Performance Analysis for Collaborative Decoding with Least-Reliable-Bit Exchange on AWGN Channels

Xin Li, Tan F. Wong, and John M. Shea
Wireless Information Networking Group
Department of Electrical & Computer Engineering
University of Florida
Gainesville, Florida 32611-6130, USA
Email: lix@ecel.ufl.edu, {twong, jshea}@ece.ufl.edu

Abstract

Collaborative decoding is an approach that achieves spatial gain and diversity by exchanging decoding information among a cluster of physically separated receivers. The least-reliable-bit (LRB) exchange scheme helps to lower the amount of information that must be exchanged in collaborative decoding, while providing performance close to maximal-ratio combining (MRC). In this paper, we analyze the error performance of collaborative decoding with the LRB exchange scheme when nonrecursive convolutional codes are used. The analysis is based on the observation that the extrinsic information generated in the collaborating decoding of these convolutional codes can be approximated by Gaussian random variables. A density evolution model based on a single maximum *a posteriori* decoder is used to obtain the statistical characteristics of the extrinsic information. With the statistical knowledge of the extrinsic information, we upper bound the error performance of the collaborative decoding process. Numerical results show that our analysis gives very good predictions of the bit error rate for collaborate decoding with LRB exchange. With properly chosen parameters, collaborative decoding can achieve the same error performance as MRC in the high signal-to-noise ratio region.

Index Terms

Collaborative decoding, distributed array, least-reliable-bit exchange, density evolution, iterative decoding analysis, max-log-MAP decoding

I. Introduction

Collaborative decoding with a distributed array, originally proposed in [1]– [4], allows a group of physically separated receivers (called nodes) to collaborate in their decoding process to obtain spatial gain and diversity. In the collaborative decoding process, these nodes exchange a small portion of their decoding information with each other, and the information from other nodes is used as *a priori* information in future decoding. The exchange and decoding process is then repeated in an iterative fashion. An information exchange scheme named least-reliable-bit (LRB) exchange was also proposed for the collaborating decoding technique in [3]. The LRB exchange scheme allows each node in the distributed array to request extrinsic information from other nodes for its unreliable decoded information bits in each iteration. It was demonstrated in [3] that collaborative decoding with the LRB exchange scheme can provide significant savings on

*This work was supported by the National Science Foundation under Grant ANI-0220287 and the Office of Naval Research under Grant N000140210554.

the cost of information exchange while still achieving performance close to that of maximal-ratio combining (MRC). Further, the collaborative decoding technique was investigated in [5] for spectral-efficient transmissions with high-order modulations.

Due to the exchange of extrinsic information in the process, knowledge of the statistical characteristics of extrinsic information from maximum *a posteriori* (MAP) decoders in collaborative decoding is important to its performance analysis. 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.

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 literature (e.g., [7] and [8]) 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 *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.

This paper is organized as follows. In Section II the system model and collaborative decoding technique with LRB exchange are described. In Section III, we model collaborative decoding as the concatenation of a MAP decoder and an information exchanging device, and employ Gaussian approximation to obtain the density evolution of the extrinsic information. In Section IV, we derive an upper bound on the BER of the collaborative decoding process. Numerical results from the analysis are shown in Section V. Conclusions are given in Section VI.

[†]Unfortunately, for recursive convolutional codes the extrinsic information generated in this process can not be approximated by a simple distribution, which makes the performance analysis difficult. Due to this difficulty, we only consider nonrecursive convolutional codes in this paper, although the proposed scheme works for recursive codes as well as nonrecursive codes.

II. System Description

A. Distributed Array and Channel Model

The model of collaborative decoding with a distributed array proposed in [1]– [4] is shown in Fig. **??**. A remote source node transmits a 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 non-recursive convolutional code and BPSK modulation. Each node in $\mathcal{M}$ use the max-log-MAP decoding algorithm [6] to process the received symbols. All nodes can perform the demodulation and decoding process individually. The channels from the source to the receiving nodes are assumed to be independent identically distributed (i.i.d.) transmission channels. Meanwhile, the receiving nodes form a local network such that they can communicate with one another on an error-free *broadcast* channel. This assumption is reasonable when 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. The performance of the forward channel is the main concern of this paper.

The forward channels are nonfading additive white Gaussian noise (AWGN) channels. Let $y_{k,i}$ denote the received signal at the kth receiving nodes corresponding to the transmitted BPSK signal x_i (i.e., $x_i \in \{+1, -1\}$) at time instant i, then it can be expressed as $y_{k,i} = x_i + n_{k,i}$, where $n_{k,i}$ for $k \in \mathcal{M}$ and all i are i.i.d. zero-mean additive Gaussian random variables with variance $E[|n_{k,i}|^2] = \sigma_n^2$. We normalize the signal energy $E[|x_i|^2] = 1$. Thus, the average SNR is $1/\sigma_n^2$. With this model the pdf $p(y_{k,i}|x_i)$, for $k \in \mathcal{M}$ and all i, is given by

$$p(y_{k,i}|x_i) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\Big(- \frac{|y_{k,i} - x_i|^2}{2\sigma_n^2}\Big). \tag{1}$$

B. Collaborative Decoding with LRB Exchange

The idea of the LRB exchange scheme is that each node requests extrinsic information from other nodes for its least reliable decoded information bits [3]. All the information collected from other nodes, called *additional information*, is used as *a priori* information in the next decoding iteration. To weaken the correlation among the additional information at different iterations, once information about 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 jth $(0 \leq j \leq I-1)$ iteration by p_j ($0 \leq p_j \leq 1$), respectively. The analysis is based on the following procedures of the LRB exchange scheme:

1) Set all information 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 the 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 set in advance. The design of the parameters $\{p_j\}_{j=0}^{I-1}$ is an interesting topic but is outside the scope of this paper. In this paper, the focus is on evaluating the error performance for collaborative decoding given the node number M and $\{p_j\}_{j=0}^{I-1}$. Note that the soft output for non-systematic codes consists only of extrinsic information and *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.

III. Gaussian-Approximated Density Evolution For Nonrecursive Convolutional Codes

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. Thus 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. The collaborative decoding process can be modeled by the joint operation of an information exchange unit and a MAP decoder unit as shown in Fig. **??**. 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 [8]. For the collaborative decoding process described in Section II, the additional information obtained from the information exchanging process has a non-Gaussian distribution. Nevertheless, we observe that the probability distribution of the extrinsic information from the MAP decoder in each iteration is still well approximated as Gaussian when nonrecursive convolutional codes are employed. Fig. **??** shows the typical histograms of the extrinsic information generated by MAP decoders at successive iterations in the collaborative decoding process for nonrecursive convolutional codes. The histograms are very close to Gaussian distributions. Thus we apply the Gaussian-approximated density evolution technique in [8], [7] to predict the behavior of the MAP decoders in collaborative decoding.

As in [8], 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 extrinsic information 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 non-Gaussian inputs. Hence, we use simulation based on the model in Fig. **??** to quantify 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 IV.

We first describe the generation of the additional information. For the jth decoding iteration, let $\xi_{k,i}^{(j)}$ denote the extrinsic information generated by the MAP decoder for the ith information bit at node k, and let $\mathcal{B}_i^{(j)}$ 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 μ and variance σ^2. When the 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 kth 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 by

$$P(|\xi_{k,i}^{(j)}| \leq T_j \big| \mathcal{B}_i^{(j)}) = p_j. \tag{2}$$

Let $\lambda_{k,i}^{(j)}$ denote the additional information for the ith bit at the kth node generated by the LRB exchange process in the jth 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 $M \geq 3$; we discuss $M = 2$ separately later. According to the LRB scheme, if bit i is a non-candidate bit in the jth iteration, then $\lambda_{k,i}^{(j)} = 0$. Otherwise, there are three possibilities for a candidate bit:

i) No node requests information for the ith bit, i.e., $\bigcap_{t\in\mathcal{M}} |\xi_{t,i}^{(j)}| > T_j$, then $\lambda_{k,i}^{(j)} = 0$;

ii) The kth node does not request information for bit i, but there is one other node requesting information for that bit. We denote this event by $\dot{\mathcal{R}}_{k,i}^{(j)}$, i.e.,

$$\dot{\mathcal{R}}_{k,i}^{(j)} = \bigcup_{r\in\mathcal{M}, r\neq k} \Big\{ |\xi_{r,i}^{(j)}| \leq T_j, \bigcap_{t\in\mathcal{M}, t\neq r} |\xi_{t,i}^{(j)}| > T_j \Big\}. \tag{3}$$

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

$$\lambda_{k,i}^{(j)} = \sum_{\substack{t\in\mathcal{M} \\ t\neq r, t\neq k, r\neq k}} \xi_{t,i}^{(j)} \triangleq \dot{\lambda}_{k,i}^{(j)}; \tag{4}$$

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

$$\ddot{\mathcal{R}}_{k,i}^{(j)} = \{|\xi_{k,i}^{(j)}| \leq T_j\} \cup \Big\{ \{|\xi_{k,i}^{(j)}| > T_j\} \cap \overline{\bigcup_{r\in\mathcal{M}, r\neq k} \Big\{ |\xi_{r,i}^{(j)}| \leq T_j, \bigcap_{t\in\mathcal{M}, t\neq r, t\neq k} |\xi_{t,i}^{(j)}| > T_j \Big\}} \Big\}. \tag{5}$$

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

$$\lambda_{k,i}^{(j)} = \sum_{t\in\mathcal{M}, t\neq k} \xi_{t,i}^{(j)} \triangleq \ddot{\lambda}_{k,i}^{(j)}. \tag{6}$$

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

$$P\big(\mathcal{B}_i^{(j+1)}\big|\mathcal{B}_i^{(j)}\big) = P\Big(\bigcap_{k\in\mathcal{M}} |\xi_{k,i}^{(j)}| > T_j \Big| \mathcal{B}_i^{(j)}\Big) = (1-p_j)^M. \tag{7}$$

From this recursive relation, we immediately obtain

$$P(\mathcal{B}_i^{(j)}) = P(\mathcal{B}_i^{(j)}, \mathcal{B}_i^{(j-1)}, \cdots, \mathcal{B}_i^{(0)}) = \prod_{l=0}^{j-1} P\big(\mathcal{B}_i^{(l+1)}\big|\mathcal{B}_i^{(l)}\big) = \prod_{l=0}^{j-1}(1-p_l)^M, \tag{8}$$

and

$$\mathcal{B}_i^{(j)} = \bigcap_{l=0}^{j-1} \bigcap_{k\in\mathcal{M}} \{|\xi_{k,i}^{(l)}| > T_l\}. \tag{9}$$

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

$$P\big(\overline{\mathcal{B}_i^{(j+1)}}\big|\mathcal{B}_i^{(j)}\big) = P\big(\dot{\mathcal{R}}_{k,i}^{(j)} \cup \ddot{\mathcal{R}}_{k,i}^{(j)}\big|\mathcal{B}_i^{(j)}\big) = P\big(\dot{\mathcal{R}}_{k,i}^{(j)}\big|\mathcal{B}_i^{(j)}\big) + P\big(\ddot{\mathcal{R}}_{k,i}^{(j)}\big|\mathcal{B}_i^{(j)}\big). \tag{10}$$

With the above arguments, we can easily simulate the additional information generated in the actual LRB process for the density evolution model in Fig. **??**. Without loss of generality, we assume that the MAP decoder in Fig. **??** is in the Mth node. Also, we assume that the block length of the code is long enough to ensure the Gaussian approximations and thresholding. In the jth iteration, the MAP decoder generates $\xi_{M,i}^{(j)}$ for the ith bit. Then the values of μ_j and σ_j^2 of the extrinsic information for the information bits are estimated. To find T_j using (2), 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.,

$$F_j(x) = P(\xi_{k,i}^{(j)} < x \big| \mathcal{B}_i^{(j)}). \tag{11}$$

Then, according to (2) we have

$$p_j = F_j(T_j) - F_j(-T_j). \tag{12}$$

For $p_j < 1$, by solving (12) numerically we can obtain T_j approximately. For $p_j = 1$, 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 $\dot{\mathcal{R}}_{M,i}^{(j)}$ or $\ddot{\mathcal{R}}_{M,i}^{(j)}$ occurs, set $\lambda_{M,i}^{(j)}$ to $\dot{\lambda}_{M,i}^{(j)}$ or $\ddot{\lambda}_{M,i}^{(j)}$ 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 priori* information, denoted by $\eta_{M,i}^{(j+1)}$, for the $(j+1)$th iteration, as

$$\eta_{M,i}^{(j+1)}=\sum_{l=0}^{j} \lambda_{M,i}^{(l)} = \begin{cases} \lambda_{M,i}^{(l)} & \text{if } \mathcal{R}_{M,i}^{(l)} \neq \varnothing, \forall 0 \leq l \leq j \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

where $\mathcal{R}_{M,i}^{(l)} \triangleq \dot{\mathcal{R}}_{M,i}^{(l)} \cup \ddot{\mathcal{R}}_{M,i}^{(l)}$. By inputting this *a priori* information to the MAP decoder and iterating the above procedure, we can obtain the statistical characteristics of the Gaussian-approximated extrinsic information for the whole collaborative decoding process. Fig. **??** shows the mean and variance of the extrinsic information and the threshold T_j estimated in our density evolution model along with simulations results for the actual collaborative decoding process for $M = 6$. In the figure, the maximum free-distance, 4-state non-recursive convolution code is used, and $\{p_j\} = \{0.1, 0.15, 0.25\}$. The results show that our density evolution model well approximates the actual collaborative decoding process with only $1/M$th of the simulation load. In the next section, we show that to evaluate the error performance for I iterations, we only need statistical knowledge of the extrinsic information in the first $I-1$ iterations.

IV. 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 Mth 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 Mth node in following derivation. From the definition (4), 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 $M \geq 3$ first below.

A. BER Upper Bound for $M \geq 3$

For $M \geq 3$, the BER of the MAP decoders in the jth $(j > 0)$ 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.,

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

where $\xi^{(j)}$ is the extrinsic information, and $\eta^{(j)}$ is the *a priori* information in the jth iteration given in (13) at the Mth node, respectively. Here, we evaluate the error performance by finding an upper bound for (14).

According to (13) and (10), we rewrite (14) as

$$\begin{aligned} P_b^{(j)} &= P(\xi^{(j)} + \eta^{(j)} < 0, \overline{\mathcal{B}^{(j)}}) + P(\xi^{(j)} < 0, \mathcal{B}^{(j)}) \\ &= \sum_{l=0}^{j-1} P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} < 0, \mathcal{B}^{(j)}). \end{aligned} \tag{15}$$

We first consider the first part in (15). Using (10), (4) and (6), we have

$$\begin{aligned} P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) &= P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0, \ddot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) \\ &\le P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0). \end{aligned} \tag{16}$$

With (3) and (9), when $p_t < 1$ (i.e., $T_t < \infty$) for $0 \le t \le l$, we upper bound the first term in (16) as follows

$$\begin{aligned} P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) &= P\Big(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \bigcup_{r \in \mathcal{M}'} \Big\{ |\xi_r^{(l)}| \le T_l, \bigcap_{t \in \mathcal{M}, t \ne r} |\xi_t^{(l)}| > T_l \Big\}, \mathcal{B}^{(l)} \Big) \\ &= \sum_{r=1}^{M-1} P\Big(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, |\xi_r^{(l)}| \le T_l, \bigcap_{t \in \mathcal{M}, t \ne r} |\xi_t^{(l)}| > T_l, \bigcap_{t=0}^{l-1} \bigcap_{k \in \mathcal{M}} \{|\xi_k^{(t)}| > T_t\} \Big) \\ &\overset{\text{(a)}}{\le} \sum_{r=1}^{M-1} P\Big(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, |\xi_r^{(l)}| \le T_l, \bigcap_{t=0}^{l-1} |\xi_r^{(t)}| > T_t \Big) \qquad (17) \\ &\overset{\text{(b)}}{=} (M-1) P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0) P\Big(|\xi_1^{(l)}| \le T_l, \bigcap_{t=0}^{l-1} |\xi_1^{(t)}| > T_t \Big), \qquad (18) \end{aligned}$$

where (a) is obtained by dropping all the events in $\{\dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}\}$ associated with $\xi_k^{(t)}$ for all t and $k \in \mathcal{M},\ k \ne r$, and (b) is due to the fact that the probabilities in the sum in (17) are equal for $1 \le r \le M-1$, and that $\xi^{(j)}$ and $\dot{\lambda}^{(l)}$ are independent of $\xi_r^{(t)}$ for all t.

To evaluate the probability $P\Big(|\xi_1^{(l)}| \le T_l, \bigcap_{t=0}^{l-1} |\xi_1^{(t)}| > T_t\Big)$ in (18), we use (9) to rewrite $P(\mathcal{B}^{(j)})$ as

$$\begin{aligned} P(\mathcal{B}^{(j)}) &= P\Big(\bigcap_{k=1}^{M} \Big\{ \bigcap_{l=0}^{j-1} |\xi_k^{(l)}| > T_l \Big\} \Big) = \prod_{k=1}^{M} P\Big(\bigcap_{l=0}^{j-1} |\xi_k^{(l)}| > T_l \Big) \\ &= \Big[P\Big(\bigcap_{l=0}^{j-1} |\xi_k^{(l)}| > T_l \Big) \Big]^M,\ k \in \mathcal{M}. \end{aligned} \tag{19}$$

By comparing (19) with (8), for all $k \in \mathcal{M}$ we have

$$P\Big(\bigcap_{l=0}^{j-1} |\xi_k^{(l)}| > T_l\Big) = \prod_{l=0}^{j-1} (1-p_l). \tag{20}$$

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

$$P\Big(|\xi_k^{(l)}| \le T_l, \bigcap_{t=0}^{l-1} |\xi_k^{(t)}| > T_t\Big) = p_l \prod_{t=0}^{l-1} (1-p_t). \tag{21}$$

Thus, with (21) and taking into account the fact $P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) \le P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0)$, we refine the upper bound (18) as

$$P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) \le \min\Big\{1,\ (M-1)p_l \prod_{t=0}^{l-1} (1-p_t)\Big\} \cdot P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0). \tag{22}$$

This bound is for the case that all p_t are not equal to 1. If there exists a $0 \le t \le l$ such that $p_t = 1$, then $P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) = 0$ because of $P(\dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) = 0$. To include this case, we rewrite the upper bound (22) as

$$P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) \le a_l \cdot P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0), \tag{23}$$

where

$$a_l = \begin{cases} 0 & \text{if } \prod_{t=0}^{l} (1-p_t) = 0 \\ \min\Big\{1,\ (M-1)p_l \prod_{t=0}^{l-1} (1-p_t)\Big\} & \text{otherwise.} \end{cases} \tag{24}$$

In the same way, we consider the probability $P(\xi^{(j)} < 0, \mathcal{B}^{(j)})$ in (15). With (8) and (20), this probability can be easily expanded and upper bounded as

$$\begin{aligned} P(\xi^{(j)} < 0, \mathcal{B}^{(j)}) &= P\Big(\xi^{(j)} < 0, \bigcap_{l=0}^{j-1} \bigcap_{k \in \mathcal{M}} |\xi_k^{(l)}| > T_l\Big) \\ &= P\Big(\xi^{(j)} < 0, \bigcap_{l=0}^{j-1} |\xi^{(l)}| > T_l\Big) P\Big(\bigcap_{l=0}^{j-1} \bigcap_{k=1}^{M-1} |\xi_k^{(l)}| > T_l\Big) \\ &\le P(\xi^{(j)} < 0, |\xi^{(j-1)}| > T_{j-1}) \cdot \prod_{l=0}^{j-1} (1-p_l)^{M-1} \\ &\le b_j \big[P(\xi^{(j-1)} < -T_{j-1}) + P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1})\big], \end{aligned} \tag{25}$$

where

$$b_j = \prod_{l=0}^{j-1} (1-p_l)^{M-1}. \tag{26}$$

By inserting (23), (25) and (16) into (15), we obtain following upper bound

$$P_b^{(j)} \leq \sum_{l=0}^{j-1} \Big[a_l P(\xi^{(j)} + \dot{\lambda}^{(l)} < 0) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0) \Big] + b_j \big[P(\xi^{(j-1)} < -T_{j-1}) + P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1}) \big], \quad (27)$$

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

B. Union Bound for Max-log-MAP Decoding

Let $\underline{\boldsymbol{u}}$ and $\underline{\boldsymbol{c}}$ denote a information bit sequence and the corresponding codeword generated by a nonrecursive convolutional code $\mathcal{C}$: $\underline{\boldsymbol{u}} \rightarrow \underline{\boldsymbol{c}}$, where $\underline{\boldsymbol{u}} = (u_0, u_1, \cdots, u_i, \cdots)$, $\underline{\boldsymbol{c}} = (c_0, c_1, \cdots, c_i, \cdots)$, 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 (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 [6]

$$\xi_k^{(j)} = \max_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^+} \big\{ -\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \big\} + \min_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^-} \big\{ \Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \big\}, \quad (28)$$

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

$$\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} = \sum_{\substack{i \in \{i : u_i = 1\} \\ i \neq k}} \eta_i^{(j)} + L_c \sum_{i \in \{i : c_i = 1\}} y_i. \quad (29)$$

In (29), $i \in \{i : u_i = 1\}$ and $i \in \{i : c_i = 1\}$ mean taking the indices of the non-zero bits in $\underline{\boldsymbol{u}}$ and $\underline{\boldsymbol{c}}$, $\eta_i^{(j)}$ is the *a priori* information of the ith information bit, and

$$L_c = 2/\sigma_n^2 \quad (30)$$

is known as the channel reliability measure. Note that since the all-zero codeword $(\underline{\mathbf{0}}, \underline{\mathbf{0}}) \in C^+$ and $\Gamma^{(j)}_{\underline{\mathbf{0}}, \underline{\mathbf{0}}} = 0$, we have

$$\max_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^+} \big\{ -\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \big\} = \max_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^+} \big\{ 0, -\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \big\} \geq 0. \quad (31)$$

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

$$\begin{aligned} P(\xi_k^{(j)} < x) &= P\Big(\max_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{c}})\in C^+} \big\{ -\Gamma_{\underline{\boldsymbol{u}},\underline{\boldsymbol{c}}}^{(j)} \big\} + \min_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{c}})\in C^-} \big\{ \Gamma_{\underline{\boldsymbol{u}},\underline{\boldsymbol{c}}}^{(j)} \big\} < x \Big) \\ &\leq P\Big(\min_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{c}})\in C^-} \big\{ \Gamma_{\underline{\boldsymbol{u}},\underline{\boldsymbol{c}}}^{(j)} \big\} < x \Big) = \frac{1}{K_c} P\Big(\bigcup_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{c}})\in C^-} \Gamma_{\underline{\boldsymbol{u}},\underline{\boldsymbol{c}}}^{(j)} < x \Big) \\ &\leq \frac{1}{K_c} \sum_{(\underline{\boldsymbol{u}},\underline{\boldsymbol{c}})\in C^-} P(\Gamma_{\underline{\boldsymbol{u}},\underline{\boldsymbol{c}}}^{(j)} < x), \end{aligned} \tag{32}$$

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

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

$$\Gamma_{w,d}^{(j)} = \sum_{i=1}^{w-1} \eta_i^{(j)} + L_c \sum_{i=0}^{d-1} y_i, \tag{33}$$

where we have considered the bit u_0 in $\underline{\boldsymbol{u}}$ without loss of generality. Thus, by using (33) and dropping the subscript k, the union bound in (32) can be written as

$$P(\xi^{(j)} < x) \leq \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} w A_{w,d} P(\Gamma_{w,d}^{(j)} < x), \tag{34}$$

where $d_{\min}$ 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. (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 (34) with $x = 0$ and the *a priori* information equal to 0.

C. Applying Max-log-MAP Decoding Union Bound to Collaborative Decoding

To apply the generalized union bound in (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 (33). From (13), we know that in the jth decoding iteration, not all the $w-1$ non-zero information bits in $\underline{\boldsymbol{u}}$ (the first non-zero bit u_0 itself is excluded) can obtain the *a priori* information. Among those bits that obtain *a priori* information, some of them obtain $\dot{\lambda}_i^{(l)}$ while the other obtain $\ddot{\lambda}_i^{(l)}$ for $l < j$. For convenience, we use A_l to denote the bit set in the $w-1$ nonzero bits of $\underline{\boldsymbol{u}}$ that obtain additional information $\lambda_i^{(l)}$ in the lth iteration for $l < j$. This means that in the lth iteration the event $\mathcal{R}_i^{(l)}$ 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 $\underline{\boldsymbol{u}}$. Further, we define $\dot{A}_l$ as the subset of A_l that the event $\dot{\mathcal{R}}_i^{(l)}$ occurs for $i \in \dot{A}_l$, which means the information bits in $\dot{A}_l$ obtain additional information $\dot{\lambda}_i^{(l)}$ in the lth iteration. Also, we define $\ddot{A}_l$ as the complementary subset of A_l that the event $\ddot{\mathcal{R}}_i^{(l)}$ occurs for $i \in \ddot{A}_l$, i.e., those bits obtain additional information $\ddot{\lambda}_i^{(l)}$ in the lth iteration. Note due to the fact that no information can be exchanged for a non-candidate bit, we have $A_l \cap A_k = \varnothing$ for $l \neq k$. With these notations, the above event can be expressed as

$$\mathcal{V}_j = \Big\{ \bigcap_{l=0}^{j-1} \Big\{ \bigcap_{i \in \dot{A}_l} \dot{\mathcal{R}}_i^{(l)}, \bigcap_{i \in \ddot{A}_l} \ddot{\mathcal{R}}_i^{(l)}, \bigcap_{i \in A_l} \mathcal{B}_i^{(l)} \Big\}, \bigcap_{i \in B_j} \mathcal{B}_i^{(j)} \Big\}, \tag{35}$$

where $B_j = \overline{\bigcup_{l=0}^{j-1} A_l}$ 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 jth decoding iteration. From (33), the error event metric associated with event $\mathcal{V}_j$ can be defined as

$$\Gamma_{w,d}^{(j)} = \sum_{l=0}^{j-1} \sum_{i \in A_l} \lambda_i^{(l)} + Y_d = \sum_{l=0}^{j-1} \Big(\sum_{i \in \dot{A}_l} \dot{\lambda}_i^{(l)} + \sum_{i \in \ddot{A}_l} \ddot{\lambda}_i^{(l)} \Big) + Y_d, \tag{36}$$

where

$$Y_d \triangleq L_c \sum_{i=0}^{d-1} y_i. \tag{37}$$

and $Y_d \sim \mathcal{N}(dL_c, 2dL_c)$.

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 (36) and the probability of $\mathcal{V}_j$ only depends on the size of $\dot{A}_l$ 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

$$|A_l| = m_l, \text{ and } |\dot{A}_l| = n_l \tag{38}$$

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

$$|\ddot{A}_l| = m_l - n_l, \tag{39}$$

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

$$V_j = \big\{ |A_l| = m_l, |\dot{A}_l| = n_l \big\}_{l=0}^{j-1}. \tag{40}$$

For convenience, we use $\Gamma_{w,d}^{(j)}(V_j)$ to denote the error event metric with a particular V_j. Then we have $\Gamma_{w,d}^{(j)}(V_j) \sim \mathcal{N}\big(\mu(V_j), \sigma^2(V_j)\big)$, where

$$\mu(V_j) = dL_c + \sum_{l=0}^{j-1} \phi_l \mu_l, \text{ and } \sigma^2(V_j) = 2dL_c + \sum_{l=0}^{j-1} \phi_l \sigma_l^2, \tag{41}$$

in which

$$\phi_l = m_l(M-1) - n_l. \tag{42}$$

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

$$0 \leq m_l \leq w_l, \tag{43}$$

where

$$w_l = w_{l-1} - m_{l-1}, \text{ and } w_0 = w - 1, \tag{44}$$

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

$$P(\Gamma_{w,d}^{(j)} < x) = \sum_{V_j}^{(2j)} P\big(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j\big), \tag{45}$$

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

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

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

$$\prod_{l=0}^{j-1} \binom{w_l}{m_l} \binom{m_l}{n_l} \tag{46}$$

possible choices of $\mathcal{A}(V_j)$. For all these choices, the probabilities of the event $\{\mathcal{V}_j = \mathcal{A}(V_j)\}$, are the same. Thus, we can upper bound (45) by upper bounding $P\big(\Gamma_{w,d}^{(j)}(V_j) < x, \mathcal{V}_j = \mathcal{A}(V_j)\big)$

for each particular choice of $\mathcal{A}(V_j)$. In a manner similar to (17) and (25), we drop all the events associated with $\dot{\lambda}_i^{(l)}$ or $\ddot{\lambda}_i^{(l)}$ in $\mathcal{V}_j$ and use (20) and (21) to obtain

$$\begin{aligned}
P\big(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j\big) &= \prod_{l=0}^{j-1} \binom{w_l}{m_l}\binom{m_l}{n_l} P\big(\Gamma_{w,d}^{(j)}(V_j) < x, \mathcal{V}_j = \mathcal{A}(V_j)\big) \\
&\leq \prod_{l=0}^{j-1} \binom{w_l}{m_l}\binom{m_l}{n_l} P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big) \\
&\quad \times P\Big(\bigcap_{l=0}^{j-1}\bigcap_{i\in \dot{A}_l}\Big\{\bigcup_{r\in\mathcal{M}'}\{|\xi_{r,i}^{(l)}| < T_l, \bigcap_{t=0}^{l-1}|\xi_{r,i}^{(t)}| < T_t\}, \bigcap_{t=0}^{l}|\xi_{M,i}^{(t)}| > T_t\Big\}\Big) \\
&\quad \times P\Big(\bigcap_{l=0}^{j-1}\bigcap_{i\in \ddot{A}_l}\bigcap_{t=0}^{l-1}|\xi_{M,i}^{(t)}| > T_t\Big) P\Big(\bigcap_{i\in B_j}\mathcal{B}_i^{(j)}\Big) \\
&\leq c'(V_j) P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big), \qquad (47)
\end{aligned}$$

where $c'(V_j)$ is calculated as

$$c'(V_j) = \prod_{l=0}^{j-1} \binom{w_l}{m_l}\binom{m_l}{n_l}\Big\{[(M-1)p_l]^{n_l}(1-p_l)^{Mw_j+n_l}\prod_{t=0}^{l-1}(1-p_t)^{m_l+n_l}\Big\}. \qquad (48)$$

On the other hand, we know that $P\big(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j\big) \leq P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big)$. Then the upper bound in (47) can be refined as

$$P\big(\Gamma_{w,d}^{(j)}(V_j) < x, |\mathcal{V}_j| = V_j\big) \leq c(V_j) P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big), \qquad (49)$$

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

$$P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big) = Q\Big(\frac{\mu(V_j) - x}{\sigma(V_j)}\Big) \qquad (50)$$

with $Q(\cdot)$ being the Gaussian Q-function. Then by inserting (49) into (45) we have

$$P(\Gamma_{w,d}^{(j)} < x) \leq \sum_{V_j}^{(2j)} c(V_j) P\big(\Gamma_{w,d}^{(j)}(V_j) < x\big). \qquad (51)$$

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

$$P(\xi^{(j)} < x) \leq \frac{1}{K_c}\sum_{d\geq d_{\min}}\sum_{w\geq 1} wA_{w,d}\sum_{V_j}^{(2j)} c(V_j) Q\Big(\frac{\mu(V_j) - x}{\sigma(V_j)}\Big). \qquad (52)$$

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

Now, we only have $P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1})$ left in (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 (28). Let

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

denote the optimal decoding sequence found by the decoder in C^{+}, meanwhile $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})^{-}_{\text{opt}}$ denote the optimal decoding sequence in C^{-}. According to max-log-MAP decoding, the final survival sequence $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})_{\text{opt}}$ is generated between $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})^{+}_{\text{opt}}$ and $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})^{-}_{\text{opt}}$. If $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})^{+}_{\text{opt}}$ 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 [9]. Thus, under the assumption that the all-zero sequence $(\underline{\mathbf{0}}, \underline{\mathbf{0}})$ is transmitted, we have $(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}})^{+}_{\text{opt}} = (\underline{\mathbf{0}}, \underline{\mathbf{0}})$ since $(\underline{\mathbf{0}}, \underline{\mathbf{0}}) \in C^{+}$. That is,

$$\max_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{+}} \left\{ -\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \right\} = \Gamma^{(j)}_{\underline{\mathbf{0}}, \underline{\mathbf{0}}} = 0.$$

With the above arguments, we can rewrite (28) as follows by dropping the first term

$$\xi^{(j)} \approx \min_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} \left\{ \Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \right\}$$

when the SNR is high. Thus, for $j > 0$ we have

$$\begin{aligned} P(\xi_k^{(j)} < 0, \xi_k^{(j-1)} > T_{j-1}) &\leq P\Big(\min_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} \left\{ \Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \right\} < 0, \min_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} \left\{ \Gamma^{(j-1)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} \right\} > T_{j-1} \Big) \\ &= \frac{1}{K_c} P\Big(\bigcup_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} \Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} < 0, \bigcap_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} \Gamma^{(j-1)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} > T_{j-1} \Big) \\ &\leq \frac{1}{K_c} \sum_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} P\Big(\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} < 0, \bigcap_{(\underline{\boldsymbol{u}}', \underline{\boldsymbol{c}}') \in C^{-}} \Gamma^{(j-1)}_{\underline{\boldsymbol{u}}', \underline{\boldsymbol{c}}'} > T_{j-1} \Big) \\ &\leq \frac{1}{K_c} \sum_{(\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}) \in C^{-}} P(\Gamma^{(j)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} < 0, \Gamma^{(j-1)}_{\underline{\boldsymbol{u}}, \underline{\boldsymbol{c}}} > T_{j-1}), \end{aligned} \quad (53)$$

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

$$P(\xi_k^{(j)} < 0, \xi_k^{(j-1)} > T_{j-1}) \leq \frac{1}{K_c} \sum_{d \geq d_{\min}} \sum_{w \geq 1} \Big[w A_{w,d}$$

$$\times \sum_{V_j}^{(2j)} c(V_j) P\big(\Gamma^{(j)}_{w,d}(V_j) < 0, \Gamma^{(j-1)}_{w,d}(V_{j-1}) > T_{j-1} \big) \Big]. \quad (54)$$

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

$$\Gamma_{w,d}^{(j)}(V_j) = \Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi, \tag{55}$$

where

$$\Psi \triangleq \sum_{i \in \dot{A}_{j-1}} \dot{\lambda}_i^{(j-1)} + \sum_{i \in \ddot{A}_{j-1}} \ddot{\lambda}_i^{(j-1)}. \tag{56}$$

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

$$\begin{aligned}
&P\big(\Gamma_{w,d}^{(j)}(V_j) < 0, \Gamma_{w,d}^{(j-1)}(V_{j-1}) > T_{j-1}\big) = P\big(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Gamma_{w,d}^{(j-1)}(V_{j-1}) > T_{j-1}\big) \\
= \;& P\big(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Psi < -T_{j-1}\big) - P\big(\Gamma_{w,d}^{(j-1)}(V_{j-1}) < T_{j-1}, \Psi < -T_{j-1}\big) \\
= \;& P\big(\Gamma_{w,d}^{(j-1)}(V_{j-1}) + \Psi < 0, \Psi < -T_{j-1}\big) - P\big(\Gamma_{w,d}^{(j-1)}(V_{j-1}) < T_{j-1}\big) P(\Psi < -T_{j-1}) \\
= \;& Q\left(\frac{\mu(V_j)}{\sigma(V_j)}, \frac{\phi_{j-1}\mu_{j-1} + T_{j-1}}{\sqrt{\phi_{j-1}}\sigma_{j-1}}; \frac{\sqrt{\phi_{j-1}}\sigma_{j-1}}{\sigma(V_j)}\right) \\
&\quad - Q\Big(\frac{\mu(V_{j-1}) - T_{j-1}}{\sigma(V_{j-1})}\Big) Q\Big(\frac{\phi_{j-1}\mu_{j-1} + T_{j-1}}{\sqrt{\phi_{j-1}}\sigma_{j-1}}\Big),
\end{aligned} \tag{57}$$

where the relations

$$\mu(V_j) = \mu(V_{j-1}) + \phi_{j-1}\mu_{j-1}, \text{ and } \sigma^2(V_j) = \sigma^2(V_{j-1}) + \phi_{j-1}\sigma_{j-1}^2$$

are used, and

$$Q(x, y; \rho) \triangleq \int_x^\infty \int_y^\infty \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(-\frac{z_1^2 + z_2^2 - 2\rho z_1 z_2}{2(1-\rho^2)}\Big) dz_1 dz_2 \tag{58}$$

is the bivariate Gaussian Q-function, for which [10] gives a simplified expression for numerical evaluation. To this point, we have upper bounded the BER $P_b^{(j)}$ for $M \geq 3$.

D. BER Upper Bound for $M = 2$

For $M = 2$, we note that the additional information $\dot{\lambda}_{k,i}^{(j)}$ defined in (4) becomes 0 for all k, i and j, and the event $\ddot{\mathcal{R}}_{k,i}^{(j)}$ defined in (5) reduces to

$$\ddot{\mathcal{R}}_{k,i}^{(j)} = \{|\xi_{k,i}^{(j)}| \leq T_j\}. \tag{59}$$

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 Mth node. Following the notation in Section IV-A, the inequality in (16) becomes

$$\begin{aligned} P(\xi^{(j)} + \lambda^{(l)} < 0, \mathcal{R}^{(l)}, \mathcal{B}^{(l)}) &= P(\xi^{(j)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0, \ddot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) \\ &\leq P(\xi^{(j)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0). \end{aligned} \quad (60)$$

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

$$\begin{aligned} P(\xi^{(j)} < 0, \dot{\mathcal{R}}^{(l)}, \mathcal{B}^{(l)}) &= P(\xi^{(j)} < 0, |\xi_1^{(l)}| \leq T_l, |\xi^{(l)}| > T_l, \mathcal{B}^{(l)}) \\ &\leq P\Big(\xi^{(j)} < 0, |\xi_1^{(l)}| \leq T_l, |\xi^{(l)}| > T_l, \bigcap_{t=0}^{l-1} |\xi_1^{(t)}| > T_t\Big) \\ &= P(\xi^{(j)} < 0, |\xi^{(l)}| > T_l) P\Big(|\xi_1^{(l)}| \leq T_l, \bigcap_{t=0}^{l-1} |\xi_1^{(t)}| > T_t\Big) \\ &\leq a_l \big[P(\xi^{(l)} < -T_l) + P(\xi^{(j)} < 0, \xi^{(l)} > T_l)\big], \end{aligned} \quad (61)$$

where a_l is the same as (24). Thus, by substituting (16) with (60), the upper bound of $P_b^{(j)}$ in (27) becomes

$$\begin{aligned} P_b^{(j)} \leq & \sum_{l=0}^{j-1} \Big\{ a_l \big[P(\xi^{(l)} < -T_l) + P(\xi^{(j)} < 0, \xi^{(l)} > T_l)\big] + P(\xi^{(j)} + \ddot{\lambda}^{(l)} < 0) \Big\} \\ & + b_j \big[P(\xi^{(j-1)} < -T_{j-1}) + P(\xi^{(j)} < 0, \xi^{(j-1)} > T_{j-1})\big]. \end{aligned} \quad (62)$$

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

$$c'(V_j) = \prod_{l=0}^{j-1} \binom{w_l}{m_l} \binom{m_l}{n_l} \Big[p_l^{m_l} (1 - p_l)^{2w_j + n_l} \prod_{t=0}^{l-1} (1 - p_t)^{m_l + n_l} \Big]. \quad (63)$$

V. Numerical Results

In this section, we first present numerical results to demonstrate tightness of the BER upper bound developed in Section IV. 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. **??** compares the upper bounds in each iteration with the simulation results for $M = 2$ 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 (32). In the figure, we also show the union bounds for MRC. 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 within that of MRC for $M = 6$. This means that most of the spatial gain can be obtained through collaborative decoding.

In Fig. **??**, 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. **??**. We compare the upper bounds with simulation results for $M = 2, 3, 4, 6,$ and 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 III 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.

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. Since the information being exchanged are soft-outputs (in LLR form) for a portion of the information 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 that amount of exchanged information. The cost of overhead due to the protocol is ignored here. If we use Θ to denote the average information exchange amount, then with the independent assumption in Section III for a set of $\{p_j\}_{j=0}^{I-1}$ the Θ is given by

$$\Theta = MN \sum_{j=0}^{I-1} p_j \prod_{l=0}^{j-1} (1 - p_j)^M, \tag{64}$$

where N is the block size of information bits. Correspondingly, the information exchange amount of MRC is $\Theta_{\text{MRC}} = MN/R_c$ with R_c being the code rate, which will be used 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 I. In Table I, $\{p_j\}$ 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. **??**, 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. **??**, 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 choices of $\{p_j\}$, full spacial diversity can be achieved by the collaborative decoding technique.

VI. CONCLUSION

We analyzed the bit error performance for collaborative decoding with LRB exchange. A density evolution model was 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 derived an upper bound for the BER of the collaborative decoding process via a generalized union bound for the max-log-MAP decoder. Numerical results demonstrated the tightness of the bounds. We also showed 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 a collaborative decoding system.

The analysis was 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, only 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 [11] is a possible solution in this case.

REFERENCES

[1] T. F. Wong, X. Li and J. M. Shea, "Distributed decoding of rectangular parity-check code," *Electronics Letters*, vol. 38, no. 22, pp. 1364–1365, Oct. 2002.

[2] T. F. Wong, X. Li and J. M. Shea, "Iterative decoding in a two-node distributed array," in *Proc. IEEE Milcom'02*, Anaheim, CA, pp. 704.2.1–5 Oct. 2002.

[3] A. Avudainayagam, J. M. Shea, T. F. Wong, and X. Li, "Reliability exchange schemes for iterative packet combining in distributed arrays," in *Proc. IEEE WCNC'03*, New Orleans, LA, Mar. 2003.

[4] A. Avudainayagam, J. M. Shea, T. F. Wong, and X. Li, "Collaborative decoding on block-fading channels," submitted to *IEEE Trans. Commun.*.

[5] X. Li, T. F. Wong, and J. M. Shea, "Bit-interleaved rectangular parity-check coded modulation with iterative demodulation in a two-node distributed array," in *Proc. IEEE ICC'03*, Anchorage, AK, May 2003.

[6] J. Hagenauer, E. Offer, and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 429–445, Mar. 1996.

[7] D. Divsalar, S. Donlinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE J. Select. Areas Commun.*, vol. 19, no. 5, pp. 891–907, May. 2001.

[8] H. E. Gamal and A. R. Hammons, Jr., "Analyzing the turbo decoder using the Gaussian approximation", *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 671–686, Feb. 2001.

[9] L. Reggiani and G. Tartara, "Probability density function of soft information," *IEEE Commun. Lett.*, vol. 6, no. 2, pp. 52–54, Feb. 2002.

[10] M. Simon and D. Divsalar, "Some new twists to problem involving the Gaussian probability integral", *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 200–210, Feb. 1998.

[11] G. McLachlan and D. Peel, *Finite mixture models*, Wiley, 2000.

TABLE I

DIFFERENT CHOICES OF $\{p_j\}$ AND CORRESPONDING AVERAGE INFORMATION EXCHANGE AMOUNT Θ WITH $M = 8$ FOR RATE $1/2$ CC$(5,7)$ CODE. Θ IS CALCULATED WITH RESPECT TO THE INFORMATION EXCHANGE AMOUNT OF MRC, Θ_{MRC}.

	No. of exchanges	Value of $\{p_j\}_{j=0}^{I-1}$	Average info. exchange amount Θ
Case 1	$I = 3$	$\{0.1, 0.15, 0.25\}$	$\Theta_1 = 0.194\Theta_{\text{MRC}}$
Case 2	$I = 3$	$\{0.055, 0.098, 1\}$	$\Theta_2 = 0.396\Theta_{\text{MRC}}$
Case 3	$I = 5$	$\{0.0405, 0.0564, 0.0897, 0.1902, 1\}$	$\Theta_3 = 0.201\Theta_{\text{MRC}}$
Case 4	$I = 1$	$\{1\}$	$\Theta_4 = 0.5\Theta_{\text{MRC}}$

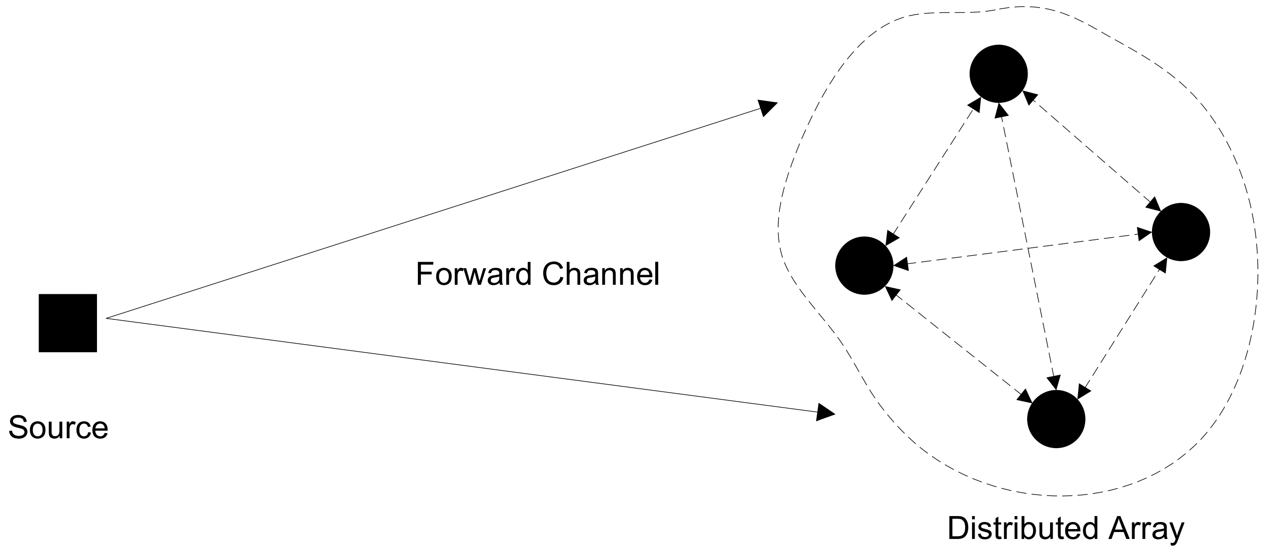


Fig. 1. Distributed array system model.

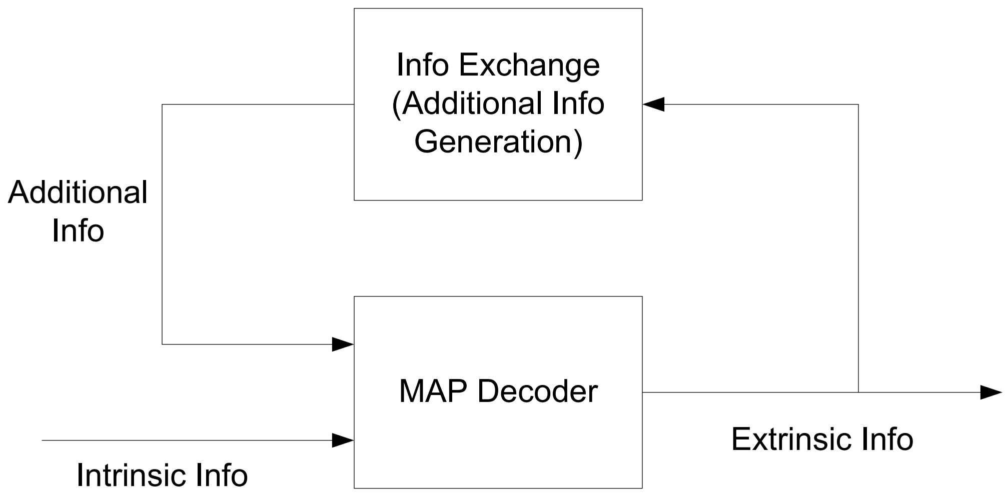


Fig. 2. System model for collaborative decoding process.

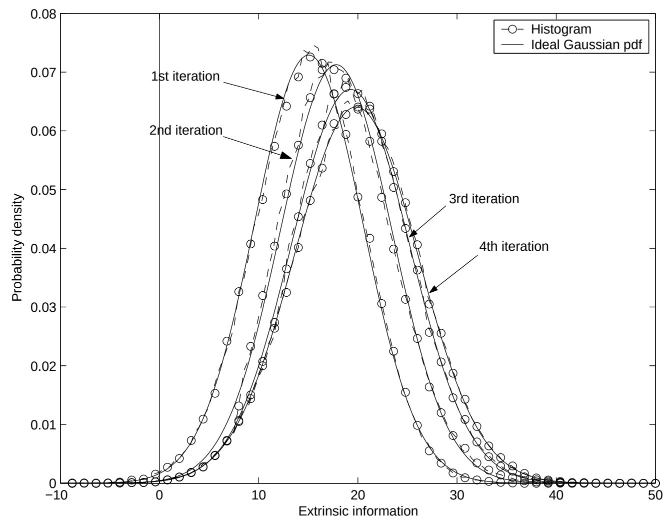
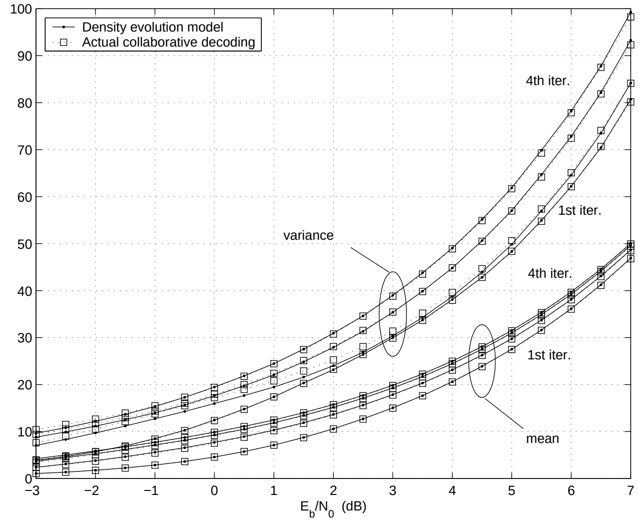
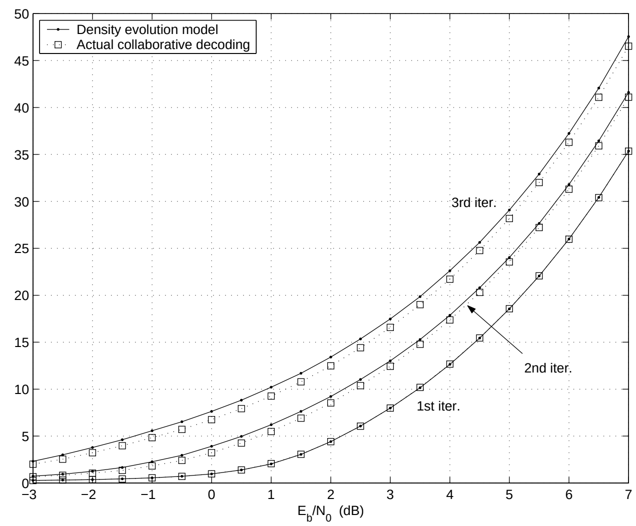


Fig. 3. Empirical pdfs of extrinsic information generated by the MAP decoders in successive iterations in collaborative decoding with the LRB exchange for $M = 6$ and $E_b/N_0 = 3$dB on AWGN channels, where the maximum free distance 4-state nonrecursive convolutional code is used.



(a) Mean and variance



(b) Threshold

Fig. 4. Mean and variance of the extrinsic information and the Threshold estimated from the density evolution model. They are compared with those from the actual collaborative decoding process.

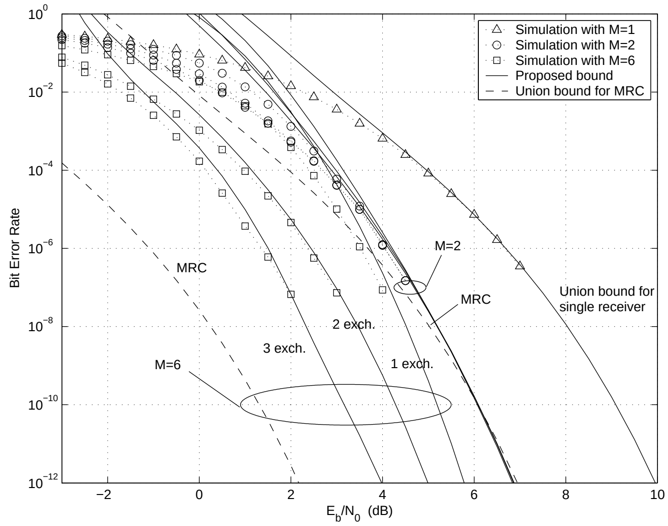


Fig. 5. Comparison of the proposed bounds, simulation results for $M = 2$ and 6 on AWGN channels, where CC$(5, 7)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used.

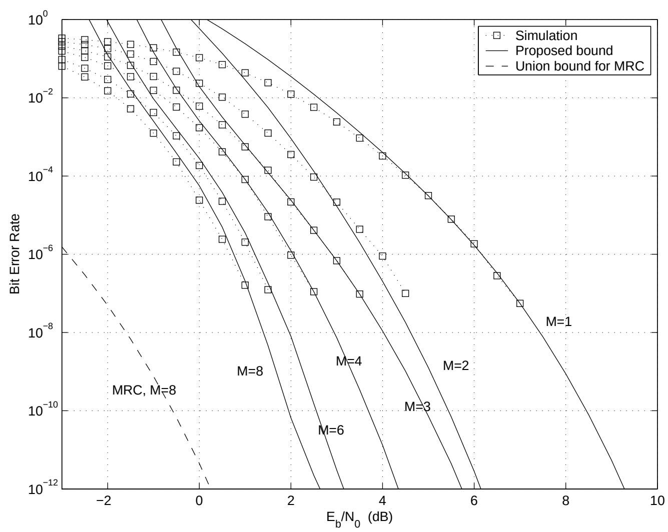
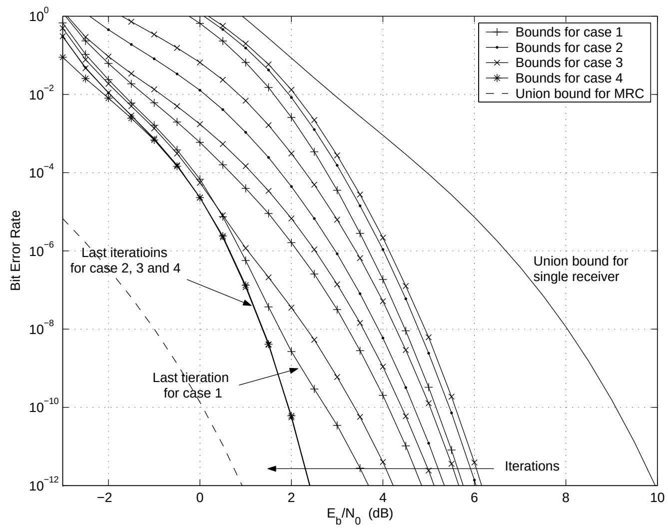
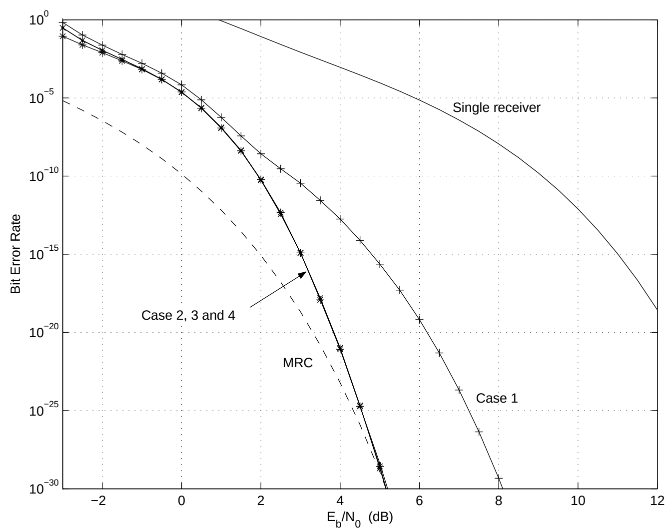


Fig. 6. Comparison of the proposed bounds, simulation results in the last iteration for $M = 2, 3, 4, 6$ and 8 on AWGN channels, where CC$(15, 17)$ and $\{p_j\} = \{0.1, 0.15, 0.25\}$ are used.



(a)



(b)

Fig. 7. Comparison of performance for $M = 8$ and $\mathrm{CC}(5,7)$ on AWGN channels with different choices of $\{p_j\}$ in Table I. (a) Comparison of BER bounds and simulation results. (b) Asymptotic performance of the 4 cases compared with MRC.