''(\cdot) \) are the first and second derivatives of the mgf \( \Phi(\cdot) \) given in (6–148). For Chernoff bound, we have

\[
P(S(V_L) + Y'_d < 0, T_{\xi}(V_L)) < \Phi(\alpha_C), \tag{6–152}
\]

where \( \alpha_C = \arg \min_{0 < \alpha < \alpha_2} \Phi(\alpha) \), for which we note that \( \alpha_C = \alpha_S \) if \( \alpha_S > 0 \). For \( V_L \) with \( \sum_{l=0}^{L-1} \nu_l = 1 \), we adopt the saddlepoint approximation (6–151) to evaluate
\[ P(S(V_L) + Y_d' < 0, \mathcal{T}_\xi(V_L)) \text{ for accuracy. For } V_j \text{ with } \sum_{l=0}^{L-1} \nu_l \geq 2, \text{ we consider using the Chernoff bound (6–152) to upper bound the probability. In this case, in order to reduce the complexity of finding } \alpha_C, \text{ we loosen (6–152) as} \]

\[ P(S(V_L) + Y_d' < 0, \mathcal{T}_\xi(V_L)) < \Phi(\tilde{\alpha}_C), \quad (6–153) \]

where \( \tilde{\alpha}_C \) is the saddle point of \( \tilde{\Phi}(s) = \Phi_Y(s) \prod_{l=0}^{L-1} \Phi_l^{\nu_l}(s) \). The above methods can be used to handle the probability \( P(\tilde{\Gamma}_d(V_j') < 0, \mathcal{T}_\xi(V_j')) \) as well.

Finally, for the probability \( P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, \mathcal{T}_\xi(V_j)) \) with \( l \leq j - 2 \), we have shown in (6–125) and (6–126) that it takes value of 0 for any \( V_j \) with \( \{\nu_l\}_{l=1}^{j-1} = \{0\} \). When \( \{\nu_l\}_{l=1}^{j-1} \neq \{0\} \), we separate it into two cases similar to the scenario of AWGN channel. For \( V_j \) with \( \sum_{l=1}^{j-1} \nu_l \geq 2 \) and \( \{\nu_l\}_{l=1}^{j-1} \neq \{0\} \), we upper bound the probability as in (6–128). For the remaining case of \( V_j \), i.e., \( \{\nu_l\}_{l=0}^{j-1} = \{0\} \) and \( \sum_{l=1}^{j-1} \nu_l = \nu_r = 1 \) for \( l \leq r < j \), according to (6–127) we have

\[ P(\tilde{\Gamma}_d(V_j) < 0, \tilde{\Gamma}_d(V_l) \geq T_l, \mathcal{T}_\xi(V_j)) < P(Y_d + \tilde{\xi}_1^{(r)} < 0, \tilde{\xi}_1^{(r)} \leq -T_r). \quad (6–154) \]

To evaluate this bound, let \( \hat{\Phi}_\Delta(s) \) denote the mgf of random variable \( Y_d + \tilde{\xi}_1^{(r)} \) conditioned on \( \tilde{\xi}_1^{(r)} \leq -T_r \). Then similar to the derivation of (6–147), we have

\[ \hat{\Phi}_\Delta(s) = \frac{\Phi_\Delta(s)}{P(\tilde{\xi}_1^{(r)} \leq -T_r)}, \quad (6–155) \]

where

\[ \Phi_\Delta(s) = \frac{G(-T_r; \hat{\mu}_r, \hat{\sigma}_r^2, \tau_r)}{(1 + L_c s - L_c s^2)^d (1 + \mu_r - \frac{\sigma_g^2}{T} s^2)^\tau_r}, \quad (6–156) \]

and \( \hat{\mu}_r, \hat{\sigma}_r^2 \) as in (6–138). Again, similar to (6–151), with the saddlepoint approximation, we have

\[ P(Y_d + \tilde{\xi}_1^{(r)} < 0, \tilde{\xi}_1^{(r)} \leq -T_r) \approx \sqrt{\frac{\Phi_\Delta(\alpha_S)}{2\pi \Phi_\Delta''(\alpha_S)}} \frac{\Phi_\Delta(\alpha_S)}{\alpha_S}, \quad (6–157) \]

where \( \alpha_S \) is the saddle point of \( \Phi_\Delta(s) \) and \( \Phi_\Delta''(\cdot) \) is the second derivative of \( \Phi_\Delta(\cdot) \).
6.6 Numerical Results

In this section, we present numerical results to examine the validity of the BER upper bounds developed in this chapter. Recall that the performance analysis procedure is primarily based on the independence assumption and statistical distribution approximations for the extrinsic information with the density evolution model. The BER bounds, strictly speaking, are approximated upper bounds for the performance of the density evolution model. The assumptions idealize the iterative decoding and information exchange procedures, which may deviate from the actual situation in the realistic collaborative decoding process. For example, the independence assumption for extrinsic information tends to be invalid in very low SNR region. As a result, the performance predicted by our analysis may become too optimistic when compared with the actual collaborative decoding with MRB information exchange. Thus, in order to verify their tightness, we will compare the BER upper bounds to the simulation results from both of the realistic collaborative decoding process and the density evolution models.

First we set the number of exchanges $I$ to 3 (i.e., 3 exchanges and 4 decoding iterations are performed in total), set $M$ to 6, and compare the cases of different information exchange parameters for $\{p_j\} = \{0.1, 0.2, 0.6\}$, $\{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$, i.e., we vary the information exchange percentage in the last exchange. Fig. 6–6 and Fig. 6–7 compare the upper bounds in each iteration with the simulation results over an AWGN channel when the non-recursive convolutional code with the generation polynomial of $[1 + D^2, 1 + D + D^2]$ (denoted by CC(5, 7)) and $[1 + D^2 + D^3, 1 + D + D^2 + D^3]$ (denoted by CC(15, 17)) are used, respectively. Meanwhile, Fig.’s 6–8 and 6–9 show the comparison for CC(5, 7) and CC(15, 17) over an independent Rayleigh fading channel, respectively.

In the figures, we use CD and DEM to denote collaborative decoding and density evolution model for short, respectively. From the figures, we see that the proposed
Figure 6–6: Comparison of the bounds and simulation results for the cases of $M = 6$ on an AWGN channel, where $CC(5, 7)$ and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.

Figure 6–7: Comparison of the bounds and simulation results for the cases of $M = 6$ on an AWGN channel, where $CC(15, 17)$ and $\{p_j\} = \{0.1, 0.2, 0.6\}, \{0.1, 0.2, 0.8\}$ and $\{0.1, 0.2, 1\}$ are used.
Figure 6–8: Comparison of the bounds and simulation results for the cases of $M = 6$ on an independent Rayleigh fading channel, where $CC(5, 7)$ and $\{p_j\} = \{0.1, 0.2, 0.6\}$, \{0.1, 0.2, 0.8\} and \{0.1, 0.2, 1\} are used.

Figure 6–9: Comparison of the bounds and simulation results for the cases of $M = 6$ on an independent Rayleigh fading channel, where $CC(15, 17)$ and $\{p_j\} = \{0.1, 0.2, 0.6\}$, \{0.1, 0.2, 0.8\} and \{0.1, 0.2, 1\} are used.
upper bounds are very tight in the middle and high SNR regions (the bounds diverge in the low SNR region due to the nature of union bound) compared with the simulation results from the density evolution model, but slightly lower than those from the collaborative coding process. As mentioned above, this phenomenon is due to the idealized assumptions and approximations used in the analysis. However, we note that with increasing SNR, the performance of the actual collaborative coding process tends to approach that of the density evolution model. Thus, the BER bounds actually illustrate the performance of collaborative decoding process in the high SNR regions (or the low BER region). The consistence with the simulation results for different choices of \( \{p_j\} \), codes and channel models verifies the validity of our analysis and the tightness of the bounds. In the figures, we also show the union bounds for MRC. The BER upper bounds show that, for \( M = 6 \), collaborative decoding with MRB exchange can achieve the performance about 2dB and 3dB within that of MRC at the BER of \( 10^{-10} \) for AWGN and Rayleigh fading channels, respectively, when \( \{p_j\} = \{0.1, 0.2, 1\} \) is used.

In Fig.s 6–10 and 6–11, we compare the bounds with the simulation results for different values of \( M \) over AWGN and Rayleigh fading channels. In the comparison, we fix the information exchange parameters to \( \{p_j\} = \{0.1, 0.2, 1\} \) and use CC(5, 7) as example. For clarity, we only show the results for the last decoding iterations for each \( M \). Similar to the previous comparison, the BER bounds match the simulation results from the density model very well. Again, at the low BER region, the BER curves of the actual collaborative decoding tends to merge with those of the density evolution model for different values of \( M \). Hence, the BER bounds approximately reflect the performance of the actual collaborative decoding in the low BER region. From the BER bounds, we see that with the choice of \( \{p_j\} \), collaborative decoding with MRB information exchange eventually can provide the performance very close to that of MRC for \( M = 2, 3, \) and 4. For Rayleigh fading channels, this means that
most of the diversity gain provided by the channel is achieved in the collaborative decoding process. According to Fig. 4–7, the average information exchange amount is only 39%, 31%, and 25% of MRC, respectively, in this case.

With the proposed BER bounds for collaborative decoding, we can illustrate the effect of different choices of \( \{p_j\} \) on the error performance in the very low BER region, which cannot be reached by simulations. Here, we fix the setting of \( M = 6 \) and rate 1/2 code CC(5, 7), and compare the 8 cases listed in Table 6–1, in which the average information exchange amount \( \Theta \) corresponding to each choice of \( \{p_j\} \) is calculated by (4–7), and is normalized by the information exchange amount of MRC \( \Theta_{\text{MRC}} \) (calculated by (4–6)) for the purpose of comparison.

In Fig. 6–12, we show the BER bounds of the last decoding iterations over an AWGN channel for all 8 cases. From the figure, we see that in case 1, collaborative decoding with MRB exchange gives the performance very close to MRC, but with the largest average information exchange amount compared to all other cases. Meanwhile,
Figure 6–11: Comparison of the proposed bounds, simulation results in the last iteration for the cases of $M = 2, 3$ and 4 on an independent Rayleigh fading channels, where $\text{CC}(5, 7)$ and $\{p_j\} = \{0.1, 0.2, 1\}$ are used.

Table 6–1: Different choices of $\{p_j\}$ and the corresponding average information exchange amount $\Theta$ with $M = 6$ for rate-1/2 $\text{CC}(5, 7)$ code. $\Theta$ is calculated with respect to the information exchange amount of MRC, $\Theta_{\text{MRC}}$.

<table>
<thead>
<tr>
<th>No. of exchanges</th>
<th>Value of ${p_j}_{j=0}^{I-1}$</th>
<th>Average info exchange amount $\Theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>$I = 1$</td>
<td>${1}$</td>
</tr>
<tr>
<td>Case 2</td>
<td>$I = 2$</td>
<td>${0.6, 0.8}$</td>
</tr>
<tr>
<td>Case 3</td>
<td>$I = 2$</td>
<td>${0.5, 1}$</td>
</tr>
<tr>
<td>Case 4</td>
<td>$I = 3$</td>
<td>${0.1, 0.2, 1}$</td>
</tr>
<tr>
<td>Case 5</td>
<td>$I = 3$</td>
<td>${0.1, 0.2, 0.8}$</td>
</tr>
<tr>
<td>Case 6</td>
<td>$I = 3$</td>
<td>${0.1, 0.2, 0.6}$</td>
</tr>
<tr>
<td>Case 7</td>
<td>$I = 3$</td>
<td>${0.2, 0.3, 1}$</td>
</tr>
<tr>
<td>Case 8</td>
<td>$I = 4$</td>
<td>${0.08, 0.12, 0.3, 1}$</td>
</tr>
</tbody>
</table>
Figure 6–12: Comparison of performance for $M = 6$ and CC(5,7) on an AWGN channel with the different choices of \( \{p_j\} \) in Table 6–1.

we note that due to $p_j = 1$ in their last exchanges, cases 3, 4, 7 and 8 significantly outperform cases 2, 5 and 6, and give performance about 2dB within that of MRC with much lower average information exchange amounts than that in case 1.

In Fig. 6–13, we compare the BER bounds of the last decoding iterations over an independent Rayleigh fading channel for the 8 cases. Similar to the case of AWGN channel, by setting $p_j = 1$ in the last exchanges, the collaborative decoding with MRB information exchange in cases 1, 3, 4, 7 and 8 significantly outperforms that in cases 2, 5 and 6. Especially, with the choices of \( \{p_j\} \) in cases 4 and 8, most of the diversity gain in the fading channel can be achieved by exchanging only 17.3% and 13.5% of the information amount in MRC, respectively. On the other hand, in cases 2, 5 and 6, the BER bounds have almost the same slope as that of single receiver in the high SNR region. This implies that by choosing $p_j < 1$ for the last exchanges, nearly no diversity gain can be obtained in collaborative decoding with MRB information exchange. However, significant performance gain can still be achieved by exchanging
the decoding information in this situation. From the figure, about 11dB, 9dB and 6dB gain are obtained in cases 2, 5, and 6, respectively, compared with a single receiver over the Rayleigh fading channel.

6.7 Summary

We have analyzed the bit error performance for collaborative decoding with MRB information exchange. A density evolution model is proposed to simplify the analysis based on the independence assumption for extrinsic information. With the Gaussian and GAL approximations, knowledge of the extrinsic information is obtained by simulating the proposed model over AWGN and Rayleigh fading channels.

Then the MRB information exchange process for each single data bit and decoding error events are analyzed according to the density evolution model. By combining the analysis with a generalized union bound for the max-log-MAP decoder, we derive a general BER upper bound for collaborative decoding with MRB information exchange. This union bound expressed in terms of the joint probabilities involving a
set of truncated independent random variables with the same distribution as that of the extrinsic information. Finally, the BER upper bound is evaluated by using the methods of moment generating function and saddlepoint approximation for AWGN channels and independent Rayleigh fading channels.

The BER upper bound provides an effective way to study the error performance of collaborative decoding with MRB information exchange, especially for the low BER region at which the BER can not be easily estimated by simulations. From the numerical results of the bound, we find that to obtain space diversity gain effectively, the information exchange parameter $p_j$ should be set to 1 in the last exchange.
CHAPTER 7
CONCLUSIONS AND FUTURE WORK

In this dissertation, we have studied a class of network-based iterative receiving diversity techniques known as collaborative decoding. These techniques are suitable for distributed arrays when error correction codes are used in the transmission process. Collaborative decoding achieves receiving diversity by exchanging decoding information among receiving nodes in a distributed array. By carefully selecting what decoding information to exchange, collaborative decoding can lower the amount of information that must be exchanged in the array, while providing performance close to that of maximum ratio combining (MRC). Based on the statistical characteristics of the output of maximum a posteriori decoders, we study two information exchange schemes for collaborative decoding: the least-reliable-bit (LRB) and most-reliable-bit (MRB) exchange schemes. Error performance of these two schemes under different transmission environments with different parameter settings is investigated and compared with Monte Carlo simulations.

To further study the collaborative decoding approach, theoretical analysis is carried out for the LRB and MRB information exchange schemes, respectively. For analytical tractability, we consider the cases in which nonrecursive convolutional codes are used in the system under the AWGN and Rayleigh fading channel models. The analysis is based on the assumption that the extrinsic information generated in the collaborating decoding process for nonrecursive convolutional codes can be approximately described by certain simple distributions. With the independence assumption for extrinsic information, we represent the collaborating decoding process by a density evolution model with a single MAP decoder, and propose a systematic method to evaluate the error performance of collaborating decoding analytically. The resulting
analysis shows that with proper choices of parameters, collaborative decoding can achieve full diversity and approach the theoretical performance bounds asymptotically.

On the other side, the analysis approach proposed in this work provides a new way for evaluating the bit error performance at each iteration in an iterative decoding process. This is significantly different from the conventional analysis approach for iterative decoding in that only the performance after a large number of decoding iterations (asymptotical performance) is considered in conventional analysis. Meanwhile, as another important contribution, the GAL approximation and related statistical methods introduced in this work for the extrinsic information generated by MAP decoding also provide a new statistical tool for the study of iterative decoding over independent Rayleigh fading channel models, which are readily employed to model practical wireless communication scenarios.

While extensive simulation study and theoretical analysis have been done for collaborative decoding, all this work should only be regarded as a preliminary study for collaborative decoding because there are still a lot of open problems left for future work, especially in the performance analysis aspects.

First of all, in our analysis, we rely on simulating the density evolution model to obtain the statistical parameters for the extrinsic information generated by the MAP decoder in the collaborative decoding process. This fact makes our analysis a semi-analytical one, which is not completely desirable. Unfortunately, to obtain the distribution and its parameters for the extrinsic information in MAP decoding analytically is a wide open problem. Necessary mathematical tools or good approximation methods for solving this problem seem still unavailable so far. Besides, due to the fact that the independence assumption for the extrinsic information may not be realistic, there is some small gap between our analysis results and those obtained in actual collaborative decoding with MRB information exchange. We hope that in
future work, this dependence can be considered in the performance analysis so that the gap can be reduced. In addition, we also hope that the analysis can be extended to the case of recursive convolutional codes.

Another important aspect is the design of parameters and their optimization for collaborative decoding. As an important purpose of performance analysis, it is desire to develop some useful design criteria for collaborative decoding based on minimizing the BER bounds or maximizing the diversity order of the bounds for Rayleigh fading channels. This may require further study of the mathematic property for the BER upper bounds proposed in this dissertation. However, this may be a difficult task due to the complicated form of the bounds. Meanwhile, unveiling the tradeoff relation between performance and information exchange amount analytically is also a touch but very meaningful task for the design of collaborative decoding systems in the future work.

Finally, further improving the performance of collaborative decoding with an as low as possible information exchange traffic load in the distribution array is also an important issue. In this work, we see that in some cases, the collaborative decoding with LRB information exchange provides better compromise in terms of performance and information exchange amount than that with MRB information exchange. However, in some other cases, we have the opposite situation. This implies that the LRB and MRB schemes may be far away from optimum. Thus, it is possible to combine the LRB and MRB information exchange schemes in some manner, or to develop some other information exchange approaches for exploiting the space diversity provided by the channels more efficiently. This may become an interesting direction in future research work.
APPENDIX A
RECTANGULAR PARITY-CHECK ENCODING AND DECODING

Rectangular parity check (RPC) codes are a special class of multidimensional parity check (MDPC) codes with the number of dimension $M$ equal to 2. Here, we describe the encoding and decoding algorithms of the general MDPC codes. The encoding and decoding algorithms for rectangular parity check codes can be immediately obtained by setting $M = 2$.

A.1 Multidimensional Parity-Check Encoding

MDPC codes are actually parallel concatenated signal parity check (SPC) codes following the structure of turbo codes. Based on the idea in [11], by generalizing the single parity check codes to multiple dimensional lattices geometrically, a class of MDPC codes can be constructed [12, 13].

Let $M$ ($M > 1$) denote the number of dimensions. Arrange the data bits into a $M$-dimensional hyper cube with size $A$ for each side, where $A$ is an integer larger than 1. By doing so, we can obtain a block of data bits $\{u_{i_1,i_2,\ldots,i_M}\}$ with block size of $A^M$ indexed by their positions along each dimension $(i_1, i_2, \ldots, i_M)$. Then apply single parity check codes to each layer (hyper plane) along each dimension of the hyper cube. For convenience, we will donate the bit $u_{i_1,i_2,\ldots,i_M}$ with $i_m = j$ as $u_{j}^{m}$. Thus, the $MA$ parity bits can be represented as

$$p_{m,j} = \sum_{i_1=1}^{A} \cdot \sum_{i_{m-1}=1}^{A} \sum_{i_{m+1}=1}^{A} \cdot \sum_{i_M=1}^{A} u_{j}^{m} \quad (A-1)$$

for $m = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, A$, where $\sum$ denotes modulo-2 sum. Appending the $MA$ parity bits $\{p_{m,j}\}$ to the $A^M$ data bits, we obtain the $M$-dimensional parity check codes. If we regard all the SPC codes along each dimension $m$ as a
component of MDPC codes, then the MDPC codes are formed by parallel concatenating all the $M$ components together. In another viewpoint, MDPC codes can also be thought as a punctured version of multidimensional product codes. Clearly, the code rate $R_c$ of MDPC is

$$R_c = \frac{1}{1 + \frac{M}{A^{M-1}}}.$$  \hspace{1cm} (A-2)

In general, MDPC codes are very high rate codes for large code block sizes.

### A.2 Iterative Multidimensional Parity-Check Decoding

The basic idea behind iterative decoding of concatenated codes using soft-in, soft-out decoders is to break up the decoding of a fairly complex and long code into steps where the transfer of soft information among the decoding steps guarantees almost no loss of information. It is well known that iterative decoding can achieve performance close to that of the maximum \textit{a posteriori} (MAP) rule.

Suppose the observation of $u_{i_1,i_2,\ldots,i_M}$ and $p_{m,j}$ at the decoder are $x_{i_1,i_2,\ldots,i_M}$ and $y_{m,j}$, respectively. Similar to the notation $u^m_j$ defined in II.A, we denote the $x_{i_1,i_2,\ldots,i_M}$ with $i_m = j$ as $x^m_j$. By applying the iterative SISO decoding rule to each SPC suggested in (A–1), the extrinsic information of each bit in log-likelihood ratio (LLR) form for the $m$th component codes at the $n$th iteration is given by

$$L^{m(n)}(u_{i_1,\ldots,i_{m-1},j,i_{m+1},\ldots,i_M}) = \left( \sum_{m,\neq} \left( L(x^m_j|u^m_j) + L^{(n)}(u^m_j) \right) \right) \boxplus L(y_{m,j}|p_{m,j});  \hspace{1cm} (A–3)$$

where $L^{(n)}(u)$ is the \textit{a priori} LLR of the data bit $u$, and $L(x|u)$ and $L(y|p)$ are the log-likelihood ratios of received symbols conditioned on the transmitted bits, which are also known as channel measurement or reliability value of the channel. Above, the symbol $\boxplus$ is defined as log-likelihood ratio addition, and the notation $\sum_{m,\neq}$ means operating the log-likelihood ratio addition over the LLRs corresponding to all the bits on the $j$th hyper plane along the $m$th dimension (i.e., $\{u^m_j\}$), except the one on the left-hand side of the equation indexed by $(i_1,\ldots,i_{m-1},j,i_{m+1},\ldots,i_M)$. A little different from [12] and [13], in order to fit the iterative demodulation and decoding
feature of the BICM system in Chapter 3, we also need to compute the extrinsic
information of parity bits

\[
L_e^{(n)}(p_{m,j}) = \sum_{\{x^m_j, u^m_j\}} \left( L(x^m_j|u^m_j) + L^{(n)}(u^m_j) \right).
\]  \hspace{1cm} (A–4)

As suggested in [11], the LLR sum can be approximated very well as

\[
\sum_{j=1}^{J} L(u_j) \approx \left( \prod_{j=1}^{J} \text{sign}(L(u_j)) \right) \cdot \min_{j=1,\ldots,J} |L(u_j)|.
\]  \hspace{1cm} (A–5)

With this approximation, the complexity of the decoding procedure in (A–3) and (A–4) becomes very low.

The a priori LLR \( L^{(n)}(u) \) is updated in the decoding for the \( m \)th component codes at the \( n \)th iteration in the following way

\[
L^{(n)}(u^m_j) = \sum_{k<m} L^{(n)}_e(u^m_j) + \sum_{k>m} L^{(n-1)}_e(u^m_j).
\]  \hspace{1cm} (A–6)

The soft output of iterative decoding after \( n \) iterations is given by

\[
L^{(n)}(\hat{u}) = L(x|u) + \sum_{m=1}^{M} L^{(m)}_e(u).
\]  \hspace{1cm} (A–7)

Making hard decision on the soft output \( L^{(n)}(\hat{u}) \), the estimate \( \hat{u} \) corresponding to the transmitted bit \( u \) can be obtained.

Owing to their concatenated structure, with low decoding complexity, MDPC codes have exhibited close to capacity performance at very high code rate.
APPENDIX B
PROOF OF EQUATION (5–28)

B.1 System Model

Let \( u \) and \( c \) denote a information bit sequence and the corresponding codeword generated by a nonrecursive convolutional code \( C: \ u \rightarrow c \), where \( u = (u_0, u_1, \cdots, u_i, \cdots) \), \( c = (c_0, c_1, \cdots, c_i, \cdots) \), and \( u_i, c_i \in \{0, 1\} \) are the data (or information) bit and coded bit, respectively.

We consider the BPSK modulation here. Thus, the transmitted signal \( x_i \) is defined as

\[
x_i = 1 - 2c_i,
\]

with \( x_i \in \{-1, 1\} \).

In the system, we use a memoryless independent fading channel model, which includes the additive white Gaussian noise (AWGN) channel as a special case, to describe the transmission environment between the source and destination. Then, the received signal \( y_i \) at the destination corresponding to the transmitted BPSK signal \( x_i \) at time instant \( i \) can be expressed as

\[
y_i = g_i x_i + n_i,
\]

where \( n_i \) for all \( i \) are i.i.d. zero-mean additive Gaussian random variables with variance \( E[|n_i|^2] = \sigma_n^2 \), and \( g_i \) is the channel fading gain. For AWGN channels \( g_i = 1 \), and for Rayleigh fading channels \( g_i \) all \( i \) are i.i.d. Rayleigh random variables with pdf of

\[
p_{g_i}(g) = 2ge^{-g^2}, \quad g \geq 0.
\]
We normalize the signal energy $E[|x_i|^2] = 1$. Thus, the average SNR is $1/\sigma_n^2$. In this channel model, we assume that perfect channel state information (CSI) is available and hence coherent detection is performed at the receiver. With this model the conditional pdf $p(y_i|x_i, g_i)$, for all $i$, with perfect CSI is given by

$$p(y_i|x_i, g_i) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left(-\frac{|y_i - g_i x_i|^2}{2\sigma_n^2}\right).$$  \hfill (B-4)

### B.2 Extrinsic Information in Max-log-MAP Decoding

#### B.2.1 Optimal Log-MAP Decoding

In order to study the extrinsic information for MAP decoding, we first consider the soft output of the MAP decoder. Let $L_k$ denote the soft output of a MAP decoder for the $k$th data bit, then according to the optimal log-MAP decoding algorithm, soft output $L_k$ is defined as the log-likelihood ratio of the a posteriori probability of the $k$th data bit, for arbitrary $k$, i.e. \cite{11},

$$L_k = \log \frac{P(\hat{u}_k = 0 | y, g)}{P(\hat{u}_k = 1 | y, g)} = \log \frac{P(y | \hat{u}_k = 0, g) P(\hat{u}_k = 0)}{P(y | \hat{u}_k = 1, g) P(\hat{u}_k = 1)} = \log \frac{\sum_{u : u_k = 0} p(y | u, g) P(\hat{u} = u)}{\sum_{u : u_k = 1} p(y | u, g) P(\hat{u} = u)},$$

where $\hat{u}_k$ and $\hat{u}$ are the decision (or estimation) of $u_k$ and $u$, $y = (y_0, y_1, \cdots, y_i, \cdots)$ and $g = (g_0, g_1, \cdots, g_i, \cdots)$ are the received signal sequence and the channel fading gain sequence, $C_k^+$ and $C_k^-$ are the set of all codewords pair $(u, c)$ that gives the decision of $u_k = 0$ and $u_k = 1$, $P(\hat{u} = u)$ and $p(y | c, g)$ are the $a$ priori probability of data bit sequence $u$ and the conditional pdf of signal sequence $y$ given the codeword $c$ and the channel fading gain sequence $g$, respectively.
Based on the assumption of independent data bits and memoryless channel model, we have

\begin{align}
P(\hat{u} = u) &= \prod_j P(\hat{u}_j = u_j), \quad \text{(B-6)} \\
p(y | c, g) &= \prod_i p(y_i | c_i, g_i). \quad \text{(B-7)}
\end{align}

With the BPSK modulation in (B-1) and the conditional pdf \( p(y_i | x_i, g_i) \) in (B-4), the conditional pdf \( p(y_i | c_i, g_i) \) can be expressed as

\[
p(y_i | c_i, g_i) = f(y_i, g_i) \exp \left( -L_c c_i g_i y_i \right), \quad \text{(B-8)}
\]

where the function \( f(y_i, g_i) \) defined as

\[
f(y_i, g_i) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp \left( -\frac{(y_i - g_i)^2}{2\sigma_n^2} \right)
\]

is independent of coded bit \( c_i \), and \( L_c \) defined as

\[
L_c = \frac{2}{\sigma_n^2}
\]

is known as the channel reliability measure. For convenience, define

\[
f(\underline{y}, g) = \prod_i f(y_i, g_i), \quad \text{(B-9)}
\]

then it is easy to see that \( f(\underline{y}, g) \) is independent of sequence \( c \). With this definition, the conditional pdf \( p(\underline{y} | c, g) \) can be rewritten as

\[
p(\underline{y} | c, g) = f(\underline{y}, g) \cdot \prod_i \exp \left( -L_c c_i g_i y_i \right) = f(\underline{y}, g) \cdot \exp \left( -L_c \sum_i c_i g_i y_i \right). \quad \text{(B-10)}
\]
Inserting (B–10) and (B–6) into (B–5) yields

\[ L_k = \log \left( \frac{\sum_{(u,c) \in C_k^+} \exp \left( -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right)}{\sum_{(u,c) \in C_k^-} \exp \left( -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right)} \right) \]

\[ = \log \left\{ \sum_{(u,c) \in C_k^+} \exp \left( -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right) \right\} \]

\[ - \log \left\{ \sum_{(u,c) \in C_k^-} \exp \left( -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right) \right\}. \]  

(B–11)

**B.2.2 Max-log-MAP Decoding**

In the suboptimal max-log-MAP decoding, the approximation

\[ \log \left( \sum_i a_i \right) \approx \max_i \{ \log a_i \} \]

is applied into the calculation of (B–11), and results in the following max-log-MAP soft output expression

\[ L_k = \max_{(u,c) \in C_k^+} \left\{ -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right\} \]

\[ - \max_{(u,c) \in C_k^-} \left\{ -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) \right\}. \]

(B–13)

We note that, the above equation will not be changed by introducing an arbitrary constant \( C \) in the following way

\[ L_k = \max_{(u,c) \in C_k^+} \left\{ -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) + C \right\} \]

\[ - \max_{(u,c) \in C_k^-} \left\{ -L_c \sum_i c_i g_i y_i + \sum_j \log P(\hat{u}_j = u_j) + C \right\}. \]

(B–14)

For convenience, we can choose

\[ C = -\sum_j \log P(\hat{u}_j = 0). \]  

(B–15)
Recall that \( u_j \in \{0, 1\} \), thus we have
\[
\sum_j \log P(\hat{u}_j = u_j) + C = - \sum_j \left( \log P(\hat{u}_j = 0) - \log P(\hat{u}_j = u_j) \right)
\]
\[
= - \sum_j u_j \log \frac{P(\hat{u}_j = 0)}{P(\hat{u}_j = 1)} = - \sum_j u_j \eta_j,
\]
where \( \eta_j \) defined as
\[
\eta_j = \log \frac{P(\hat{u}_j = 0)}{P(\hat{u}_j = 1)}
\]
is the a priori information of the \( k \)th data bit. Applying (B–16) and (B–17) to (B–14), the soft output of max-log-MAP decoding becomes
\[
L_k = \max_{(u, c) \in C_k^+} \left\{ - L_c \sum_i c_i g_i y_i - \sum_j u_j \eta_j \right\} - \max_{(u, c) \in C_k^-} \left\{ - L_c \sum_i c_i g_i y_i - \sum_j u_j \eta_j \right\}
\]
\[
= \max_{(u, c) \in C_k^+} \left\{ - L_c \sum_i c_i g_i y_i - \sum_j u_j \eta_j \right\} + \min_{(u, c) \in C_k^-} \left\{ L_c \sum_i c_i g_i y_i + \sum_j u_j \eta_j \right\}
\]
\[
= \max_{(u, c) \in C_k^+} \left\{ - \Lambda_{u, c} \right\} + \min_{(u, c) \in C_k^-} \left\{ \Lambda_{u, c} \right\},
\]
where \( \Lambda_{u, c} \) defined as
\[
\Lambda_{u, c} = L_c \sum_i c_i g_i y_i + \sum_j u_j \eta_j
\]
is the error event metric for \( (u, c) \) including the a priori information of the \( k \)th data bit, \( \eta_k \). Explicitly, we rewrite (B–19) as
\[
\Lambda_{u, c} = L_c \sum_i c_i g_i y_i + \sum_{j, j \neq k} u_j \eta_j + u_k \eta_k = \Gamma_{u, c} + u_k \eta_k,
\]
where
\[
\Gamma_{u, c} = L_c \sum_i c_i g_i y_i + \sum_{j, j \neq k} u_j \eta_j
\]
is the error event metric for \( (u, c) \) without including the a priori information of the \( k \)th data bit, \( \eta_k \). Recall that, for \( (u, c) \in C_k^+ \), \( u_k = 0 \), and for \( (u, c) \in C_k^- \), \( u_k = 1 \). Thus, with (B–20) we have
\[
\Lambda_{u, c} = \Gamma_{u, c}
\]
for \((\mathbf{u}, \mathbf{c}) \in C^+_k\), and

\[
\Lambda_{\mathbf{u}, \mathbf{c}} = \Gamma_{\mathbf{u}, \mathbf{c}} + \eta_k
\]

(B–23)

for \((\mathbf{u}, \mathbf{c}) \in C^-_k\), respectively.

By inserting (B–22) and (B–23) into (B–18), we have

\[
L_k = \max_{(\mathbf{u}, \mathbf{c}) \in C^+_k} \left\{ -\Gamma_{\mathbf{u}, \mathbf{c}} \right\} + \min_{(\mathbf{u}, \mathbf{c}) \in C^-_k} \left\{ \Gamma_{\mathbf{u}, \mathbf{c}} + \eta_k \right\}
\]

\[
= \max_{(\mathbf{u}, \mathbf{c}) \in C^+_k} \left\{ -\Gamma_{\mathbf{u}, \mathbf{c}} \right\} + \min_{(\mathbf{u}, \mathbf{c}) \in C^-_k} \left\{ \Gamma_{\mathbf{u}, \mathbf{c}} \right\} + \eta_k.
\]

(B–24)

With the above and according to the definition of extrinsic information \(\xi_k\) for the bit \(\hat{u}_k\), i.e.,

\[
L_k = \xi_k + \eta_k,
\]

(B–25)

it is easy to see that the extrinsic information \(\xi_k\) in max-log-MAP decoding can be represented as

\[
\xi_k = \max_{(\mathbf{u}, \mathbf{c}) \in C^+_k} \left\{ -\Gamma_{\mathbf{u}, \mathbf{c}} \right\} + \min_{(\mathbf{u}, \mathbf{c}) \in C^-_k} \left\{ \Gamma_{\mathbf{u}, \mathbf{c}} \right\},
\]

(B–26)

where the error event metric \(\Gamma_{\mathbf{u}, \mathbf{c}}\) is given by (B–21). Since \(u_j\) and \(c_i \in \{0, 1\}\), we can rewrite \(\Gamma_{\mathbf{u}, \mathbf{c}}\) as

\[
\Gamma_{\mathbf{u}, \mathbf{c}} = \sum_{i \in \{i : u_i = 1\}} \eta_i + L_c \sum_{i \in \{i : c_i = 1\}} g_i y_i,
\]

(B–27)

which is convenient for us to represent the following derivations in terms of the Hamming weights of \(\mathbf{u}\) and \(\mathbf{c}\).
We first consider the last $2j$-fold summation in $\sum_{W_j}$ defined by (6.79), i.e.,

\[ m_0 \sum_{n_0=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_j-1} n_0 M' \sum_{\tau_0=n_0}^{n_1} \cdots \sum_{\nu_{j-1}=n_{j-1}}^{n_j-1} M', \]

(C-1)

where $M' = M - 1$. Since the range of $\nu_l$ depends only on $n_l$ for each $l$, we can rewrite (C-1) as

\[ m_0 \sum_{n_0=0}^{m_1} \cdots \sum_{n_{j-1}=0}^{m_j-1} n_0 M' \sum_{\tau_0=n_0}^{n_1} \cdots \sum_{\nu_{j-1}=n_{j-1}}^{n_j-1} M', \]

(C-2)

which contains $j$ pairs of 2-fold sums with indices $\{n_l, \nu_l\}_{l=0}^{j-1}$. Then we consider switching the summation order within each pair of the sums. In the pair of sums with indices $n_l$ and $\nu_l$, $n_l$ ranges from 0 to $m_l$, and $\nu_l$ takes value from $n_l$ to $n_l M'$ for any given $n_l$. We denote these by $n_l \in [0, m_l]$ and $\nu_l \in [n_l, n_l M']$, respectively. Clearly, for all the possible values of $n_l$, $\nu_l$ runs through the region $[0, m_l M']$. On the other hand, for any fixed value of $\nu_l \in [0, m_l M']$, we have

\[ n_l \in \left[ \left\lfloor \frac{\nu_l}{M'} \right\rfloor, \nu_l \right] \cap [0, m_l] = \left[ \left\lfloor \frac{\nu_l}{M'} \right\rfloor, \min\{\nu_l, m_l\} \right]. \]

Thus, the summation order for $\{n_l, \nu_l\}$ can be switched as

\[ m_l \sum_{n_l=0}^{n_l M'} \sum_{\nu_l=n_l}^{n_l M'} \min\{\nu_l, m_l\}, \text{ for } 0 \leq l \leq j - 1. \]

(C-3)
Figure C–1: Summation order switch procedure for $\{\nu_l, m_l\}_{l=0}^{j-1}$.

Plug (C–3) into (C–2), and with some rearrangements, we obtain

$$\sum_{V_j} = \sum_{m_0=0}^{w_0} \sum_{m_1=0}^{w_1} \cdots \sum_{m_{j-1}=0}^{w_{j-1}} m_0 M' m_1 M' \cdots m_{j-1} M' \sum_{n_0=[\nu_0 M]}^{n_0} \sum_{n_1=[\nu_1 M]}^{n_1} \cdots \sum_{n_{j-1}=[\nu_{j-1} M]}^{n_{j-1}} w_0'' \cdots w_{j-1}''', \quad (C–4)$$

where $w_l''' = \min\{\nu_l, m_l\}$ for all $l$.

Now we consider the first $2j$-fold summation in (C–4). Similar to (C–2), we can rewrite the first $2j$-fold summation in (C–4) as $j$ pairs of 2-fold sums with each pair indexed by $m_l$ and $\nu_l$. Then by using the same argument for (C–3), we switch the summation order within each of these pairs. Thus, the first $2j$-fold summation in (C–4) is equivalent to

$$\sum_{V_j} = \sum_{\nu_0=0}^{w_0 M'} \sum_{\nu_1=0}^{w_1 M'} \cdots \sum_{\nu_{j-1}=0}^{w_{j-1} M'} \sum_{m_0=[\nu_0 M']}^{m_0} \sum_{m_1=[\nu_1 M']}^{m_1} \cdots \sum_{m_{j-1}=[\nu_{j-1} M']}^{m_{j-1}} w_{l+1}''', \quad (C–5)$$

Since $w_{l+1} = w_l - m_l = w_0 - \sum_{t=0}^{l} m_t$, the summation range of $\nu_l$ depends on $\{m_k\}$ for all $k < l$. In order to switch the summation order so that summation on $\{\nu_l\}_{l=0}^{j-1}$ can be placed in front of that on $\{m_l\}_{l=0}^{j-1}$, we use the procedure shown in Fig. C–1 to achieve the purpose as follows. In the first round of switches, we perform switch on the summation order for each pair of $\{m_l, \nu_{l+1}\}$ for $l = 0, 1, \cdots, j - 2$. Then
we switch the summation order for the pair of \( \{m_l, \nu_{l+2}\} \), for \( l = 0, 1, \cdots, j - 3 \), in the second round. In the \( k \)th round, we switch the order for each pair of \( \{m_l, \nu_{l+k}\} \), for \( l = 0, 1, \cdots, j - k - 1 \). By repeating this procedure until the \((j - 1)\)th round, the desired summation order can be obtained. To perform this procedure, we first consider a general 2-fold summation on the integer pair \( \{a, b\} \)

\[
\sum_{a=x}^{y} \sum_{b=0}^{(y-a)M'} f(a, b),
\]

where \( x \) and \( y \) (\( y \geq x \)) are arbitrary integers independent of \( a \) and \( b \), and \( f(a, b) \) is an arbitrary function of \( \{a, b\} \). Since \( b \in [0, (y-a)M'] \) for any \( a \in [x, y] \), it is easy to see that \( b \) runs through \( [0, (y-x)M'] \), and for any \( b \in [0, (y-x)M'] \), \( a \) runs through \( [x, y - \lceil \frac{b}{M'} \rceil] \). Thus, we have

\[
\sum_{a=x}^{y} \sum_{b=0}^{(y-a)M'} f(a, b) = \sum_{b=0}^{(y-x)M'} \sum_{a=x}^{y - \lceil \frac{b}{M'} \rceil} f(a, b). \tag{C–6}
\]

Now, we consider the procedure in Fig. C–1. For the first round switches, let \( a = m_l \), \( b = \nu_{l+1} \), \( x = \lceil \frac{\nu_l}{M'} \rceil \), and \( y = w_l \). By using \( w_{l+1} = w_l - m_l \) and (C–6), we have

\[
\sum_{m_l=\lfloor \frac{\nu_l}{M'} \rfloor}^{w_l} \sum_{\nu_{l+1}=0}^{w_{l+1}M'} = \sum_{\nu_{l+1}=0}^{w_l} \sum_{m_l=\lfloor \frac{\nu_l}{M'} \rfloor}^{w_{l+1}M'} \quad \text{for } l = 0, 1, \cdots, j - 2. \tag{C–7}
\]

Similarly, for the second round switches, let \( a = m_l \), \( b = \nu_{l+2} \), \( x = \lceil \frac{\nu_l}{M'} \rceil \), and \( y = w_l - \lceil \frac{\nu_{l+1}}{M'} \rceil \). Then by applying \( w_{l+1} = w_l - m_l \) and (C–6) to the summation on \( \{m_l, \nu_{l+2}\} \) obtained from (C–7), we have

\[
\sum_{m_l=\lceil \frac{\nu_l}{M'} \rceil}^{w_l - \lfloor \frac{\nu_{l+1}+1}{M'} \rfloor} \sum_{\nu_{l+2}=0}^{w_{l+1}M'} = \sum_{\nu_{l+2}=0}^{w_l - \lfloor \frac{\nu_{l+1}+1}{M'} - \lfloor \frac{\nu_{l+1}+2}{M'} \rfloor \rfloor} \sum_{m_l=\lceil \frac{\nu_l}{M'} \rceil}^{w_{l+1}M'} \quad \text{for } l = 0, 1, \cdots, j - 3. \tag{C–8}
\]
By using mathematical induction, we can easily prove that the switching in the \(k\)th \((1 \leq k \leq j - 1)\) round is given by

\[
\sum_{m_l=\left\lceil \frac{\nu_l}{M'} \right\rceil}^{\nu_l} \sum_{\nu_{l+k}=0}^{t} = \sum_{m_l=\left\lceil \frac{\nu_l}{M'} \right\rceil}^{\nu_l} \sum_{\nu_{l+k}=0}^{t} \sum_{m_l=\left\lceil \frac{\nu_l}{M'} \right\rceil}^{\nu_l} \text{for } l = 0, 1, \cdots, j-k-1.
\]  

(C–9)

As shown in Fig. C–1, the summation order will become \(\{\nu_0, \nu_1, \cdots, \nu_{j-1}, m_0, m_1, \cdots, m_{j-1}\}\) after the \((j - 1)\)th round of switches. We also note that the final summation range for \(\nu_k\) and \(m_{j-k-1}\) are calculated at the two edges of the \(k\)th round (i.e., \(l = 0\) and \(l = j - k - 1\) in (C–9)), respectively. Thus, the resulting summation is given by

\[
\sum_{\nu_0=0}^{w'_0} \sum_{\nu_1=0}^{w'_1} \cdots \sum_{\nu_{j-1}=0}^{w'_{j-1}} \sum_{m_0=\left\lceil \frac{\nu_0}{M'} \right\rceil}^{\nu_0} \sum_{m_1=\left\lceil \frac{\nu_1}{M'} \rceil}^{\nu_1} \cdots \sum_{m_{j-1}=\left\lceil \frac{\nu_{j-1}}{M'} \right\rceil}^{\nu_{j-1}},
\]  

(C–10)

where \(w'_0 = w_0\). This summation is equivalent to the first \(2j\)-fold summation in (C–4).

Hence, by replacing it with (C–10), we finally obtain

\[
\sum_{W_j} = \sum_{\nu_0=0}^{w'_0} \sum_{\nu_1=0}^{w'_1} \cdots \sum_{\nu_{j-1}=0}^{w'_{j-1}} \sum_{m_0=\left\lceil \frac{\nu_0}{M'} \right\rceil}^{\nu_0} \sum_{m_1=\left\lceil \frac{\nu_1}{M'} \rceil}^{\nu_1} \cdots \sum_{m_{j-1}=\left\lceil \frac{\nu_{j-1}}{M'} \right\rceil}^{\nu_{j-1}}.
\]  

(C–11)
APPENDIX D

NUMERICAL EVALUATION OF GAL CDF

Suppose that the random variable $X$ has a GAL distribution defined in (6–1), i.e., $X \sim \mathcal{GAL}(\mu, \sigma^2, \tau)$, for $\mu \in \mathbb{R}$ and $\sigma, \tau \geq 0$. Let $G(t; \mu, \sigma^2, \tau)$ denote the cdf of $X$, i.e.,

$$G(t; \mu, \sigma^2, \tau) = P(X \leq t), \quad t \in \mathbb{R}. \quad \text{(D–1)}$$

In general, no closed-form expression exists for the function $G(t; \mu, \sigma^2, \tau)$. Hence, we consider its numerical evaluation here.

Due to the complicated form of the GAL pdf, it is usually difficult to evaluate $G(t; \mu, \sigma^2, \tau)$ efficiently with high accuracy through direct numerical integration of the GAL pdf. To avoid this difficulty, we utilize the mixture representation of (6–4) for a GAL random variable to compute its cdf, i.e.,

$$X = \mu W + \sigma \sqrt{W} Z, \quad \text{(D–2)}$$

where $W$ and $Z$ are two statistically independent random variables with the properties that $Z \sim \mathcal{N}(0, 1)$ and $W$ has a Chi-square distribution with $2\tau$ degrees of freedom, i.e., the pdf of $W$ is

$$g(x) = \frac{x^{\tau-1}}{\Gamma(\tau)} e^{-x}, \quad x > 0. \quad \text{(D–3)}$$

With this mixture representation, we know that when $\sigma^2 = 0$ and $\tau = 0$, $X$ will reduce to a Gamma and Gaussian random variable, respectively. For these two cases, numerical evaluating the cdf of $X$ is well studied. Thus, we only consider the case of
σ², τ ≠ 0 here. For this case, we rewrite the cdf \( G(t; \mu, \sigma^2, \tau) \) as

\[
G(t; \mu, \sigma^2, \tau) = P(\mu W + \sigma \sqrt{W} Z < t) = P(Z < \frac{t - \mu W}{\sigma \sqrt{W}})
\]

\[
= \int_0^\infty Q\left(\frac{\mu x - t}{\sigma \sqrt{x}}\right) g(x) dx = \frac{1}{\Gamma(\tau)} \int_0^\infty Q\left(\frac{\mu x - t}{\sigma \sqrt{x}}\right)x^{\tau-1}e^{-x} dx, \tag{D-4}
\]

where \( Q(\cdot) \) is the Gaussian \( Q \)-function, i.e.,

\[
Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz. \tag{D-5}
\]

We note that the integral in (D–4) is of the typical form that can be computed by Gauss-Laguerre quadrature [50] efficiently. In this method, integrals of the form

\[
I(\alpha) = \int_0^\infty x^\alpha e^{-x} f(x) dx, \tag{D-6}
\]

for an arbitrary function \( f(x) \) bounded in \([0, +\infty)\), can be computed by a \( n \)-point quadrature as

\[
I(\alpha) \approx \sum_{i=1}^n w_i(\alpha) f(x_i(\alpha)), \tag{D-7}
\]

where \( x_i(\alpha) \) are the zeros (abscissas) of the associated Laguerre polynomial \( L_n^\alpha(x) \) with the weighting function

\[
w(x, \alpha) = x^\alpha e^{-x}, \tag{D-8}
\]

and \( w_i(\alpha) \) are the corresponding weights.

Now we separate the computation of (D–4) into two cases: \( \tau \geq 1 \) and \( 0 < \tau < 1 \).

For \( \tau \geq 1 \), in order to avoid the dependency of \( x_i(\alpha) \) and \( w_i(\alpha) \) on \( \tau \), we choose the weight function as \( w(x, 0) = e^{-x} \), and \( f(x) \) as

\[
f(x) = x^{\tau-1}Q\left(\frac{\mu x - t}{\sigma \sqrt{x}}\right). \tag{D-9}
\]
It is easy to check that for $\tau \geq 1$, $f(x)$ in (D–9) is bounded for $x \in [0, +\infty)$. Thus, the integral in (D–4) can be computed as

$$\int_0^\infty e^{-x}f(x)dx \approx \sum_{i=1}^{n} w_i(0)f(x_i(0)) \quad (\text{D–10})$$

where the Laguerre abscissas $x_i(0)$ and weights $w_i(0)$ can be found in the literature, e.g., [51].

For the case of $0 < \tau < 1$, we note that $f(x)$ defined in (D–9) has a singular point at $x = 0$ when $t > 0$. In this case, using (D–10) to calculate the integral will introduce large error. To avoid this problem, we use the weight function in (D–8) with $\alpha = \tau - 1$. Correspondingly, $f(x)$ becomes

$$f(x) = Q\left(\frac{\mu x - t}{\sigma \sqrt{x}}\right), \quad (\text{D–11})$$

which guarantees to be bounded. Then, the integral in (D–4) is given by

$$\int_0^\infty x^\alpha e^{-x}f(x)dx \approx \sum_{i=1}^{n} w_i(\alpha)f(x_i(\alpha)), \quad -1 < \alpha = \tau - 1 < 0.$$ 

For this integral, [52] gives an efficient method and the necessary coefficient tables to compute the Laguerre abscissas $x_i(\alpha)$ and weights $w_i(\alpha)$ by Chebyshev expansions.
REFERENCES


[52] P. Lambin and J. P. Vigneron, “Tables for the Gaussian computation of \( \int_{0}^{\infty} x^\alpha e^{-x} f(x) \, dx \) for values of \( \alpha \) varying continuously between \(-1\) and \(+1\),” *Mathematics of Computation*, vol. 33, no. 146, pp. 805–811, Apr. 1979.
BIOGRAPHICAL SKETCH

Xin Li received his B.S. and M.S. degrees in electrical engineering from Northwestern Polytechnical University, Xi’an, China, in 1996, and from the Shanghai Jiao Tong University, Shanghai, China, in 1999, respectively. He joined the Wireless Information Networking Group (WING) at the University of Florida Department of Electrical and Computer Engineering in 2000. He is currently pursuing the Ph.D. degree as a graduate research assistant. His research interests include diversity techniques in wireless communication, iterative coding/decoding, modulation/demodulation, equalization, channel estimation, timing and carrier synchronization, QoS and high throughput MAC protocols for WLAN, adaptive signal processing, and statistical signal processing.