

Density Evolution for Asymmetric Memoryless Channels

Chih-Chun Wang, Sanjeev R. Kulkarni, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

Abstract—Density evolution (DE) is one of the most powerful analytical tools for low-density parity-check (LDPC) codes and graph codes with message passing decoding algorithms. With channel symmetry as one of its fundamental assumptions, density evolution has been widely and successfully applied to different channels, including binary erasure channels (BECs), binary symmetric channels (BSCs), binary additive white Gaussian noise (BiAWGN) channels, etc. This paper generalizes density evolution for *asymmetric* memoryless channels, which in turn broadens the applications to general memoryless channels, e.g., *z*-channels, composite white Gaussian noise channels, etc. The central theorem underpinning this generalization is the convergence to perfect projection for any fixed-size supporting tree. A new iterative formula of the same complexity is then presented and the necessary theorems for the performance concentration theorems are developed. Several properties of the new density evolution method are explored, including stability results for general asymmetric memoryless channels. Simulations, code optimizations, and possible new applications suggested by this new density evolution method are also provided. This result is also used to prove the typicality of linear LDPC codes among the coset code ensemble when the minimum check node degree is sufficiently large. It is shown that the convergence to perfect projection is essential to the belief propagation (BP) algorithm even when only symmetric channels are considered. Hence, the proof of the convergence to perfect projection serves also as a completion of the theory of classical density evolution for symmetric memoryless channels.

Index Terms—Asymmetric channels, density evolution (DE), low-density parity-check (LDPC) codes, rank of random matrices, sum-product algorithms, *z*-channels.

I. INTRODUCTION

SINCE the advent of turbo codes [1] and the rediscovery of low-density parity-check (LDPC) codes [2], [3] in the mid 1990s, graph codes [4] have attracted significant attention because of their capacity-approaching error correcting capability and the inherent low-complexity ($\mathcal{O}(n)$ or $\mathcal{O}(n \log(n))$, where n is the codeword length) of message passing decoding algorithms [3]. The near-optimal performance of graph codes is

Manuscript received August 15, 2003; revised August 22, 2005. This work was supported in part by the National Science Foundation under Grants CCR-9980590 and CCR-0312413, the Army Research Laboratory Communications Technology Alliance under Contract DAAD19-01-2-0011, the Army Research Office under Contract DAAD19-00-1-0466, and the New Jersey Center for Pervasive Information Technologies. The material in this paper was presented in part at the 3rd International Symposium on Turbo Codes and Related Topics, Brest, France, September 1–5, 2003, and in part at the 39th Annual Conference on Information Sciences and Systems, Baltimore, MD, March 16–18, 2005.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: chihw@princeton.edu; kulkarni@princeton.edu; poor@princeton.edu).

Communicated by G. Zémor, Associate Editor for Coding Theory.

Digital Object Identifier 10.1109/TIT.2005.858931

generally based on pseudorandom interconnections and Pearl's belief propagation (BP) algorithm [5], which is a distributed message-passing algorithm efficiently computing *a posteriori* probabilities in cycle-free inference networks. Turbo codes can also be viewed as a variation of LDPC codes, as discussed in [3] and [6].

Due to their simple arithmetic structure, completely parallel decoding algorithms, excellent error correcting capability [7], and acceptable encoding complexity [8], [9], LDPC codes have been widely and successfully applied to different channels, including binary erasure channels (BECs) [10]–[12], binary symmetric channels (BSCs), binary-input additive white Gaussian noise channels (BiAWGNs) [3], [13], Rayleigh-fading channels [14], Markov channels [15], partial response channels/intersymbol interference channels [16]–[19], and bit-interleaved coded modulation [21]. Except for the finite-length analysis of LDPC codes over the BEC [22], the analysis of iterative message-passing decoding algorithms is asymptotic (when the block length tends to infinity) [13], [23]. Under the optimal maximum-likelihood (ML) decoding algorithm, both the finite-length analysis and the asymptotic analysis for LDPC codes and other ensembles of turbo-like codes become tractable and rely on the weight distribution of these ensembles (see, e.g., [24]–[26]). Various Gallager-type bounds on ML decoders for different finite LDPC code ensembles have been established in [27].

In essence, the density evolution method proposed by Richardson *et al.* in [13] is an asymptotic analytical tool for LDPC codes. As the codeword length tends to infinity, the random codebook will be more and more likely to be cycle free, under which condition the input messages of each node are independent. Therefore, the probability density of messages passed can be computed iteratively. A performance concentration theorem and a cycle-free convergence theorem, providing the theoretical foundation of the density evolution method, are proved in [13]. The behavior of codes with block length $>10^4$ is well predicted by this technique, and thus degree optimization for LDPC codes becomes tractable. Near optimal LDPC codes have been found in [7] and [23]. In [16], Kavčić *et al.* generalized the density evolution method to intersymbol interference channels, by introducing the ensemble of *coset codes*, i.e., the parity-check equations are *randomly* selected as even or odd parities. Kavčić *et al.* also proved the corresponding fundamental theorems for the new coset code ensemble.

Because of the symmetry of the BP algorithm and the symmetry of parity-check constraints in LDPC codes, the decoding error probability will be independent of the transmitted

codeword in the symmetric channel setting. Thus, in [13], an all-zero transmitted codeword is assumed and the probability density of the messages passed depends only on the noise distribution. Nevertheless, in symbol-dependent asymmetric channels, which are the subject of this paper, the noise distribution is codeword dependent, and thus, some codewords are more noise resistant than others. As a result, the all-zero codeword cannot be assumed. Instead of using a larger coset code ensemble as in [16], we circumvent this problem by averaging over all valid codewords, which is straightforward and has practical interpretations as the averaged error probability. Our results apply to all binary input, memoryless, symbol-dependent channels (e.g., z-channels, binary asymmetric channels (BASCs), composite binary-input white Gaussian noise channels (composite BiAWGNCs), etc.) and can be generalized to LDPC codes over $\text{GF}(q)$ or \mathbb{Z}_m [28]–[30]. The theorem of convergence to *perfect projection* is provided to justify this codeword-averaged approach in conjunction with the existing theorems. New results on monotonicity, symmetry, stability (a necessary and a sufficient condition), and convergence rate analysis of the codeword-averaged density evolution method are also provided. Our approach based on the linear¹ code ensemble will be linked to that of the coset code ensemble [16] by proving the typicality of linear¹ LDPC codes when the minimum check node degree is sufficiently large, which was first conjectured in [21]. All of the above generalizations are based on the convergence to perfect projection, which will serve also as a theoretical foundation for the BP algorithms even when only symmetric channels are considered.

This paper is organized as follows. The formulations of and background on channel models, LDPC code ensembles, the BP algorithm, and density evolution, are provided in Section II. In Section III, an iterative formula is developed for computing the evolution of the codeword-averaged probability density. In Section IV, we state the theorem of convergence to perfect projection, which justifies the iterative formula. A detailed proof will be given in Appendix I. Monotonicity, symmetry, and stability theorems are stated and proved in Section V. Section VI consists of simulations and discussion of possible applications of our new density evolution method. Section VII proves the typicality of linear LDPC codes and revisits BP for symmetric channels. Section VIII concludes the paper.

II. FORMULATIONS

A. Symbol-Dependent Asymmetric Channels

The memoryless, symbol-dependent channels we consider here are modeled as follows. Let \mathbf{x} and \mathbf{y} denote a transmitted codeword vector and a received signal vector of codeword length n , where x_i and y_i are the i th transmitted symbol and

¹LDPC codes are, by definition, linear codes since only even parity-check equations are considered. Nonetheless, by taking both even and odd parity-check equations into consideration, the extended LDPC “coset” code has been proven to have important practical and theoretical value in many applications [16]. To be explicit on whether only even parity-check equations are considered or an extended set of parity-check equations is involved, two terms, “linear LDPC codes” and “LDPC coset codes,” will be used whenever a comparison is made, even though the adjective, linear, is redundant for traditional LDPC codes.

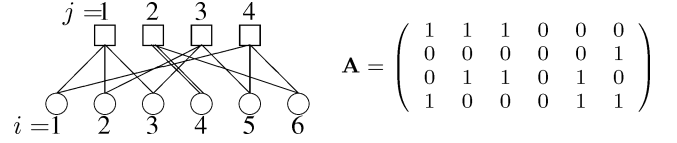


Fig. 1. A realization of the code ensemble $C^6(2, 3)$.

received signal, respectively, taking values in $\text{GF}(2)$ and the reals, respectively. The channel is memoryless and is specified by the conditional probability density function

$$f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^n f(y_i|x_i).$$

Two common examples are as follows.

Example 1: [Binary Asymmetric Channels (BASC)]

$$f(y|x) = \begin{cases} (1 - \epsilon_0)\delta(y) + \epsilon_0\delta(y - 1), & \text{if } x = 0 \\ \epsilon_1\delta(y) + (1 - \epsilon_1)\delta(y - 1), & \text{if } x = 1 \end{cases}$$

where ϵ_0, ϵ_1 are the crossover probabilities and $\delta(y)$ is the Dirac delta function. Note: if $\epsilon_0 = 0$, the above collapses to the z-channel.

Example 2: [Composite BiAWGNCs]

$$f(y|x) = \begin{cases} \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} \left(e^{-\frac{(y-3/\sqrt{5})^2}{2\sigma^2}} + e^{-\frac{(y+3/\sqrt{5})^2}{2\sigma^2}} \right), & \text{if } x=0 \\ \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} \left(e^{-\frac{(y-1/\sqrt{5})^2}{2\sigma^2}} + e^{-\frac{(y+1/\sqrt{5})^2}{2\sigma^2}} \right), & \text{if } x=1 \end{cases}$$

which corresponds to a bit-level subchannel of the 4 pulse amplitude modulation (4PAM) with Gray mapping.

B. Linear LDPC Code Ensembles

The linear LDPC codes of length n are actually a special family of parity-check codes, such that all codewords can be specified by the following even parity-check equation in $\text{GF}(2)$:

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

where \mathbf{A} is an $m \times n$ sparse matrix in $\text{GF}(2)$ with the number of nonzero elements linearly proportional to n . To facilitate our analysis, we use a code ensemble rather than a fixed code. Our linear code ensemble is generated by equiprobable edge permutations in a regular bipartite graph.

As illustrated in Fig. 1, the bipartite graph model consists of a bottom row of variable nodes (corresponding to codeword bits) and a top row of check nodes (corresponding to parity-check equations). Suppose we have n variable nodes on the bottom and each of them has d_v sockets. There are $m := \frac{nd_v}{d_c}$ check nodes on the top and each of them has d_c sockets. With these fixed $(n + m)$ nodes, there are a total of $(nd_v)!$ possible configurations obtained by connecting these $nd_v = md_c$ sockets on each side, assuming all sockets are distinguishable.² The re-

²When assuming all variable/check node sockets are indistinguishable, the number of configurations can be upper-bounded by $\frac{(nd_v)!}{(d_c!)^m}$.

sulting graphs (multigraphs) will be regular and bipartite with degrees denoted by (d_v, d_c) , and can be mapped to parity-check codes with the convention that the variable bit v is involved in parity-check equation c if and only if the variable node v and the check node c are connected by an odd number of edges. We consider a regular code ensemble $\mathcal{C}^n(d_v, d_c)$ putting equal probability on each of the possible configurations of the regular bipartite graphs described above. One realization of the codebook ensemble $\mathcal{C}^6(2, 3)$ is shown in Fig. 1. For practical interest, we assume $d_c > 2$.

For each graph in $\mathcal{C}^n(d_v, d_c)$, the parity-check matrix \mathbf{A} is an $m \times n$ matrix over $\text{GF}(2)$, with $A_{j,i} = 1$ if and only if there is an *odd* number of edges between variable node i and check node j . Any valid codeword \mathbf{x} satisfies the parity-check equation $\mathbf{Ax} = \mathbf{0}$. For future use, we let i and j denote the indices of the i th variable node and the j th check node. $\{j_{i_0,c}\}_{c \in [1, d_v]}$ denotes all check nodes connecting to variable node i_0 and similarly with $\{i_{j_0,v}\}_{v \in [1, d_c]}$.

Besides the regular graph case, we can also consider irregular code ensembles. Let λ and ρ denote the finite order *edge degree distribution* polynomials

$$\lambda(x) = \sum_k \lambda_k x^{k-1}$$

$$\rho(x) = \sum_k \rho_k x^{k-1}$$

where λ_k or ρ_k is the fraction of edges connecting to a degree k variable or check node, respectively. By assigning equal probability to each possible configuration of irregular bipartite graphs with degree distributions λ and ρ (similarly to the regular case), we obtain the equiprobable, irregular, bipartite graph ensemble $\mathcal{C}^n(\lambda, \rho)$. For example: $\mathcal{C}^n(3, 6) = \mathcal{C}^n(x^2, x^5)$.

C. Message Passing Algorithms and Belief Propagation

The message passing decoding algorithm is a distributed algorithm such that each variable/check node has a processor, which takes all incoming messages from its neighbors as inputs, and outputs new messages back to all its neighbors. The algorithm can be completely specified by the variable and check node message maps, Ψ_v and Ψ_c , which may or may not be stationary (i.e., the maps remain the same as time evolves) or uniform (i.e., node independent). The message passing algorithm can be executed sequentially or in parallel depending on the order of the activations of different node processors. Henceforth, we consider only parallel message passing algorithms complying with the *extrinsic* principle (adapted from turbo codes), i.e., the new message sending to node i (or j) does not depend on the received message from the same node i (or j) but depends only on other received messages.

A BP algorithm is a message passing algorithm whose variable and check node message maps are derived from Pearl's inference network [5]. Under the cycle-free assumption on the inference network, BP calculates the exact marginal *a posteriori* probabilities, and thus, we obtain the optimal maximum *a posteriori* probability (MAP) decisions. Let m_0 denote the initial message from the variable nodes, and $\{m_k\}$ denote the messages from its neighbors excluding that from the destination

node. The entire BP algorithm with messages representing the corresponding log-likelihood ratio (LLR) is as follows:

$$m_0 := \ln \frac{\text{P}(y_i | x_i = 0)}{\text{P}(y_i | x_i = 1)}$$

$$\Psi_v(m_0, m_1, \dots, m_{d_v-1}) := \sum_{j=0}^{d_v-1} m_j \quad (1)$$

$$\Psi_c(m_1, \dots, m_{d_c-1}) := \ln \left(\frac{1 + \prod_{i=1}^{d_c-1} \tanh \frac{m_i}{2}}{1 - \prod_{i=1}^{d_c-1} \tanh \frac{m_i}{2}} \right). \quad (2)$$

We note that the BP algorithm is based only on the cycle-free assumption³ and is actually independent of the channel model. The initial message m_0 depends only on the single-bit LLR function and can be calculated under asymmetric $f(y_i | x_i)$. As a result, the BP algorithm remains the same for memoryless, symbol-dependent channels.

Example: For BASCs

$$m_0 = \begin{cases} \ln \frac{1-\epsilon_0}{\epsilon_1}, & \text{if } y_i = 0 \\ \ln \frac{\epsilon_0}{1-\epsilon_1}, & \text{if } y_i = 1. \end{cases}$$

We assume that the BP is executed in parallel and each *iteration* is a "round" in which all variable nodes send messages to all check nodes and then the check nodes send messages back. We use l to denote the number of iterations that have been executed.

D. Density Evolution

For a symmetric channel and any message-passing algorithm, the probability density of the transmitted messages in each iteration can be calculated iteratively with a concrete theoretical foundation [13]. The iterative formula and related theorems are termed "density evolution." Since the BP algorithm performs extremely well under most circumstances and is of great importance, sometimes the term "density evolution" is reserved for the corresponding analytical method for BP algorithms.

III. DENSITY EVOLUTION: NEW ITERATIVE FORMULA

In what follows, we use the BP algorithm as the illustrative example for our new iterative density evolution formula.

With the assumption of channel symmetry and the inherent symmetry of the parity-check equations in LDPC codes, the probability density of the messages in any symmetric message passing algorithm will be codeword independent, i.e., for different codewords, the densities of the messages passed differ only in parities, but all of them are of the same shape [13, Lemma 1].

In the symbol-dependent setting, the symmetry of the channel may not hold. Even though the BP mappings remain the same for asymmetric channels, the densities of the messages for different transmitted codewords are of different shapes and the density for the all-zero codeword cannot represent the behavior when other codewords are transmitted. To circumvent this problem, we *average* the density of the messages over all valid codewords. However, directly averaging over all codewords takes 2^{n-m}

³An implicit assumption will be revisited in Section VII-B.

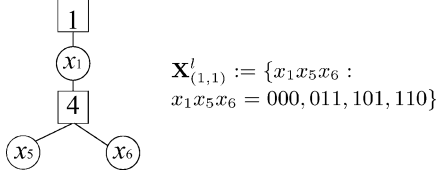


Fig. 2. Illustrations of $\mathcal{N}_{(1,1)}^{2l}$ and $\mathbf{X}_{(1,1)}^l$, $l = 2$.

times more computations, which ruins the efficiency of the iterative formula for density evolution. Henceforth, we provide a new iterative formula for the codeword-averaged density evolution which increases the number of computations only by a constant factor; the corresponding theoretical foundations are provided in this section and in Section IV.

A. Preliminaries

We consider the density of the message passed from variable node i to check node j . The probability density of this message is denoted by $P_{(i,j)}^{(l)}(\mathbf{x})$ where the superscript l denotes the l th iteration and the appended argument \mathbf{x} denotes the actual transmitted codeword. For example, $P_{(i,j)}^{(1)}(\mathbf{0})$ is the density of the initial message m_0 from variable node i to check node j assuming the all-zero codeword is transmitted. $P_{(i,j)}^{(2)}(\mathbf{0})$ is the density from i to j in the second iteration, and so on. We also denote by $Q_{(j,i)}^{(l)}(\mathbf{x})$ the density of the message from check node j to variable node i in the l th iteration.

With the assumption that the corresponding graph is tree-like until depth $2(l-1)$, we define the following quantities. Fig. 2 illustrates these quantities for the code in Fig. 1 with $i = j = 1$ and $l = 2$.

- $\mathcal{N}_{(i,j)}^{2l}$ denotes the tree-like subset of the graph⁴ $G = (\mathcal{V}, \mathcal{E})$ with root edge (i, j) and depth $2(l-1)$, named as the supporting tree. A formal definition is: $\mathcal{N}_{(i,j)}^{2l}$ is the subgraph induced by $\mathcal{V}_{(i,j)}^{2l}$, where

$$\mathcal{V}_{(i,j)}^{2l} := \{v \in \mathcal{V} : d(v, i) = d(v, j) - 1 \in [0, 2(l-1)]\} \quad (3)$$

where $d(v, i)$ is the shortest distance between node v and variable node i . In other words, $\mathcal{N}_{(i,j)}^{2l}$ is the depth $2(l-1)$ tree spanned from edge (i, j) . Let $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{V}}$ denote the number of variable nodes in $\mathcal{N}_{(i,j)}^{2l}$ (including variable node i). $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{C}}$ denotes the number of check nodes in $\mathcal{N}_{(i,j)}^{2l}$ (check node j is excluded by definition).

- $\mathbf{X} = \{\mathbf{x} \in \{0, 1\}^n : \mathbf{A}\mathbf{x} = \mathbf{0}\}$ denotes the set of all valid codewords, and the information source selects each codeword equiprobably from \mathbf{X} .
- \mathbf{x}_i and $\mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}}$ are the projections of codeword $\mathbf{x} \in \mathbf{X}$ on bit i and on the variable nodes in the supporting tree $\mathcal{N}_{(i,j)}^{2l}$, respectively.
- $\mathbf{X}_{(i,j)}^l$ denotes the set of all strings of length $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{V}}$ satisfying the $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{C}}$ check node constraints in $\mathcal{N}_{(i,j)}^{2l}$.

⁴The calligraphic \mathcal{V} in $G = (\mathcal{V}, \mathcal{E})$ denotes the set of all vertices, including both variable nodes and check nodes. Namely, a node $v \in \mathcal{V}$ can be a variable/check node.

\mathbf{x}^l denotes any element of $\mathbf{X}_{(i,j)}^l$ (the subscript (i, j) is omitted if there is no ambiguity). The connection between \mathbf{X} , the valid codewords, and $\mathbf{X}_{(i,j)}^l$, the tree-satisfying strings, will be clear in the following remark and in Definition 1.

- For any set of codewords (or strings) \mathbf{W} , the average operator $\langle \cdot \rangle_{\mathbf{W}}$ is defined as

$$\langle g(\mathbf{x}) \rangle_{\mathbf{W}} = \frac{1}{|\mathbf{W}|} \sum_{\mathbf{x} \in \mathbf{W}} g(\mathbf{x}).$$

- With a slight abuse of notation for $P_{(i,j)}^{(l)}(\mathbf{x})$, we define

$$P_{(i,j)}^{(l)}(x) := \left\langle P_{(i,j)}^{(l)}(\mathbf{x}) \right\rangle_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}_i = x\}}$$

$$P_{(i,j)}^{(l)}(\mathbf{x}^l) := \left\langle P_{(i,j)}^{(l)}(\mathbf{x}) \right\rangle_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}} = \mathbf{x}^l\}}.$$

Namely, $P_{(i,j)}^{(l)}(x)$ and $P_{(i,j)}^{(l)}(\mathbf{x}^l)$ denote the density averaged over all compatible codewords with projections being x and \mathbf{x}^l , respectively.

Remark: For any tree-satisfying string $\mathbf{x}^l \in \mathbf{X}_{(i,j)}^l$, there may or may not be a codeword \mathbf{x} with projection $\mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}} = \mathbf{x}^l$, since the codeword \mathbf{x} must satisfy *all* check nodes, but the string \mathbf{x}^l needs to satisfy only $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{C}}$ constraints. Those check nodes outside $\mathcal{N}_{(i,j)}^{2l}$ may limit the projected space $\mathbf{X}|_{\mathcal{N}_{(i,j)}^{2l}}$ to a strict subset of $\mathbf{X}_{(i,j)}^l$. For example, the second row of $\mathbf{A}\mathbf{x} = \mathbf{0}$ in Fig. 1 implies $x_6 = 0$. Therefore, two of the four elements of $\mathbf{X}_{(1,1)}^2$ in Fig. 2 are invalid/impossible projections of $\mathbf{x} \in \mathbf{X}$ on $\mathcal{N}_{(1,1)}^{2l}$. Thus, $\mathbf{X}|_{\mathcal{N}_{(1,1)}^{2l}}$ is a proper subset of $\mathbf{X}_{(1,1)}^2$.

To capture this phenomenon, we introduce the notion of a *perfectly projected* $\mathcal{N}_{(i,j)}^{2l}$.

Definition 1 (Perfectly Projected $\mathcal{N}_{(i,j)}^{2l}$): The supporting tree $\mathcal{N}_{(i,j)}^{2l}$ is perfectly projected, if for any $\mathbf{x}^l \in \mathbf{X}_{(i,j)}^l$

$$\frac{|\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}} = \mathbf{x}^l\}|}{|\mathbf{X}|} = \frac{1}{|\mathbf{X}_{(i,j)}^l|}. \quad (4)$$

That is, if we choose $\mathbf{x} \in \mathbf{X}$ equiprobably, $\mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}}$ will appear uniformly among all elements in $\mathbf{x}^l \in \mathbf{X}_{(i,j)}^l$. Thus, by looking only at the projections on $\mathcal{N}_{(i,j)}^{2l}$, it is as if we were choosing \mathbf{x}^l from $\mathbf{X}_{(i,j)}^l$ equiprobably and there were only $|\mathcal{N}_{(i,j)}^{2l}|_{\mathcal{C}}$ check node constraints and no others.

The example in Figs. 1 and 2 is obviously not perfectly projected.

Since the message emitted from node i to j in the l th iteration depends only on the received signals of the supporting tree, $\mathbf{y}|_{\mathcal{N}_{(i,j)}^{2l}}$, the codeword-dependent $P_{(i,j)}^{(l)}(\mathbf{x})$ actually depends only on the projection $\mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}}$, not on the entire codeword \mathbf{x} . That is,

$$P_{(i,j)}^{(l)}(\mathbf{x}) = P_{(i,j)}^{(l)}(\mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}}). \quad (5)$$

An immediate implication of $\mathcal{N}_{(i,j)}^{2l}$ being a perfect projection and (5) is

$$\begin{aligned}
P_{(i,j)}^{(l)}(x) &:= \left\langle P_{(i,j)}^{(l)}(\mathbf{x}) \right\rangle_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_i = x\}} \\
&= \frac{1}{|\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_i = x\}|} \sum_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_i = x\}} P_{(i,j)}^{(l)}(\mathbf{x}) \\
&= \frac{1}{|\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_i = x\}|} \\
&\quad \cdot |\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_{\mathcal{N}_{(i,j)}^{2l}} = \mathbf{x}^l, \mathbf{x}^l|_i = x\}| \\
&\quad \cdot \sum_{\{\mathbf{x}^l \in \mathbf{X}_{(i,j)}^l, \mathbf{x}^l|_i = x\}} P_{(i,j)}^{(l)}(\mathbf{x}^l) \\
&= \left\langle P_{(i,j)}^{(l)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l \in \mathbf{X}_{(i,j)}^l : \mathbf{x}^l|_i = x\}}. \tag{6}
\end{aligned}$$

Because of these two useful properties (5) and (6), throughout this subsection we assume that $\mathcal{N}_{(i,j)}^{2l}$ is perfectly projected. The convergence of $\mathcal{N}_{(i,j)}^{2l}$ to a perfect projection in probability is dealt with in Section IV. We will have all the preliminaries necessary for deriving the new density evolution after introducing the following self-explanatory lemma.

Lemma 1 (Linearity of Density Transformation): For any random variable A with distribution P_A , if $g : A \mapsto g(A)$ is measurable, then $B = g(A)$ is a random variable with distribution $P_B = T_g(P_A) := P_A \circ g^{-1}$. Furthermore, the density transformation T_g is linear, i.e., if $P_B = T_g(P_A)$ and $Q_B = T_g(Q_A)$, then $\alpha P_B + (1 - \alpha) Q_B = T_g(\alpha P_A + (1 - \alpha) Q_A)$, $\forall \alpha \in [0, 1]$.

B. New Formula

In the l th iteration, the probability of sending an incorrect message (averaged over all possible codewords) from variable node i_0 to check node j_0 is

$$\begin{aligned}
p_e^{(l)}(i_0, j_0) &= \frac{1}{|\mathbf{X}|} \left(\sum_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_{i_0} = 0\}} \int_{m=-\infty}^0 P_{(i_0, j_0)}^{(l)}(\mathbf{x})(dm) \right. \\
&\quad \left. + \sum_{\{\mathbf{x} \in \mathbf{X} : \mathbf{x}|_{i_0} = 1\}} \int_{m=0}^{\infty} P_{(i_0, j_0)}^{(l)}(\mathbf{x})(dm) \right) \\
&= \frac{1}{2} \left(\int_{m=-\infty}^0 P_{(i_0, j_0)}^{(l)}(0)(dm) \right. \\
&\quad \left. + \int_{m=0}^{\infty} P_{(i_0, j_0)}^{(l)}(1)(dm) \right). \tag{7}
\end{aligned}$$

Motivated by (7), we concentrate on finding an iterative formula for the density pair $P_{(i_0, j_0)}^{(l)}(0)$ and $P_{(i_0, j_0)}^{(l)}(1)$. Throughout this section, we also assume $\mathcal{N}_{(i_0, j_0)}^{2l}$ is tree-like (cycle free) and perfectly projected.

Let $1_{\{\cdot\}}$ denote the indicator function. By an auxiliary function $\gamma(m)$

$$\gamma(m) := \left(1_{\{m \leq 0\}}, \ln \coth \left| \frac{m}{2} \right| \right) \tag{8}$$

and letting the domain of the first coordinate of $\gamma(m)$ be GF(2), (2) for Ψ_c can be written as

$$\Psi_c(m_1, \dots, m_{d_c-1}) = \gamma^{-1} \left(\sum_{v=1}^{d_c-1} \gamma(m_v) \right). \tag{9}$$

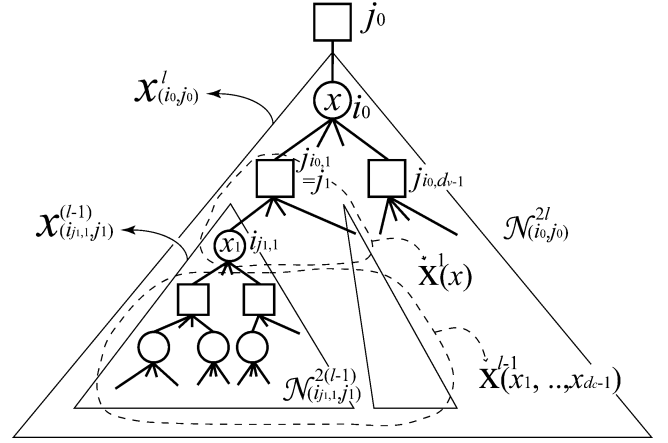


Fig. 3. Illustration of various quantities used in Section III-B.

By (1), (9), and the independence among the input messages, the classical density evolution for BP algorithms [23, eq.(9)] is as follows:

$$P_{(i_0, j_0)}^{(l)}(\mathbf{x}) = P_{(i_0, j_0)}^{(0)}(\mathbf{x}) \otimes \left(\bigotimes_{c=1}^{d_v-1} Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}) \right) \tag{10}$$

$$Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}) = \Gamma^{-1} \left(\bigotimes_{v=1}^{d_c-1} \Gamma \left(P_{(i_j, v, j_{i_0, c})}^{(l-1)}(\mathbf{x}) \right) \right) \tag{11}$$

where \otimes denotes the convolution operator on probability density functions, which can be implemented efficiently using the Fourier transform. $\Gamma := T_\gamma$ is the density transformation functional based on γ , defined in Lemma 1. Fig. 3 illustrates many helpful quantities used in (10), (11), and throughout this section.

By (5), (10), and the perfect projection assumption, we have

$$P_{(i_0, j_0)}^{(l)}(\mathbf{x}^l) = P_{(i_0, j_0)}^{(0)}(\mathbf{x}|_{i_0}) \otimes \left(\bigotimes_{c=1}^{d_v-1} Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}^l) \right). \tag{12}$$

Further simplification can be made such that

$$\begin{aligned}
P_{(i_0, j_0)}^{(l)}(x) &\stackrel{(a)}{=} \left\langle P_{(i_0, j_0)}^{(l)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l : \mathbf{x}^l|_{i_0} = x\}} \\
&\stackrel{(b)}{=} \left\langle P_{(i_0, j_0)}^{(0)}(x) \otimes \left(\bigotimes_{c=1}^{d_v-1} Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}^l) \right) \right\rangle_{\{\mathbf{x}^l : \mathbf{x}^l|_{i_0} = x\}} \\
&\stackrel{(c)}{=} P_{(i_0, j_0)}^{(0)}(x) \otimes \left\langle \bigotimes_{c=1}^{d_v-1} Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l : \mathbf{x}^l|_{i_0} = x\}} \\
&\stackrel{(d)}{=} P_{(i_0, j_0)}^{(0)}(x) \otimes \left(\bigotimes_{c=1}^{d_v-1} \left\langle Q_{(j_{i_0, c}, i_0)}^{(l-1)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l : \mathbf{x}^l|_{i_0} = x\}} \right) \\
&\stackrel{(e)}{=} P_{(i_0, j_0)}^{(0)}(x) \otimes \left(\left\langle Q_{(j_{i_0, 1}, i_0)}^{(l-1)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l : \mathbf{x}^l|_{i_0} = x\}} \right)^{\otimes (d_v-1)} \tag{13}
\end{aligned}$$

where (a) follows from (6), (b) follows from (12), and (c) follows from the linearity of convolutions. The fact that the subtrees generated by edges $(j_{i_0, c}, i_0)$ are completely disjoint implies that, by the perfect projection assumption on $\mathcal{N}_{(i_0, j_0)}^{2l}$, the distributions of strings on different subtrees are independent. As a result, the average of the convolutional products (over these

strings) equals the convolution of the averaged distributions, yielding (d). Finally, (e) follows from the fact that the distributions of messages from different subtrees are identical according to the perfect projection assumption.

To simplify $\langle Q_{(j_{i_0,1}, i_0)}^{(l-1)}(\mathbf{x}^l) \rangle_{\{\mathbf{x}^l: \mathbf{x}^l|_{i_0}=x\}}$, we need to define some new notation. We use j_1 to represent $j_{i_0,1}$ for simplicity. Denote by $\{\mathcal{N}_{(i_{j_1,v}, j_1)}^{2(l-1)}\}_{v \in [1, d_c-1]}$ the collection of all $d_c - 1$ subtrees rooted at $(i_{j_1,v}, j_1)$, $v \in [1, d_c - 1]$, and by $\mathbf{X}^{l-1}_{(i_{j_1,v}, j_1)}$ the strings compatible to $\mathcal{N}_{(i_{j_1,v}, j_1)}^{2(l-1)}$. We can then consider

$$\mathbf{X}^1(x) = \left\{ (x_1, \dots, x_{d_c-1}) : \left(\sum_{v=1}^{d_c-1} x_v \right) + x = 0 \right\}$$

containing the strings satisfying parity-check constraint j_1 given $x_{i_0} = x$, and

$$\begin{aligned} \mathbf{X}^{l-1}(x_1, \dots, x_{d_c-1}) \\ := \left\{ \left(\mathbf{x}_{(i_{j_1,1}, j_1)}^{l-1}, \dots, \mathbf{x}_{(i_{j_1, d_c-1}, j_1)}^{l-1} \right) : \right. \\ \left. \begin{aligned} \mathbf{x}_{(i_{j_1,1}, j_1)}^{l-1} \Big|_{i_{j_1,1}} &= x_1, \dots, \\ \mathbf{x}_{(i_{j_1, d_c-1}, j_1)}^{l-1} \Big|_{i_{j_1, d_c-1}} &= x_{d_c-1} \end{aligned} \right\} \end{aligned}$$

is the collection of the concatenations of substrings, in which the leading symbols of the substrings are (x_1, \dots, x_{d_c-1}) . All these quantities are illustrated in Fig. 3.

Note the following two properties: i) For any v , the message m_v from variable $i_{j_1,v}$ to check node j_1 depends only on $\mathbf{x}_{(i_{j_1,v}, j_1)}^{l-1}$; and ii) with the leading symbols $\{x_v\}_{v \in [1, d_c-1]}$ fixed and the perfect projection assumption, the projection on the strings $\{\mathbf{x}_{(i_{j_1,v}, j_1)}^{l-1}\}_{v \in [1, d_c-1]}$ are independent, and thus, the averaged convolution of densities is equal to the convolution of the averaged densities. By repeatedly applying Lemma 1 and the above two properties, we have

$$\begin{aligned} & \left\langle Q_{(j_{i_0,1}, i_0)}^{(l-1)}(\mathbf{x}^l) \right\rangle_{\{\mathbf{x}^l: \mathbf{x}^l|_{i_0}=x\}} \\ &= \left\langle \Gamma^{-1} \left(\bigotimes_{v=1}^{d_c-1} \Gamma \left(P_{(i_{j_1,v}, j_{i_0,c})}^{(l-1)}(\mathbf{x}^l) \right) \right) \right\rangle_{\{\mathbf{x}^l: \mathbf{x}^l|_{i_0}=x\}} \\ &= \left\langle \Gamma^{-1} \left(\bigotimes_{v=1}^{d_c-1} \Gamma \left(P_{(i_{j_1,v}, j_{i_0,c})}^{(l-1)}(\mathbf{x}_{(i_{j_1,v}, j_1)}^{l-1}) \right) \right) \right\rangle_{\{\mathbf{x}^l: \mathbf{x}^l|_{i_0}=x\}} \\ &= \Gamma^{-1} \left(\frac{1}{2^{d_c-2}} \sum_{\mathbf{x}^1 \in \mathbf{X}^1(x)} \left\langle \bigotimes_{v=1}^{d_c-1} \Gamma \left(P_{(i_{j_1,v}, j_{i_0,c})}^{(l-1)}(\mathbf{x}_{(i_{j_1,v}, j_1)}^{l-1}) \right) \right\rangle_{\mathbf{X}^{l-1}(\mathbf{x}^1)} \right) \\ &= \Gamma^{-1} \left(\frac{1}{2^{d_c-2}} \sum_{\mathbf{x}^1 \in \mathbf{X}^1(x)} \bigotimes_{v=1}^{d_c-1} \left(\left\langle P_{(i_{j_1,v}, j_{i_0,c})}^{(l-1)}(\mathbf{x}_{(i_{j_1,v}, j_1)}^{l-1}) \right\rangle_{\mathbf{X}^{l-1}(\mathbf{x}^1)} \right) \right) \\ &= \Gamma^{-1} \left(\frac{1}{2^{d_c-2}} \sum_{\mathbf{x}^1 \in \mathbf{X}^1(x)} \bigotimes_{v=1}^{d_c-1} \Gamma \left(P_{(i_{j_1,v}, j_{i_0,c})}^{(l-1)}(x_v) \right) \right). \quad (14) \end{aligned}$$

By (13), (14), and dropping the subscripts during the density evolution, a new density evolution formula for $P^{(l)}(x)$, $\forall x = 0, 1$, is as follows:

$$\begin{aligned} P^{(l)}(x) &= P^{(0)}(x) \otimes \left(Q^{(l-1)}(x) \right)^{\otimes (d_c-1)} \\ Q^{(l-1)}(x) &= \Gamma^{-1} \left(\frac{1}{2^{d_c-2}} \sum_{\mathbf{x}^1 \in \mathbf{X}^1(x)} \bigotimes_{v=1}^{d_c-1} \Gamma \left(P^{(l-1)}(x_v) \right) \right). \end{aligned}$$

With the help of the linearity of distribution transformations and convolutions, the preceding equations can be further simplified and the desired efficient iterative formulas become

$$\begin{aligned} P^{(l)}(x) &= P^{(0)}(x) \otimes \left(Q^{(l-1)}(x) \right)^{\otimes (d_c-1)} \\ Q^{(l-1)}(x) &= \Gamma^{-1} \left(\left(\Gamma \left(\frac{P^{(l-1)}(0) + P^{(l-1)}(1)}{2} \right) \right)^{\otimes (d_c-1)} \right. \\ &\quad \left. + (-1)^x \left(\Gamma \left(\frac{P^{(l-1)}(0) - P^{(l-1)}(1)}{2} \right) \right)^{\otimes (d_c-1)} \right). \end{aligned}$$

The preceding formula can be easily generalized to the irregular code ensembles $\mathcal{C}^n(\lambda, \rho)$

$$\begin{aligned} P^{(l)}(x) &= P^{(0)}(x) \otimes \lambda \left(Q^{(l-1)}(x) \right) \\ Q^{(l-1)}(x) &= \Gamma^{-1} \left(\rho \left(\Gamma \left(\frac{P^{(l-1)}(0) + P^{(l-1)}(1)}{2} \right) \right) \right. \\ &\quad \left. + (-1)^x \rho \left(\Gamma \left(\frac{P^{(l-1)}(0) - P^{(l-1)}(1)}{2} \right) \right) \right) \quad (15) \end{aligned}$$

which has the same complexity as the classical density evolution for symmetric channels.

Remark: The above derivation relies heavily on the perfect projection assumption, which guarantees that uniformly averaging over all codewords is equivalent to uniformly averaging over the tree-satisfying strings. Since the tree-satisfying strings are well-structured and symmetric, we are on solid ground to move the average inside the classical density evolution formula.

IV. DENSITY EVOLUTION: FUNDAMENTAL THEOREMS

As stated in Section III, the tree-like until depth $2l$ and the perfect projection assumptions are critical in our analysis. The use of code ensembles rather than fixed codes facilitates the analysis but its relationship to fixed codes needs to be explored. We restate two necessary theorems from [13], and give a novel perfect projection convergence theorem, which is essential to our new density evolution method. With these theorems, a concrete theoretical foundation will be established.

Theorem 1 (Convergence to the Cycle-Free Case, [13]): Fix l, i_0 , and j_0 . For any (d_v, d_c) , there exists a constant $\alpha > 0$ such that for all $n \in \mathbb{N}$, the code ensemble $\mathcal{C}^n(d_v, d_c)$ satisfies

$$\begin{aligned} \mathbb{P} \left(\mathcal{N}_{(i_0, j_0)}^{2l} \text{ is cycle-free} \right) &\geq 1 - \alpha \left(\frac{\{(d_v - 1)(d_c - 1)\}^{2l}}{n} \right) \end{aligned}$$

where $\mathcal{N}_{(i_0, j_0)}^{2l}$ is the support tree as defined by (3).

Theorem 2 (Convergence to Perfect Projection in Probability): Fix l, i_0 , and j_0 . For any regular, bipartite, equiprobable graph ensemble $\mathcal{C}^n(d_v, d_c)$, we have

$$\mathbb{P}\left(\mathcal{N}_{(i_0, j_0)}^{2l} \text{ is perfectly projected}\right) = 1 - \mathcal{O}(n^{-0.1}).$$

Remark: The above two theorems focus only on the properties of equiprobable regular bipartite graph ensembles, and are independent of the channel type of interest.

Theorem 3 (Concentration to the Expectation, [13]): With fixed transmitted codeword \mathbf{x} , let Z denote the number of wrong messages (those m 's such that $m(-1)^x < 0$). There exists a constant $\beta > 0$ such that for any $\epsilon > 0$, over the code ensemble $\mathcal{C}^n(d_v, d_c)$ and the channel realizations \mathbf{y} , we have

$$\mathbb{P}\left(\left|\frac{Z - \mathbb{E}\{Z\}}{nd_v}\right| > \frac{\epsilon}{2}\right) \leq 2e^{-\beta\epsilon^2 n}. \quad (16)$$

Furthermore, β is independent of $f_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})$, and thus is independent of \mathbf{x} .

Theorem 3 can easily be generalized to symbol-dependent channels in the following corollary.

Corollary 1: Over the equiprobable codebook \mathbf{X} , the code ensemble⁵ $\mathcal{C}^n(d_v, d_c)$, and channel realizations \mathbf{y} , (16) still holds.

Proof: Since the constant β in Theorem 3 is independent of the transmitted codeword \mathbf{x} , after averaging over the equiprobable codebook \mathbf{X} , the inequality still holds. That is,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{Z - \mathbb{E}\{Z\}}{nd_v}\right| > \frac{\epsilon}{2}\right) &= \mathbb{E}_{\mathbf{x}}\left\{\mathbb{P}\left(\left|\frac{Z - \mathbb{E}\{Z\}}{nd_v}\right| > \frac{\epsilon}{2} \middle| \mathbf{x}\right)\right\} \\ &\leq \mathbb{E}_{\mathbf{x}}\left\{2e^{-\beta\epsilon^2 n}\right\} = 2e^{-\beta\epsilon^2 n}. \quad \square \end{aligned}$$

Now we have all the prerequisite of proving the theoretical foundation of our codeword-averaged density evolution.

Theorem 4 (Validity of Codeword-Averaged DE): Consider any regular, bipartite, equiprobable graph ensemble $\mathcal{C}^n(d_v, d_c)$ with fixed l, i_0 , and j_0 . $p_e^{(l)}(i_0, j_0)$ is derived from (7) and the codeword-averaged density evolution after l iterations. The probability over equiprobable codebook \mathbf{X} , the code ensemble $\mathcal{C}^n(d_v, d_c)$, and the channel realizations \mathbf{y} , satisfies

$$\mathbb{P}\left(\left|\frac{Z}{nd_v} - p_e^{(l)}(i_0, j_0)\right| > \epsilon\right) = e^{-\epsilon^2 \mathcal{O}(n)}, \quad \forall \epsilon > 0.$$

⁵The only valid codeword for all code instances of the ensemble is the all-zero codeword. Therefore, a fixed bit string is in general not a valid codeword for most instances of the code ensemble, which hampers the averaging over the code ensemble. This, however, can be circumvented by the following construction. We first use Gaussian elimination to index the codewords $1, \dots, 2^{nR}$ for any code instance in the code ensemble. And we then fix the index instead of the codeword. The statements and the proof of Theorem 3 hold verbatim after this slight modification.

Proof: We note that $\frac{Z}{nd_v}$ is bounded between 0 and 1. By observing that

$$\begin{aligned} &\left(\frac{Z}{nd_v}\right) \mathbb{1}\left\{\mathcal{N}_{(i_0, j_0)}^{2l} \text{ is cycle-free and perfectly projected}\right\} \\ &\leq \left(\frac{Z}{nd_v}\right) \\ &\leq \left(\frac{Z}{nd_v} - 1\right) \mathbb{1}\left\{\mathcal{N}_{(i_0, j_0)}^{2l} \text{ is cycle-free and perfectly projected}\right\} \\ &\quad + 1 \end{aligned}$$

and using Theorems 1 and 2, we have

$$\lim_{n \rightarrow \infty} \mathbb{E}\left\{\frac{Z}{nd_v}\right\} = p_e^{(l)}(i_0, j_0).$$

Then by Corollary 1, the proof is complete. \square

The Proof of Theorem 2 will be included in Appendix I.

V. MONOTONICITY, SYMMETRY, AND STABILITY

In this section, we prove the monotonicity, symmetry, and stability of our codeword-averaged density evolution method on BP algorithms. Since the codeword-averaged density evolution reduces to the traditional one when the channel of interest is symmetric, the following theorems also reduce to those (in [23] and [13]) for symmetric channels.

A. Monotonicity

Proposition 1 (Monotonicity With Respect to l): Let $p_e^{(l)}$ denote the bit-error probability of the codeword-averaged density evolution defined in (7). Then $p_e^{(l+1)} \leq p_e^{(l)}$, for all $l \in \mathbb{N}$.

Proof: We first note that the codeword-averaged approach can be viewed as concatenating a bit-to-sequence random mapper with the observation channels, and the larger the tree structure is, the more observation/information the decision maker has. Since the BP decoder is the optimal MAP decoder for the tree structure of interest, the larger the tree is, the smaller the error probability will be. The proof is thus complete. \square

Proposition 2 (Monotonicity With Respect to Degraded Channels): Let $f(y|x)$ and $g(y|x)$ denote two different channel models, such that $g(y|x)$ is degraded with respect to (w.r.t.) $f(y|x)$. The corresponding decoding error probabilities, $p_{e,f}^{(l)}$ and $p_{e,g}^{(l)}$, are defined in (7). Then for any fixed l , we have $p_{e,f}^{(l)} \leq p_{e,g}^{(l)}$.

Proof: By taking the same point of view that the codeword-averaged approach is a concatenation of a bit-to-sequence random mapper with the observation channels, this theorem can be easily proved by the channel degradation argument. \square

B. Symmetry

We will now show that even though the evolved density is derived from asymmetric channels, there are still some symmetry

properties inherent in the symmetric structure of BP algorithms. We define the symmetric distribution pair as follows.

Definition 2 (Symmetric Distribution Pairs): Two probability measures P and Q are a symmetric pair if for any integrable function h , we have

$$\int h(m)dP(m) = \int e^{-m}h(-m)dQ(m).$$

A distribution P_s is *self-symmetric* if (P_s, P_s) is a symmetric pair.

Proposition 3: Let $I(m) := -m$ be a parity reversing function, and let $P^{(l)}(0)$ and $P^{(l)}(1)$ denote the resulting density functions from the codeword-averaged density evolution. Then $P^{(l)}(0)$ and $P^{(l)}(1) \circ I^{-1}$ are a symmetric pair for all $l \in \mathbb{N}$.

Remark: In the symmetric channel case, $P^{(l)}(0)$ and $P^{(l)}(1)$ differ only in parity (Lemma 1, [13]). Thus, $P^{(l)}(0) = P^{(l)}(1) \circ I^{-1}$ is self-symmetric ([23, Theorem 3]).

Proof: We note that by the equiprobable codeword distribution and the perfect projection assumption $P^{(l)}(0)$ and $P^{(l)}(1)$ act on the random variable m , given by

$$m := \ln \frac{P(x=0|\mathbf{y}^l)}{P(x=1|\mathbf{y}^l)} = \ln \frac{P(\mathbf{y}^l|x=0)}{P(\mathbf{y}^l|x=1)}$$

where \mathbf{y}^l is the received signal on the subset \mathcal{N}^{2l} and P is the distribution over channel realizations and equiprobable codewords. Then by a change of measure

$$\begin{aligned} & \int h(m)P^{(l)}(0)(dm) \\ &= \mathbb{E}_{x=0} \left\{ h \left(\ln \frac{P(\mathbf{y}^l|x=0)}{P(\mathbf{y}^l|x=1)} \right) \right\} \\ &= \mathbb{E}_{x=1} \left\{ \frac{P(\mathbf{y}^l|x=0)}{P(\mathbf{y}^l|x=1)} h \left(\ln \frac{P(\mathbf{y}^l|x=0)}{P(\mathbf{y}^l|x=1)} \right) \right\} \\ &= \int e^m h(m)P^{(l)}(1)(dm). \end{aligned} \quad (17)$$

This completes the proof. \square

Corollary 2:

$$\langle P^{(l)} \rangle := \frac{P^{(l)}(0) + P^{(l)}(1) \circ I^{-1}}{2}$$

is self-symmetric for all l , i.e., $(\langle P^{(l)} \rangle, \langle P^{(l)} \rangle)$ is a symmetric pair.

C. Stability

Rather than looking only at the error probability $p_e^{(l)}$ of the evolved densities $P^{(l)}(0)$ and $P^{(l)}(1)$, we also focus on its Chernoff bound

$$\text{CBP}^{(l)}(x) := \int e^{-\frac{(-1)^x m}{2}} P^{(l)}(x)(dm).$$

By letting $h(m) = e^{-\frac{m}{2}}$ and by (17), we have $\text{CBP}^{(l)}(0) = \text{CBP}^{(l)}(1)$. The averaged $\langle \text{CBP}^{(l)} \rangle$ then becomes

$$\begin{aligned} \langle \text{CBP}^{(l)} \rangle &:= \frac{1}{2} \left(\text{CBP}^{(l)}(0) + \text{CBP}^{(l)}(1) \right) \\ &= \text{CBP}^{(l)}(0) = \text{CBP}^{(l)}(1) \\ &= \int e^{-\frac{m}{2}} \langle P^{(l)} \rangle (dm). \end{aligned} \quad (18)$$

We state three properties which can easily be derived from the self-symmetry of $\langle P^{(l)} \rangle$. Proofs can be found in [31], [23], and [30].

- $\langle \text{CBP}^{(l)} \rangle = \min_s \int e^{-s \cdot m} \langle P^{(l)} \rangle (dm)$.
- The density of $e^{-m/2} \langle P^{(l)} \rangle (dm)$ is symmetric with respect to $m = 0$.
- $2p_e^{(l)} \leq \langle \text{CBP}^{(l)} \rangle \leq 2\sqrt{p_e^{(l)}(1-p_e^{(l)})}$. This justifies the use of $\langle \text{CBP}^{(l)} \rangle$ as our performance measure.

Thus, we consider $\langle \text{CBP}^{(l)} \rangle$, the Chernoff bound of $p_e^{(l)}$. With the regularity assumption that $\int_{\mathbf{R}} e^{sm} \langle P^{(0)} \rangle (dm) < \infty$ for all s in some neighborhood of zero, we state the necessary and sufficient stability conditions as follows.

Theorem 5 (Sufficient Stability Condition): Let

$$r := \langle \text{CBP}^{(0)} \rangle = \int_{\mathbf{R}} e^{-m/2} \langle P^{(0)} \rangle (dm).$$

Suppose $\lambda_2 \rho'(1)r < 1$, and let ϵ^* be the smallest strictly positive root of the following equation

$$\lambda(\rho'(1)\epsilon)r = \epsilon.$$

If for some l_0 , $\langle \text{CBP}^{(l_0)} \rangle < \epsilon^*$, then

$$\langle \text{CBP}^{(l)} \rangle = \begin{cases} \mathcal{O} \left((\lambda_2 \rho'(1)r)^l \right), & \text{if } \lambda_2 > 0 \\ \mathcal{O} \left(e^{-\mathcal{O}((k_\lambda - 1)^l)} \right), & \text{if } \lambda_2 = 0 \end{cases}$$

where $k_\lambda = \min\{k : \lambda_k > 0\}$. In both cases: $\lambda_2 = 0$ and $\lambda_2 > 0$, $\lim_{l \rightarrow \infty} \langle \text{CBP}^{(l)} \rangle = 0$.

Corollary 3: For any noise distribution $f(y|x)$ with Bhattacharyya noise parameter $r := \langle \text{CBP}^{(0)} \rangle$, if there is no $\epsilon \in (0, r)$ such that

$$\lambda(\rho'(1)\epsilon)r = \epsilon$$

then $\mathcal{C}(\lambda, \rho)$ will have arbitrarily small bit-error rate (BER) as n tends to infinity. The corresponding r can serve as an inner bound of the achievable region for general asymmetric memoryless channels. Further discussion of finite-dimensional bounds on the achievable region can be found in [30].

Theorem 6 (Necessary Stability Condition): Let $r := \langle \text{CBP}^{(0)} \rangle$. If $\lambda_2 \rho'(1)r > 1$, then $\lim_{l \rightarrow \infty} p_e^{(l)} > 0$.

- *Remark 1:* $\langle \text{CBP}^{(0)} \rangle$ is the Bhattacharyya noise parameter and is related to the cutoff rate R_0 by $R_0 = 1 - \log_2(1 + \langle \text{CBP}^{(0)} \rangle)$. Further discussion of $\langle \text{CBP}^{(0)} \rangle$ for turbo-like and LDPC codes can be found in [25], [31], [30].

- *Remark 2:* The stability results are first stated in [23] without the convergence rate statement and the stability region ϵ^* . Since we focus on general asymmetric channels (with symmetric channels as a special case), our convergence rate and stability region ϵ^* results also apply to the symmetric channel case. Benefitting from considering its Chernoff version, we will provide a simple proof, which did not appear in [23].
- *Remark 3:* ϵ^* can be used as a stopping criterion for the iterations of the density evolution. Moreover, ϵ^* is lower-bounded by $\frac{1-\lambda_2\rho'(1)r}{\lambda(\rho'(1)r-\lambda_2\rho'(1)r)}$, which is a computationally efficient substitute for ϵ^* .

Proof of Theorem 5: We define the Chernoff bound of the density of the messages emitting from check nodes $\text{CBQ}^{(l)}(x)$ in a fashion similar to $\text{CBP}^{(l)}(x)$

$$\text{CBQ}^{(l)}(x) := \int e^{-\frac{(-1)^x m}{2}} Q^{(l)}(x)(dm).$$

First consider the case in which $d_c = 3$. We then have

$$\begin{aligned} \Psi_c(m_1, m_2) &= \ln \left(\frac{1 + \tanh \frac{m_1}{2} \tanh \frac{m_2}{2}}{1 - \tanh \frac{m_1}{2} \tanh \frac{m_2}{2}} \right) \\ &= \ln \frac{e^{m_1} e^{m_2} + 1}{e^{m_1} + e^{m_2}}. \end{aligned}$$

To simplify the analysis, we assume the all-zero codeword is transmitted and then generalize the results to nonzero codewords. Suppose the distributions of m_1 and m_2 are $P_1^{(l)}(0)$ and $P_2^{(l)}(0)$, respectively. The $\text{CBQ}^{(l)}(0)$ becomes

$$\begin{aligned} \text{CBQ}^{(l)}(0) &= \int e^{-\frac{\Psi_c(m_1, m_2)}{2}} P_1^{(l)}(0)(dm_1) \times P_2^{(l)}(0)(dm_2) \\ &= \int \sqrt{\frac{e^{m_1} + e^{m_2}}{e^{m_1} e^{m_2} + 1}} P_1^{(l)}(0)(dm_1) \times P_2^{(l)}(0)(dm_2) \\ &\leq \int \sqrt{e^{m_1} + e^{m_2}} P_1^{(l)}(0)(dm_1) \times P_2^{(l)}(0)(dm_2) \\ &\leq \int \sqrt{e^{m_1}} + \sqrt{e^{m_2}} P_1^{(l)}(0)(dm_1) \times P_2^{(l)}(0)(dm_2) \\ &= \text{CBP}_1^{(l)}(0) + \text{CBP}_2^{(l)}(0) \end{aligned} \quad (19)$$

where the last inequality follows from the fact that $\forall \alpha, \beta \geq 0, \sqrt{\alpha + \beta} \leq \sqrt{\alpha} + \sqrt{\beta}$. Since any check node with $d_c > 3$ can be viewed as the concatenation of many check nodes with $d_c = 3$, by induction and by assuming the all-zero codeword is transmitted, we have

$$\text{CBQ}^{(l)}(0) \leq (d_c - 1)\text{CBP}^{(l)}(0). \quad (20)$$

Since $\text{CBP}^{(l)}(0) = \text{CBP}^{(l)}(1)$ as in (18), the averaging over all possible codewords does not change (20). By further incorporating the check node degree polynomial ρ , we have

$$\begin{aligned} \forall x \in \{0, 1\}, \quad \text{CBQ}^{(l)}(x) &\leq \sum_k \rho_k(k-1) \langle \text{CBP}^{(l)} \rangle \\ &= \rho'(1) \langle \text{CBP}^{(l)} \rangle. \end{aligned}$$

By (15) and the fact that the moment generating function of the convolution equals the product of individual moment generating functions, we have

$$\begin{aligned} \text{CBP}^{(l+1)}(x) &= \text{CBP}^{(0)}(x) \sum_k \lambda_k \left(\text{CBQ}^{(l)}(x) \right)^{k-1} \\ &\leq \text{CBP}^{(0)}(x) \lambda \left(\rho'(1) \langle \text{CBP}^{(l)} \rangle \right) \end{aligned}$$

which is equivalent to

$$\langle \text{CBP}^{(l+1)} \rangle \leq \langle \text{CBP}^{(0)} \rangle \lambda \left(\rho'(1) \langle \text{CBP}^{(l)} \rangle \right). \quad (21)$$

The sufficient stability theorem follows immediately from (21), the iterative upper bound formula. \square

Remark: Inequality (21) is a one-dimensional iterative bound for general asymmetric memoryless channels. In [30], this iterative upper bound will be further strengthened to

$$\langle \text{CBP}^{(l+1)} \rangle \leq \langle \text{CBP}^{(0)} \rangle \lambda \left(1 - \rho \left(1 - \langle \text{CBP}^{(l)} \rangle \right) \right)$$

which is tight for BECs and holds for symmetric channels as well.

Proof of Theorem 6: We prove this result by the erasure decomposition technique used in [23].

The erasure decomposition lemma in [23] states that, for any $l_0 > 0$, and any symmetric channel f with LLR distribution $P^{(l_0)}$, there exists a BEC g with LLR distribution $B^{(l_0)}$ such that f is physically degraded with respect to g . Furthermore, $B^{(l_0)}$ is of the following form:

$$B^{(l_0)} = 2\epsilon\delta_0 + (1 - 2\epsilon)\delta_\infty$$

for all $\epsilon \leq p_e^{(l_0)}$, where δ_x is the Dirac-delta measure centered at x . It can be easily shown that this erasure decomposition lemma holds even when f corresponds to an asymmetric channel with LLR distributions $\{P^{(l_0)}(x)\}_{x=0,1}$, and $p_e^{(l_0)}$ computed from (7).

We can then assign

$$B^{(l_0)}(0) := B^{(l_0)} \quad \text{and} \quad B^{(l_0)}(1) := B^{(l_0)} \circ I^{-1}$$

to distinguish the distributions for different transmitted symbols x .

Suppose $r\lambda_2\rho'(1) > 1$ and $\lim_{l \rightarrow \infty} p_e^{(l)} = 0$. Then, for any $\epsilon > 0, \exists l_0 > 0$, such that $p_e^{(l_0)} \leq \epsilon$. For simplicity, we assume $p_e^{(l_0)} = \epsilon$. The physically better BEC is described as above. If during the iteration procedure (15) we replace the density $P^{(l_0)}(x)$ with $B^{(l_0)}(x)$, then the resulting density will be

$$\begin{aligned} P_B^{(l_0+\Delta l)}(0) &= 2\epsilon(\lambda_2\rho'(1))^{\Delta l} P^{(0)}(0) \otimes \left(\langle P^{(0)} \rangle \right)^{\otimes(\Delta l-1)} \\ &\quad + (1 - 2\epsilon(\lambda_2\rho'(1))^{\Delta l}) \delta_\infty + \mathcal{O}(\epsilon^2) \\ P_B^{(l_0+\Delta l)}(1) &= 2\epsilon(\lambda_2\rho'(1))^{\Delta l} P^{(0)}(1) \otimes \left(\langle P^{(0)} \rangle \circ I^{-1} \right)^{\otimes(\Delta l-1)} \\ &\quad + (1 - 2\epsilon(\lambda_2\rho'(1))^{\Delta l}) \delta_{-\infty} + \mathcal{O}(\epsilon^2) \end{aligned}$$

and the averaged error probability $p_{e,B}^{(l_0+\Delta l)}$ is

$$\begin{aligned} p_{e,B}^{(l_0+\Delta l)} &:= \int_{-\infty}^0 \frac{P_B^{(l_0+\Delta l)}(0) + P_B^{(l_0+\Delta l)}(1) \circ I^{-1}}{2} (dm) \\ &= \mathcal{O}(\epsilon^2) + 2\epsilon(\lambda_2\rho'(1))^{\Delta l} \int_{-\infty}^0 d \left(\langle P^{(0)} \rangle \right)^{\otimes \Delta l}. \end{aligned} \quad (22)$$

By the fact that $r = \langle \text{CBP}^{(0)} \rangle$ is the Chernoff bound on $\int_{-\infty}^0 d\langle P^{(0)} \rangle$, the regularity condition and the Chernoff theorem, for any $\epsilon' > 0$, there exists a large enough Δl such that

$$\int_{-\infty}^0 d \left(\langle P^{(0)} \rangle \right)^{\otimes \Delta l} \geq (r - \epsilon')^{\Delta l}.$$

With a small enough ϵ' , we have $\lambda_2 \rho'(1)(r - \epsilon') > 1$. Thus, with large enough Δl , we have

$$p_{e,B}^{(l_0+\Delta l)} > \mathcal{O}(\epsilon^2) + 2\epsilon.$$

With small enough ϵ or equivalently large enough l_0 , we have

$$p_{e,B}^{(l_0+\Delta l)} > \mathcal{O}(\epsilon^2) + 2\epsilon > \epsilon = p_e^{(l_0)}.$$

However, by the monotonicity with respect to physically degraded channels we have, $p_e^{(l_0+\Delta l)} \geq p_{e,B}^{(l_0+\Delta l)} > p_e^{(l_0)}$, which contradicts the monotonicity of $p_e^{(l)}$ with respect to l . From the above reasoning, if $r\lambda_2\rho'(1) > 1$, then $\lim_{l \rightarrow \infty} p_e^{(l)} > 0$, which completes the proof. \square

Remark: From the sufficient stability condition, for those codes with $\lambda_2 > 0$, the convergence rate is exponential in l , i.e., $\text{BER} = O((r\lambda_2\rho'(1))^l)$. However, the number of bits involved in the \mathcal{N}^{2l} tree is $O(((d_v - 1)(d_c - 1))^l)$, which is usually much faster than the reciprocal of the decrease rate of $\text{BER} = O((r\lambda_2\rho'(1))^l)$. As a result, we conjecture that the average performance of the code ensemble with $\lambda_2 > 0$ will have bad *block-error* probabilities. This is confirmed later in Fig. 5(b) and theoretically proved for the BEC in [32]. The converse is stated and proved in the following corollary.

Corollary 4: Let $E\{Z_B^{(l)}\}$ denote the block-error probability of codeword length n after l iterations of the BP algorithm, which is averaged over equiprobable codewords, channel realizations, and the code ensemble $\mathcal{C}^n(\lambda, \rho)$. If $\lambda_2 = 0$ and l_n satisfying $\ln \ln n = o(l_n)$ and $l_n = o(\ln n)$,

$$\lim_{n \rightarrow \infty} E\{Z_B^{(l_n)}\} = 0.$$

Proof: This result can be proven directly by the cycle-free convergence theorem, the super-exponential *bit* convergence rate with respect to l , and the union bound. \square

A similar observation is also made and proved in [25], in which it is shown that the interleaving gain exponent of the block-error rate is $-J + 2$, where J is the number of parallel constituent codes. The variable node degree d_v is the number of parity-check equations (parity-check subcodes) in which a variable bit participates. In a sense, an LDPC code is similar to d_v parity-check codes interleaved together. With $d_v = 2$, good interleaving gain for the block-error probability is not expected.

VI. SIMULATIONS AND DISCUSSION

It is worth noting that for asymmetric channels, different codewords will have different error-resisting capabilities. In this section, we consider the averaged performance. We can obtain codeword-independent performance by adding a random number to the information message before encoding and then subtracting it after decoding. This approach, however, introduces higher computational cost.

A. Simulation Settings

With the help of the sufficient condition of the stability theorem (Theorem 5), we can use ϵ^* to set a stopping criterion for the iterations of the density evolution. We use the 8-bit quantized density evolution method with $(-15, 15)$ being the domain of the LLR messages. We will determine the largest thresholds such that the evolved Chernoff bound $\langle \text{CBP}^{(l)} \rangle$ hits ϵ^* within 100 iterations, i.e., $\langle \text{CBP}^{(100)} \rangle < \epsilon^*$. Better performance can be achieved by using more iterations, which, however, is of less practical interest. For example, the 500-iteration threshold of our best code for z-channels, 12B (described below), is 0.2785, compared to the 100-iteration threshold 0.2731. Five different code ensembles with rate 1/2 are extensively simulated, including regular (3, 6) codes, regular (4, 8) codes, 12A codes, 12B codes, and 12C codes as follows:

- 12A: 12A is a rate-1/2 code ensemble found by Richardson, *et al.* in [23], which is the best known degree distribution optimized for the symmetric BiAWGNC, having maximum degree constraints $\max d_v \leq 12$ and $\max d_c \leq 9$. Its degree distributions are

$$\begin{aligned} \lambda(x) &= 0.24426x + 0.25907x^2 + 0.01054x^3 \\ &\quad + 0.05510x^4 + 0.01455x^7 + 0.01275x^9 \\ &\quad + 0.40373x^{11} \\ \rho(x) &= 0.25475x^6 + 0.73438x^7 + 0.01087x^8. \end{aligned}$$

- 12B: 12B is a rate-1/2 code ensemble obtained by minimizing the hitting time of ϵ^* in z-channels, through hill-climbing and linear programming techniques. The maximum degree constraints are also $\max d_v \leq 12$ and $\max d_c \leq 9$. The differences between 12A and 12B are 1) 12B is optimized for the z-channels with our code-word-averaged density evolution, and 12A is optimized for the symmetric BiAWGNC. 2) 12B is optimized with respect to the hitting time of ϵ^* (depending on (λ, ρ)) rather than a fixed small threshold. The degree distributions of 12B are

$$\begin{aligned} \lambda(x) &= 0.236809x + 0.309590x^2 + 0.032789x^3 \\ &\quad + 0.007116x^4 + 0.000001x^5 + 0.413695x^{11} \\ \rho(x) &= 0.000015x^5 + 0.464854x^6 + 0.502485x^7 \\ &\quad + 0.032647x^8. \end{aligned}$$

- 12C: 12C a rate-1/2 code ensemble similar to 12B, but with λ_2 being hardwired to 0, which is suggested by the convergence rate in the sufficient stability condition. The degree distributions of 12C are

$$\begin{aligned} \lambda(x) &= 0.861939x^2 + 0.000818x^3 + 0.000818x^4 \\ &\quad + 0.000818x^5 + 0.000818x^6 + 0.000818x^7 \\ &\quad + 0.000218x^8 + 0.077898x^9 + 0.055843x^{10} \\ &\quad + 0.000013x^{11} \\ \rho(x) &= 0.000814x^4 + 0.560594x^5 + 0.192771x^6 \\ &\quad + 0.145207x^7 + 0.100613x^8. \end{aligned}$$

TABLE I
THRESHOLDS OF DIFFERENT CODES AND CHANNELS, WITH PRECISION 10^{-4}

Codes	BEC		BSC		Z-channels		BiAWGNC		Stability
	ϵ	$\langle CBP \rangle$	ϵ	$\langle CBP \rangle$	ϵ_1	$\langle CBP \rangle$	σ	$\langle CBP \rangle$	$\langle CBP \rangle$
(3,6)	0.4294	0.4294	0.0837	0.5539	0.2305	0.4828	0.8790	0.5235	—
(4,8)	0.3834	0.3834	0.0764	0.5313	0.1997	0.4497	0.8360	0.4890	—
12A	0.4682	0.4682	0.0937	0.5828	0.2710	0.5233	0.9384	0.5668	0.6060
12B	0.4753	0.4753	0.0939	0.5834	0.2731	0.5253	0.9362	0.5653	0.6247
12C	0.4354	0.4354	0.0862	0.5613	0.2356	0.4881	0.8878	0.5303	—
Sym. Info. Rate	0.5000	0.5000	0.1100	0.6258	0.2932	0.5415	0.9787	0.5933	—
Capacity	0.5000	0.5000	0.1100	0.6258	0.3035	0.5509	0.9787	0.5933	—

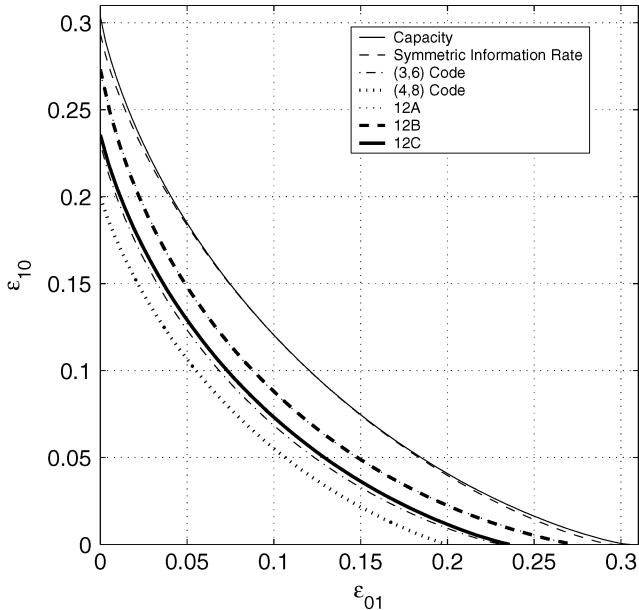


Fig. 4. Asymptotic thresholds and the achievable regions of different codes in binary asymmetric channels.

Four different channels are considered, including the BEC, BSC, z-channel, and BiAWGNC. Z-channels are simulated by binary asymmetric channels with very small ϵ_0 ($\epsilon_0 = 0.00001$) and different values of ϵ_1 . Table I summarizes the thresholds with precision 10^{-4} . Thresholds are not only presented by their conventional channel parameters, but also by their Bhattacharyya noise parameters (Chernoff bounds). The column “stability” lists the maximum $r := \langle CBP^{(0)} \rangle$ such that $r\lambda_2\rho'(1) < 1$, which is an upper bound on the $\langle CBP^{(0)} \rangle$ values of decodable channels. Further discussion of the relationship between $\langle CBP^{(0)} \rangle$ and the decodable threshold can be found in [30].

From Table I, we observe that 12A outperforms 12B in Gaussian channels (for which 12A is optimized), but 12B is superior in z-channels for which it is optimized. The above behavior promises room for improvement with codes optimized for different channels, as was also shown in [14].

Fig. 4 demonstrates the asymptotic thresholds of these codes in BASCs with the curves of 12A and 12B being very close together. It is seen that 12B is slightly better when $\epsilon_0, \epsilon_1 \rightarrow 0$ or $\epsilon_0 \approx \epsilon_1$. We notice that all the achievable regions of these codes are bounded by the symmetric mutual information rate (with a $(1/2, 1/2)$ *a priori* distribution), which was also

suggested in [16]. The difference between the symmetric mutual information rate and the capacity for asymmetric channels is generally indistinguishable from the practical point of view. For example, in [33], it was shown that the ratio between the symmetric mutual information rate and the capacity is lower bounded by $\frac{\epsilon \ln 2}{2} \approx 0.942$. [34] further proved that the absolute difference is upper-bounded by 0.011 bit/symbol. Further discussion of capacity achieving codes with nonuniform *a priori* distributions can be found in [35] and [29].

Fig. 5(a) and (b) considers several fixed finite codes in z-channels. We arbitrarily select graphs from the code ensemble with codeword lengths $n = 1000$ and $n = 10000$. Then, with these graphs (codes) fixed, we find the corresponding parity matrix \mathbf{A} , use Gaussian elimination to find the generator matrix \mathbf{G} , and transmit different codewords by encoding equiprobably selected information messages. Belief propagation decoding is used with 40 iterations for each codeword. 10000 codewords are transmitted, and the overall bit/block-error rates versus different ϵ_1 are plotted for different code ensembles and codeword lengths. Our new density evolution predicts the waterfall region quite accurately when the BERs are of primary interest. Though there are still gaps between the performance of finite codes and our asymptotic thresholds, the performance gaps between different finite-length codes are very well predicted by the differences between their asymptotic thresholds. From the above observations and the underpinning theorems, we see that our new density evolution is a successful generalization of the traditional one from both practical and theoretical points of view.

Fig. 5(b) exhibits the block error rate of the same 10000-codeword simulation. The conjecture of bad block-error probabilities for $\lambda_2 > 0$ codes is confirmed. Besides the conjectured bad block-error probabilities, Fig. 5(a) and (b) also suggests that codes with $\lambda_2 = 0$ will have a better error floor compared to those with $\lambda_2 > 0$, which can be partly explained by the comparatively slow convergence speed stated in the sufficient stability condition for $\lambda_2 > 0$ codes. 12C is so far the best code we have for $\lambda_2 = 0$. However, its threshold is not as good as those of 12A and 12B. If good block-error rates and low error floor are our major concerns, 12C (or other codes with $\lambda_2 = 0$) can still be competitive choices. Recent results in [36] shows that the error floor for codes with $\lambda_2 > 0$ can be lowered by carefully arranging the degree-2 variable nodes in the corresponding graph while keeping a similar waterfall threshold.

Fig. 6(a) and (b) illustrates the BERs versus different BASC settings with 2000 transmitted codewords. Our computed density evolution threshold is again highly correlated with the

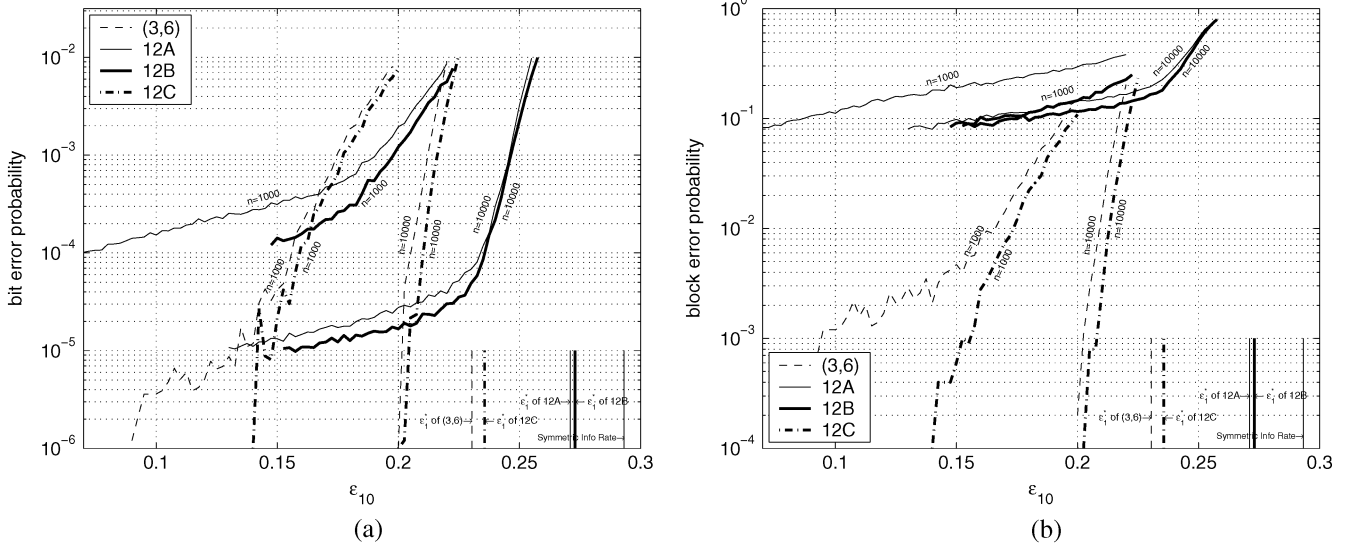


Fig. 5. Bit/block-error rates versus ϵ_1 with fixed $\epsilon_0 = 0.00001$. Computed thresholds for symmetric mutual information rate, (3, 6), 12A, 12B, and 12C codes are 0.2932, 0.2305, 0.2710, 0.2730, and 0.2356, respectively. Forty iterations of BP algorithms were performed. 10 000 codewords were used for the simulations. (a) Bit-error rates, (b) Block-error rates.

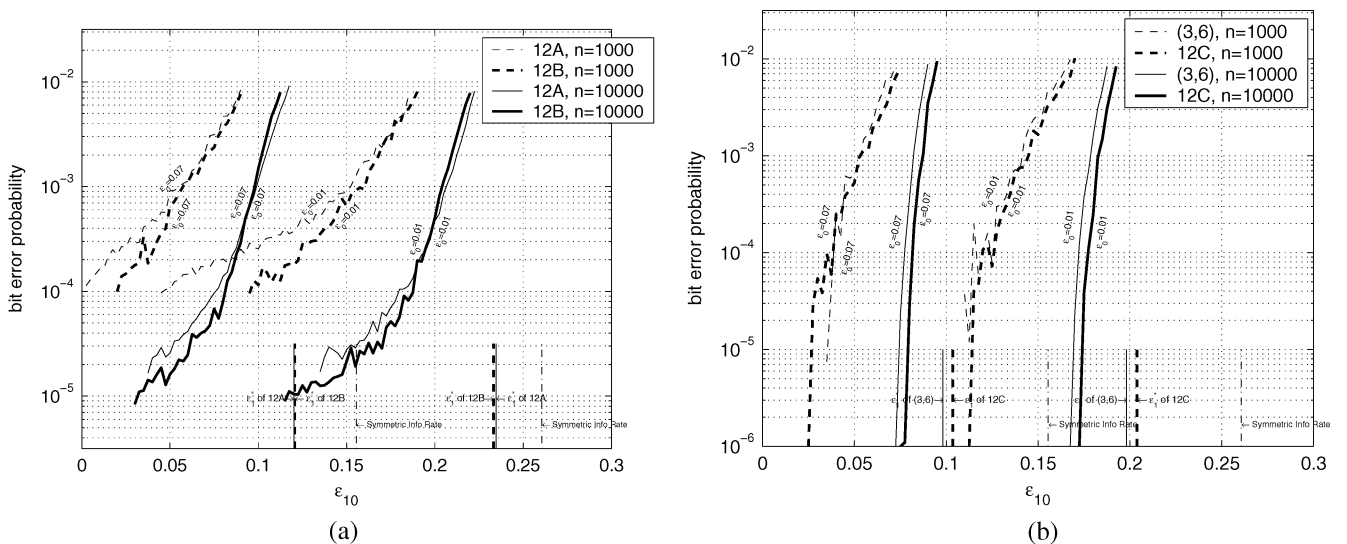


Fig. 6. Bit-error rates versus ϵ_1 for $\epsilon_0 = 0.01$ and $\epsilon_0 = 0.7$. The DE thresholds of (12A, 12B, 12C, (3, 6)) are (0.2346, 0.2332, 0.2039, 0.1981) for $\epsilon_0 = 0.01$ and (0.1202, 0.1206, 0.1036, 0.0982) for $\epsilon_0 = 0.07$. Forty iterations of BP algorithms were performed. 2000 codewords were used for the simulations. (a) 12A and 12B. (b) 12C and regular (3, 6) codes.

performance of finite-length codes for different asymmetric channel settings.

We close this section by highlighting two applications of our results.

- 1) Error Floor Analysis: “The error floor” is a characteristic of iterative decoding algorithms, which is of practical importance and may not be able to be determined solely by simulations. More analytical tools are needed to find error floors for corresponding codes. Our convergence rate statements in the sufficient stability condition may shed some light on finding codes with low error floors.
- 2) Capacity-Approaching Codes for General Nonstandard Channels: Various *very good* codes (capacity approaching) are known for standard channels, but very good codes for nonstandard channels are not yet known.

It is well known that one can construct capacity-approaching codes by incorporating symmetric-information-rate-approaching linear codes with the symbol mapper and demapper as an inner code [29], [35], [37]. Understanding density evolution for general memoryless channels allows us to construct such symmetric-information-rate-approaching codes (for asymmetric memoryless channels), and thus to find capacity-approaching codes after concatenating the inner symbol mapper and demapper. It is worth noting that intersymbol interference channels are dealt with by Kavčić *et al.* in [16] using the coset code approach. It will be of great help if a unified framework for asymmetric channels with memory can be found by incorporating both coset codes and codeword averaging approaches.

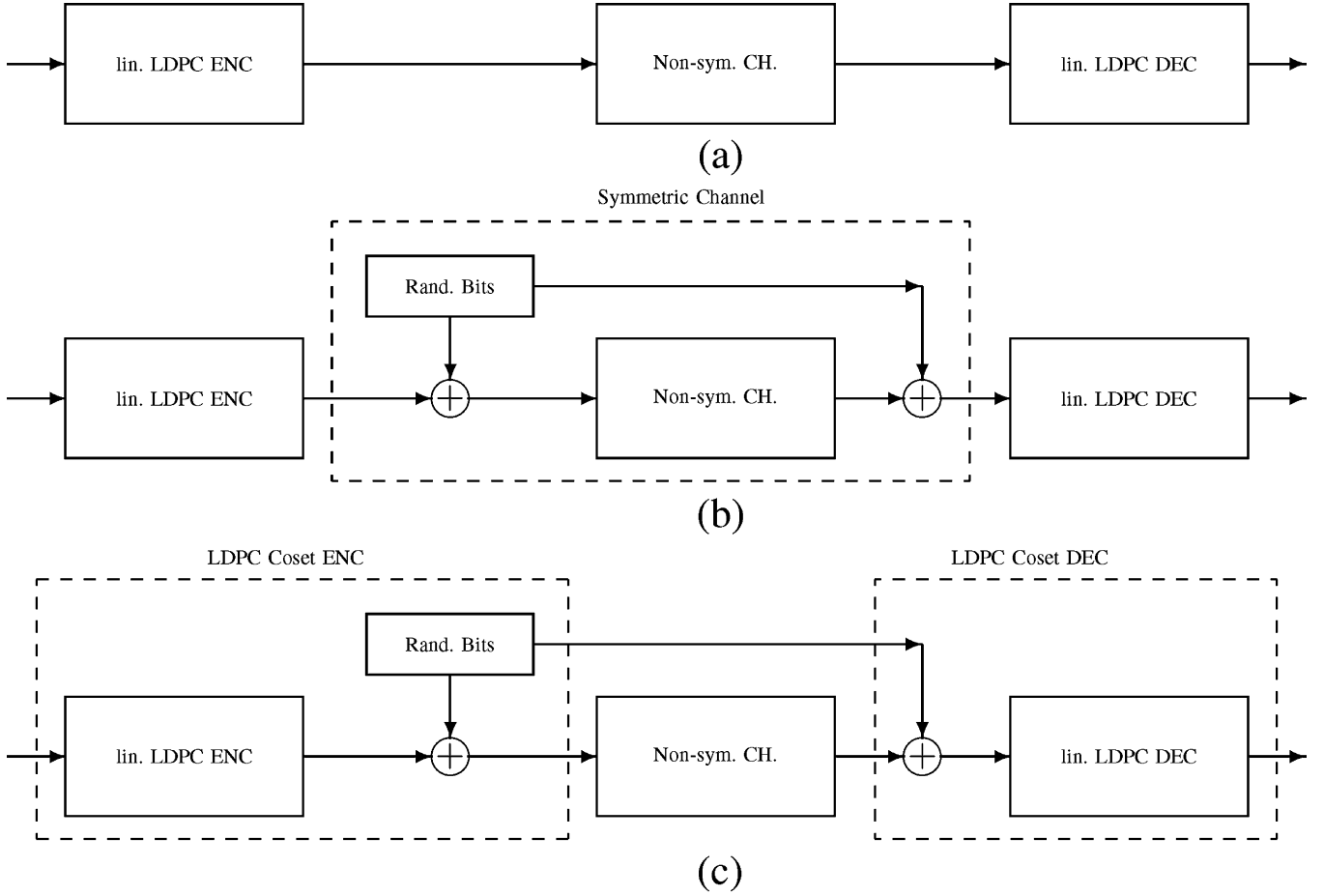


Fig. 7. Comparison of the approaches based on codeword averaging and the coset code ensemble. (a) Linear code ensemble versus asymmetric channels. (b) Linear code ensemble versus symmetrized channels. (c) Coset code ensemble versus asymmetric channels.

VII. FURTHER IMPLICATIONS OF GENERALIZED DENSITY EVOLUTION

A. Typicality of Linear LDPC Codes

One reason that asymmetric channels are often overlooked is we can always transform an asymmetric channel into a symmetric channel. Depending on different points of view, this channel-symmetrizing technique is termed the coset code argument [16] or dithering/the independent and identically distributed (i.i.d.) channel adapter [21], as illustrated in Fig. 7(c) and (b). Our generalized density evolution provides a simple way to directly analyze the linear LDPC code ensemble on asymmetric channels, as in Fig. 7(a).

As shown in Theorems 5 and 6, the necessary and sufficient stability conditions of linear LDPC codes for asymmetric channels, Fig. 7(a), are identical to those of the coset code ensemble, Fig. 7(c). Monte Carlo simulations based on finite-length codes ($n = 10^4$) [21] further show that the codeword-averaged performance in Fig. 7(a) is nearly identical⁶ to the performance of Fig. 7(c) when the same encoder/decoder pair is used. The above two facts suggest a close relationship between linear codes and the coset code ensemble, and it was conjectured in [21] that the scheme in Fig. 7(a) should always have the same/similar performance as those illustrated by Fig. 7(c). This short subsection is

⁶That is, it is within the precision of the Monte Carlo simulation.

devoted to the question whether the systems in Fig. 7(a) and (c) are equivalent in terms of performance. In sum, the performance of the linear code ensemble is very unlikely to be identical to that of the coset code ensemble. However, when the minimum

$$d_{c,\min} := \{k \in \mathbb{N} : \rho_k > 0\}$$

is sufficiently large, we can prove that their performance discrepancy is theoretically indistinguishable. In practice, the discrepancy for $d_{c,\min} \geq 6$ is $< 0.05\%$.

Let $P_{a.p.}^{(l)}(0) := P^{(l)}(0)$ and $P_{a.p.}^{(l)}(1) := P^{(l)}(1) \circ I^{-1}$ denote the two evolved densities with *aligned parity*, and similarly define $Q_{a.p.}^{(l)}(0) := Q^{(l)}(0)$ and $Q_{a.p.}^{(l)}(1) := Q^{(l)}(1) \circ I^{-1}$. Our main result in (15) can be rewritten in the following form:

$$\begin{aligned} P_{a.p.}^{(l)}(x) &= P_{a.p.}^{(0)}(x) \otimes \lambda \left(Q_{a.p.}^{(l-1)}(x) \right) \\ Q_{a.p.}^{(l-1)}(x) &= \Gamma^{-1} \left(\rho \left(\Gamma \left(\frac{P_{a.p.}^{(l-1)}(0) + P_{a.p.}^{(l-1)}(1)}{2} \right) \right) \right. \\ &\quad \left. + (-1)^x \rho \left(\Gamma \left(\frac{P_{a.p.}^{(l-1)}(0) - P_{a.p.}^{(l-1)}(1)}{2} \right) \right) \right). \end{aligned} \quad (23)$$

Let $p_{e,\text{linear}}^{(l)}$ denote the corresponding bit-error probability of the linear codes after l iterations. For comparison, the traditional

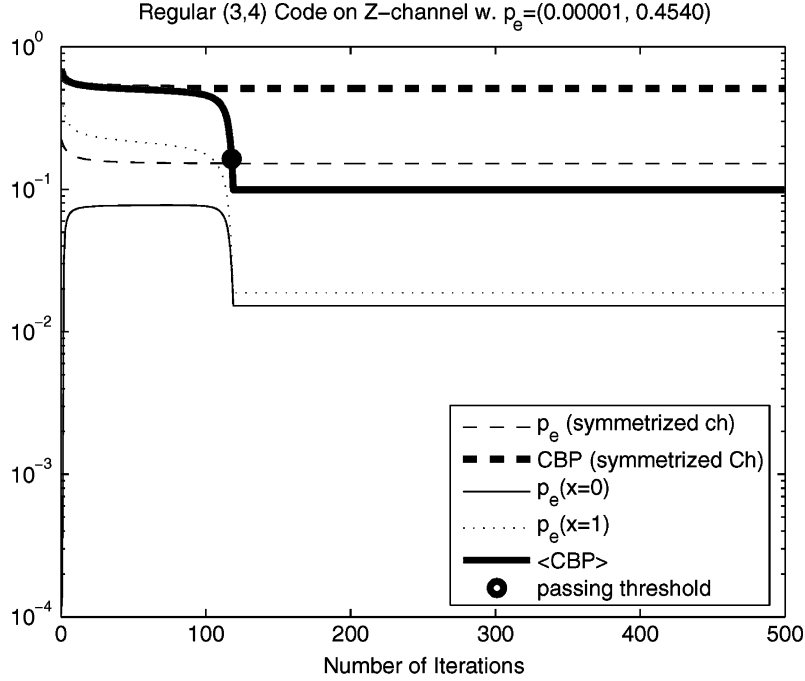


Fig. 8. Density evolution for z-channels with the linear code ensemble and the coset code ensemble.

formula of density evolution for the symmetrized channel (the coset code ensemble) is as follows:

$$\begin{aligned} P_{\text{coset}}^{(l)} &= P_{\text{coset}}^{(0)} \otimes \lambda \left(Q_{\text{coset}}^{(l-1)} \right) \\ Q_{\text{coset}}^{(l-1)} &= \Gamma^{-1} \left(\rho \left(\Gamma \left(P_{\text{coset}}^{(l-1)} \right) \right) \right) \end{aligned} \quad (24)$$

where

$$P_{\text{coset}}^{(0)} = \frac{\sum_{x=0,1} P_{\text{a.p.}}^{(0)}(x)}{2}.$$

Similarly, let $p_{e,\text{coset}}^{(l)}$ denote the corresponding bit error probability.

It is clear from (24) that when the channel of interest is symmetric, namely, $P_{\text{a.p.}}^{(0)}(0) = P_{\text{a.p.}}^{(0)}(1)$, then

$$P_{\text{coset}}^{(l)} = P_{\text{a.p.}}^{(l)}(0) = P_{\text{a.p.}}^{(l)}(1), \quad \text{for all } l \in \mathbb{N}.$$

However, for asymmetric channels, since the variable-node iteration involves convolution of several densities given the same x value, the difference between $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$ will be amplified after each variable node iteration. Hence, it is very unlikely that the decodable thresholds of linear codes and coset codes will be analytically identical, namely

$$\lim_{l \rightarrow \infty} p_{e,\text{linear}}^{(l)} = 0 \stackrel{?}{\iff} \lim_{l \rightarrow \infty} p_{e,\text{coset}}^{(l)} = 0.$$

Fig. 8 demonstrates the traces of the evolved densities for the regular (3, 4) code on z-channels. With the one-way crossover probability being 0.4540, the generalized density evolution for

 TABLE II
THRESHOLD COMPARISON $p_{1 \rightarrow 0}^*$ OF LINEAR AND COSET LDPC CODES ON Z-CHANNELS

(λ, ρ)	(x^2, x^3)	(x^2, x^5)
Linear	0.4540	0.2305
Coset	0.4527	0.2304

(λ, ρ)	$(x^2, 0.5x^2 + 0.5x^3)$	$(x^2, 0.5x^4 + 0.5x^5)$
Linear	0.5888	0.2689
Coset	0.5908	0.2690

linear codes is able to converge within 179 iterations, while the coset code ensemble shows no convergence within 500 iterations. This demonstrates the possible performance discrepancy, though we do not have analytical results proving that the latter will not converge after further iterations. Table II compares the decodable thresholds such that the density evolution enters the stability region within 100 iterations. We notice that the larger $d_{c,\text{min}}$ is, the smaller the discrepancy is. This phenomenon can be characterized by the following theorem.

Theorem 7: Consider asymmetric memoryless channels and a fixed pair of finite-degree polynomials λ and ρ . The shifted version of the check node polynomial is denoted as $\rho_{\Delta} = x^{\Delta} \cdot \rho$ where $\Delta \in \mathbb{N}$. Let $P_{\text{coset}}^{(l)}$ denote the evolved density from the coset code ensemble with degrees (λ, ρ_{Δ}) , and

$$\langle P^{(l)} \rangle = \frac{1}{2} \sum_{x=0,1} P_{\text{a.p.}}^{(l)}(x)$$

denote the averaged density from the linear code ensemble with degrees (λ, ρ_{Δ}) . For any $l_0 \in \mathbb{N}$, $\lim_{\Delta \rightarrow \infty} \langle P^{(l)} \rangle \stackrel{\mathcal{D}}{=} P_{\text{coset}}^{(l)}$ in distribution for all $l \leq l_0$, with the convergence rate for each iteration being $\mathcal{O}(\text{const}^{\Delta})$ for some $\text{const} < 1$.

Corollary 5 (The Typicality Results for Z-Channels): For any $\epsilon > 0$, there exists a $\Delta \in \mathbb{N}$ such that

$$\left| \sup \left\{ p_{1 \rightarrow 0} : \lim_{l \rightarrow \infty} p_{e, \text{linear}}^{(l)} = 0 \right\} - \sup \left\{ p_{1 \rightarrow 0} : \lim_{l \rightarrow \infty} p_{e, \text{coset}}^{(l)} = 0 \right\} \right| < \epsilon.$$

Namely, the asymptotic decodable thresholds of the linear and the coset code ensemble are arbitrarily close when the minimum check node degree $d_{c, \min}$ is sufficiently large.

Similar corollaries can be constructed for other channel models with different types of noise parameters, e.g., the σ^* in the composite BiAWGNC. A proof of Corollary 5 is found in Appendix III.

Proof of Theorem 7: Since the functionals in (23) and (24) are continuous with respect to convergence in distribution, we need only to show that $\forall l \in \mathbb{N}$

$$\begin{aligned} \lim_{\Delta \rightarrow \infty} Q_{\text{a.p.}}^{(l-1)}(0) &\stackrel{\mathcal{D}}{=} \lim_{\Delta \rightarrow \infty} Q_{\text{a.p.}}^{(l-1)}(1) \\ &\stackrel{\mathcal{D}}{=} \Gamma^{-1} \left(\rho \left(\Gamma \left(\frac{P_{\text{a.p.}}^{(l-1)}(0) + P_{\text{a.p.}}^{(l-1)}(1)}{2} \right) \right) \right) \\ &= \frac{Q_{\text{a.p.}}^{(l-1)}(0) + Q_{\text{a.p.}}^{(l-1)}(1)}{2} \end{aligned} \quad (25)$$

where $\stackrel{\mathcal{D}}{=}$ denotes convergence in distribution. Then, by inductively applying this weak convergence argument, for any bounded l_0 , $\lim_{\Delta \rightarrow \infty} \langle P^{(l)} \rangle \stackrel{\mathcal{D}}{=} P_{\text{coset}}^{(l)}$ in distribution for all $l \leq l_0$. Without loss of generality,⁷ we may assume $\rho_{\Delta} = x^{\Delta}$ and prove the weak convergence of distributions on the domain

$$\begin{aligned} \gamma(m) &:= \left(1_{\{m \leq 0\}}, \ln \coth \left| \frac{m}{2} \right| \right) \\ &= (\gamma_1, \gamma_2) \in \text{GF}(2) \times \mathbb{R}^+ \end{aligned}$$

on which the check node iteration becomes

$$\gamma_{\text{out}, \Delta} = \gamma_{\text{in}, 1} + \gamma_{\text{in}, 2} + \dots + \gamma_{\text{in}, \Delta}.$$

Let P'_0 denote the density of $\gamma_{\text{in}}(m)$ given that the distribution of m is $P_{\text{a.p.}}^{(l-1)}(0)$ and let P'_1 similarly correspond to $P_{\text{a.p.}}^{(l-1)}(1)$. Similarly, let $Q'_{0, \Delta}$ and $Q'_{1, \Delta}$ denote the output distributions on $\gamma_{\text{out}, \Delta}$ when the check node degree is $\Delta + 1$. It is worth noting that any pair of $Q'_{0, \Delta}$ and $Q'_{1, \Delta}$ can be mapped bijectively to the LLR distributions $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$.

Let $\Phi_{P'}(k, r) := \mathbb{E}_{P'} \{ (-1)^{k\gamma_1} e^{ir\gamma_2} \}$, $\forall k \in \mathbb{N}, r \in \mathbb{R}$, denote the Fourier transform of the density P' . Proving (25) is equivalent to showing that

$$\forall k \in \mathbb{N}, r \in \mathbb{R}, \quad \lim_{\Delta \rightarrow \infty} \Phi_{Q'_{0, \Delta}}(k, r) = \lim_{\Delta \rightarrow \infty} \Phi_{Q'_{1, \Delta}}(k, r).$$

However, to deal with the strictly growing average of the “limit distribution,” we concentrate on the distribution of the normalized output $\frac{\gamma_{\text{out}, \Delta}}{\Delta}$ instead. We then need to prove that

$$\forall k \in \mathbb{N}, r \in \mathbb{R}, \quad \lim_{\Delta \rightarrow \infty} \Phi_{Q'_{0, \Delta}} \left(k, \frac{r}{\Delta} \right) = \lim_{\Delta \rightarrow \infty} \Phi_{Q'_{1, \Delta}} \left(k, \frac{r}{\Delta} \right).$$

⁷We also need to assume that $\forall x, P_{\text{a.p.}}^{(l-1)}(x)(m = 0) = 0$ so that $\ln \coth \left| \frac{m}{2} \right| \in \mathbb{R}^+$ almost surely. This assumption can be relaxed by separately considering the event that $m_{\text{in}, i} = 0$ for some $i \in \{1, \dots, d_c - 1\}$.

We first note that for all $x = 0, 1$, $Q'_{x, \Delta}$ is the averaged distribution of $\gamma_{\text{out}, \Delta}$ when the inputs $\gamma_{\text{in}, i}$ are governed by $P_{\text{a.p.}}^{(l)}(x_i)$ satisfying $\sum_{i=1}^{\Delta} x_i = x$. From this observation, we can derive the following iterative equations: $\forall \Delta \in \mathbb{N}$

$$\begin{aligned} \Phi_{Q'_{0, \Delta}} \left(k, \frac{r}{\Delta} \right) &= \frac{\Phi_{Q'_{0, \Delta-1}} \left(k, \frac{r}{\Delta} \right) \Phi_{P'_0} \left(k, \frac{r}{\Delta} \right) + \Phi_{Q'_{1, \Delta-1}} \left(k, \frac{r}{\Delta} \right) \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right)}{2} \\ \Phi_{Q'_{1, \Delta}} \left(k, \frac{r}{\Delta} \right) &= \frac{\Phi_{Q'_{0, \Delta-1}} \left(k, \frac{r}{\Delta} \right) \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right) + \Phi_{Q'_{1, \Delta-1}} \left(k, \frac{r}{\Delta} \right) \Phi_{P'_0} \left(k, \frac{r}{\Delta} \right)}{2}. \end{aligned}$$

By induction, the difference thus becomes

$$\begin{aligned} \Phi_{Q'_{0, \Delta}} \left(k, \frac{r}{\Delta} \right) - \Phi_{Q'_{1, \Delta}} \left(k, \frac{r}{\Delta} \right) &= \left(\Phi_{Q'_{0, \Delta-1}} \left(k, \frac{r}{\Delta} \right) - \Phi_{Q'_{1, \Delta-1}} \left(k, \frac{r}{\Delta} \right) \right) \\ &\quad \cdot \left(\frac{\Phi_{P'_0} \left(k, \frac{r}{\Delta} \right) - \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right)}{2} \right) \\ &= 2 \left(\frac{\Phi_{P'_0} \left(k, \frac{r}{\Delta} \right) - \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right)}{2} \right)^{\Delta}. \end{aligned} \quad (26)$$

By Taylor’s expansion and the BASC decomposition argument in [30], we can show that for all $k \in \mathbb{N}, r \in \mathbb{R}$, and for all possible P'_0 and P'_1 , the quantity in (26) converges to zero with convergence rate $\mathcal{O}(\text{const}^{\Delta})$ for some $\text{const} < 1$. A detailed derivation of the convergence rate is given in Appendix IV. Since the limit of the right-hand side of (26) is zero, the proof of weak convergence is complete. The exponentially fast convergence rate $\mathcal{O}(\text{const}^{\Delta})$ also justifies the fact that even for moderate $d_{c, \min} \geq 6$, the performances of linear and coset LDPC codes are very close. \square

Remark 1: Consider any nonperfect message distribution, namely, $\exists x_0$ such that $P_{\text{a.p.}}^{(l-1)}(x_0) \neq \delta_{\infty}$. A persistent reader may notice that $\forall x, \lim_{\Delta \rightarrow \infty} Q_{\text{a.p.}}^{(l-1)}(x) \stackrel{\mathcal{D}}{=} \delta_0$, namely, as Δ becomes large, all information is erased after passing a check node of large degree. If this convergence (erasure effect) occurs earlier than the convergence of $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$, the performances of linear and coset LDPC codes are “close” only when the code is “useless.”⁸ To quantify the convergence rate, we consider again the distributions on γ and their Fourier transforms. For the average of the output distributions $Q_{\text{a.p.}}^{(l-1)}(x)$, we have

$$\begin{aligned} \frac{\Phi_{Q'_{0, \Delta}} \left(k, \frac{r}{\Delta} \right) + \Phi_{Q'_{1, \Delta}} \left(k, \frac{r}{\Delta} \right)}{2} &= \left(\frac{\Phi_{Q'_{0, \Delta-1}} \left(k, \frac{r}{\Delta} \right) + \Phi_{Q'_{1, \Delta-1}} \left(k, \frac{r}{\Delta} \right)}{2} \right) \\ &\quad \cdot \left(\frac{\Phi_{P'_0} \left(k, \frac{r}{\Delta} \right) + \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right)}{2} \right) \\ &= \left(\frac{\Phi_{P'_0} \left(k, \frac{r}{\Delta} \right) + \Phi_{P'_1} \left(k, \frac{r}{\Delta} \right)}{2} \right)^{\Delta}. \end{aligned} \quad (27)$$

⁸To be more precise, it corresponds to an extremely high-rate code and the information is erased after every check node iteration.

By Taylor's expansion and the BASIC decomposition argument, one can show that the limit of (27) exists and the convergence rate is $\mathcal{O}(\Delta^{-1})$. (A detailed derivation is included in Appendix IV.) This convergence rate is much slower than the exponential rate $\mathcal{O}(\text{const}^\Delta)$ in the Proof of Theorem 7. Therefore, we do not need to worry about the case in which the required Δ for the convergence of $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$ is excessively large so that $\forall x \in \text{GF}(2), Q_{\text{a.p.}}^{(l-1)}(x) \stackrel{\mathcal{D}}{\approx} \delta_0$.

Remark 2: The intuition behind Theorem 7 is that when the minimum d_c is sufficiently large, the parity-check constraint becomes relatively less stringent. Thus we can approximate the density of the outgoing messages for linear codes by assuming all bits involved in that particular parity-check equation are "independently" distributed among $\{0,1\}$, which leads to the formula for the coset code ensemble. On the other hand, extremely large d_c is required for a check node iteration to completely destroy all information coming from the previous iteration. This explains the difference between their convergence rates: $\mathcal{O}(\text{const}^\Delta)$ versus $\mathcal{O}(\Delta^{-1})$.

Fig. 9 illustrates the weak convergence predicted by Theorem 7 and depicts the convergence rates of

$$Q_{\text{a.p.}}^{(l-1)}(0) \rightarrow Q_{\text{a.p.}}^{(l-1)}(1) \quad \text{and} \quad \frac{Q_{\text{a.p.}}^{(l-1)}(0) + Q_{\text{a.p.}}^{(l-1)}(1)}{2} \rightarrow \delta_0.$$

Our typicality result can be viewed as a complementing theorem of the concentration theorem in [16, Corollary 2.2], where a constructive method of finding a typical coset-defining syndrome is not specified. Besides the theoretical importance, we are now on a solid basis to interchangeably use the linear LDPC codes and the LDPC coset codes when the check node degree is of moderate size. For instance, from the implementation point of view, the hardware uniformity of linear codes makes them a superior choice compared to any other coset code. We can then use the fast density evolution [38] plus the coset code ensemble to optimize the degree distribution for the linear LDPC codes. Or instead of simulating the codeword-averaged performance of linear LDPC codes, we can simulate the error probability of the all-zero codeword in the coset code ensemble, in which the efficient LDPC encoder [8] is not necessary.

B. Revisiting the Belief Propagation Decoder

Two known facts about the BP algorithm and the density evolution method are as follows. First, the BP algorithm is optimal for any cycle-free network, since it exploits the independence of the incoming LLR message. Second, by the cycle-free convergence theorem, the traditional density evolution is able to predict the behavior of the BP algorithm (designed for the tree structure) for l_0 iterations, even when we are focusing on a Tanner graph of a finite-length LDPC code, which inevitably has many cycles. The performance of BP, predicted by density evolution, is outstanding so that we "implicitly assume" that the BP (designed for the tree structure) is optimal for the first l_0 iterations in terms of minimizing the *codeword-averaged* BER. Theoretically, to be able to minimize the codeword-averaged BER, the optimal decision rule inevitably must exploit the global knowledge about all possible codewords, which is, however, not available to the BP decoder. A question of interest is whether BP is indeed optimal for the first l_0 iterations?

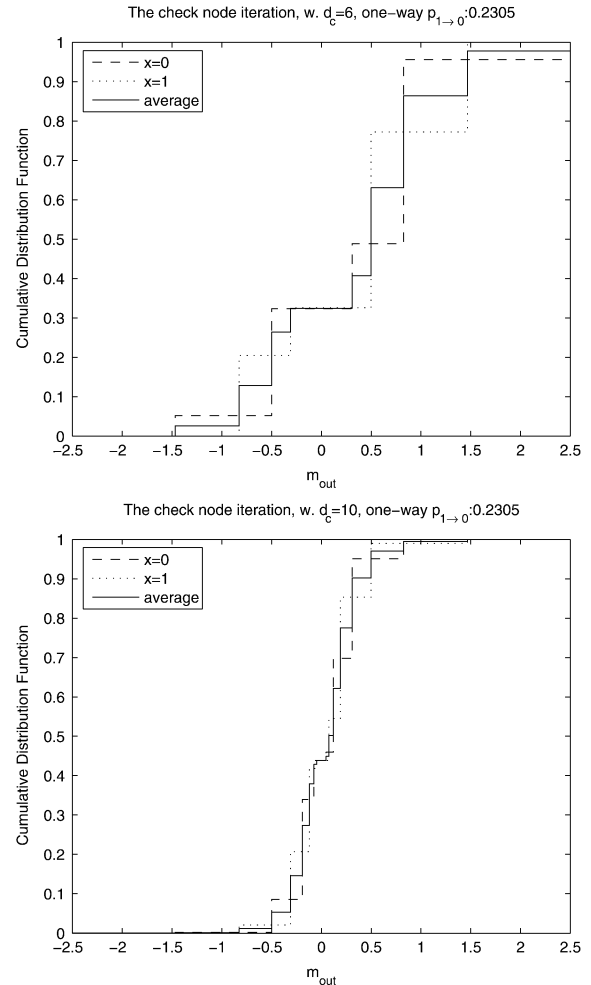


Fig. 9. Illustration of the weak convergence of $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$. One can see that the convergence of $Q_{\text{a.p.}}^{(l-1)}(0)$ and $Q_{\text{a.p.}}^{(l-1)}(1)$ is faster than the convergence of $(Q_{\text{a.p.}}^{(l-1)}(0) + Q_{\text{a.p.}}^{(l-1)}(1))/2$ and δ_0 .

Namely, with only local knowledge about possible codewords, whether BP has the same performance as the optimal detector with the global information about the entire codebook and unlimited computational power when we are only interested in the first l_0 iterations? The answer is a straightforward corollary to Theorem 2, the convergence to perfect projection, which provides the missing link regarding the optimality of BP when only local observations (on the \mathcal{N}^{2l_0}) are available.

Theorem 8 (Local Optimality of the BP Decoder): Fix $i, l_0 \in \mathbb{N}$. For sufficiently large codeword length n , almost all instances in the random code ensemble have the property that the BP decoder for x_i after l_0 iterations, $\hat{X}_{\text{BP}}(\mathbf{Y}^{l_0})$, coincides with the optimal MAP bit detector $\hat{X}_{\text{MAP}, l_0}(\mathbf{Y}^{l_0})$, where l_0 is a fixed integer. The MAP bit detector $\hat{X}_{\text{MAP}, l_0}(\cdot)$ uses the same number of observations as in $\hat{X}_{\text{BP}}(\cdot)$ but is able to exploit the global knowledge about the entire codebook.

Proof: When the support tree $\mathcal{N}_{(i,j)}^{2l_0}$ is perfectly projected, the local information about the tree-satisfying strings is equivalent to the global information about the entire codebook. Therefore, the extra information about the entire codebook does not benefit the decision maker, and $\hat{X}_{\text{BP}}(\cdot) = \hat{X}_{\text{MAP}, l_0}(\cdot)$. Theorem 2 shows that $\mathcal{N}_{(i,j)}^{2l_0}$ converges to perfect projection in probability, which in turn implies that for sufficiently large

$$\mathbf{A} = \left(\begin{array}{c|c|c} \mathbf{I}_{1 \times 1} \otimes (1, 1, 1, 1, 1) & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{I}_{5 \times 5} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \mathbf{I}_{10 \times 10} \otimes (1, 1, 1, 1) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{40 \times 40} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \mathbf{A}'_{80 \times (n-45)} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}''_{\left(\frac{3n}{5}-91\right) \times (n-45)} \end{array} \right)$$

$$\mathbf{A} = \left(\begin{array}{c|c|c} \mathbf{A}_{l_n} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} \mid \mathbf{I}_{(5 \cdot 8^{l_n-1}) \times (5 \cdot 8^{l_n-1})} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \mathbf{I}_{(10 \cdot 8^{l_n-1}) \times (10 \cdot 8^{l_n-1})} (1, 1, 1) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{(5 \cdot 8^{l_n}) \times (5 \cdot 8^{l_n})} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \mathbf{A}' \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}'' \end{array} \right),$$

Using \otimes to denote the Kronecker product (whether it represents convolution or Kronecker product should be clear from the context), (29) can be further expressed as shown by the first expression at the top of the page, where $\mathbf{I}_{a \times a}$ denotes the $a \times a$ identity matrix, $\left(\mathbf{A}'_{80 \times (n-45)} \right)$ is the incidence matrix of the equiprobable, bipartite subgraph, in which all $(n-45)$ variable nodes have degree d_v , 80 check nodes have degree $d_c - 1$, and $\left(\frac{3n}{5} - 91\right)$ check nodes have degree d_c . Conditioning on a more general event that the graph is cycle-free until depth $2(l_n + 1)$ rather than $2 \cdot 2$, we will have the second expression at the top of the page, where \mathbf{A}_{l_n} corresponds to the incidence matrix of the cycle-free graph of depth $2l_n$. $\left(\mathbf{A}' \right)$ is the incidence matrix with rows (check nodes) in \mathbf{A}' and \mathbf{A}'' having degree $(d_c - 1)$ and d_c . For convenience, we denote the blocks in \mathbf{A} as

$$\mathbf{A} = \left(\begin{array}{c|c|c|c} \mathbf{A}_{l_n} & \mathbf{0} & \mathbf{0} & \\ \hline \mathbf{0} \mid \mathbf{T}_{l_n} & \mathbf{U}_{l_n+1} & \mathbf{0} & \\ \hline \mathbf{0} & \mathbf{T}_{l_n+1} & \mathbf{A}' & \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}'' & \end{array} \right).$$

Then \mathcal{N}^{2l_n} is not perfectly projected if and only if there exists a nonzero row vector $(\mathbf{r}|\mathbf{0}|\mathbf{0})$ such that

$$(\mathbf{r}|\mathbf{0}|\mathbf{0}) \in \text{RowSpace} \left(\begin{array}{c|c|c|c} \mathbf{0} & \mathbf{T}_{l_n} & \mathbf{U}_{l_n+1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{T}_{l_n+1} & \mathbf{A}' & \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}'' & \end{array} \right) \quad (30)$$

and

\mathbf{r} is not in the row space of \mathbf{A}_{l_n}

or equivalently $(\mathbf{r}|\mathbf{0}|\mathbf{0})$ is not in $\text{RowSpace}(\mathbf{A}_{l_n}|\mathbf{0}|\mathbf{0})$. (31)

Equations (31) and (30) say that there exists a constraint \mathbf{r} on the variable nodes of \mathcal{N}^{2l_n} , which is not from the linear combination of those check node equations within \mathcal{N}^{2l_n} , but rather is imposed by the parity-check equations outside \mathcal{N}^{2l_n} . It can be easily proved that if the matrix $\left(\mathbf{A}' \right)$ is of full row rank, then

no such \mathbf{r} exists and \mathcal{N}^{2l_n} is perfectly projected.¹⁰ Instead of proving $\left(\mathbf{A}' \right)$ is of full rank, we take a different approach, which takes care of the constraint propagation.

From (30), we know that, for $(\mathbf{r}|\mathbf{0}|\mathbf{0})$ to exist, there must exist a nonzero row vector $(\mathbf{0}|\mathbf{s}|\mathbf{0})$ such that

$$(\mathbf{0}|\mathbf{s}|\mathbf{0}) \in \text{RowSpace} \left(\begin{array}{c|c|c} \mathbf{0} & \mathbf{T}_{l_n+1} & \mathbf{A}' \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{A}'' \end{array} \right) \quad (32)$$

and

$$\begin{aligned} \mathbf{s} &\in \text{RowSpace}(\mathbf{U}_{l_n+1}) \\ &= \text{RowSpace}(\mathbf{I}_{(10 \cdot 8^{l_n-1}) \times (10 \cdot 8^{l_n-1})} \otimes (1, 1, 1, 1)). \end{aligned} \quad (33)$$

From (33), the 1's in \mathbf{s} must be aligned such that four neighboring bits should have the same value; for example,

$$\mathbf{s} = (111100001111000000001111 \dots 00001111).$$

Any nonzero \mathbf{s} satisfying (32) is generated by \mathbf{T}_{l_n+1} . By applying the row symmetry in \mathbf{A}' , we see that the 1's in any \mathbf{s} are uniformly distributed among all these $5 \cdot 8^{l_n}$ bits. Therefore, conditioning on the event that there exists a not-all-one \mathbf{s} satisfying (32), the probability that \mathbf{s} satisfies (33) is

$$\begin{aligned} &\mathbb{P}(\mathbf{s} \text{ satisfies (33)} \mid \exists \mathbf{s} \text{ satisfies (32) and is not } \mathbf{1}) \\ &= \mathbb{P}(\text{the 1's in } \mathbf{s} \text{ are aligned} \mid \exists \mathbf{s} \text{ satisfies (32) and is not } \mathbf{1}) \\ &= \sum_{a=1}^{10 \cdot 8^{l_n-1} - 1} \frac{\binom{10 \cdot 8^{l_n-1}}{a}}{\binom{5 \cdot 8^{l_n}}{4a}} \cdot \mathbb{P}(\text{there are } 4a \text{ ones in } \mathbf{s}) \\ &\leq \frac{\binom{10 \cdot 8^{l_n-1}}{1}}{\binom{5 \cdot 8^{l_n}}{4}} = \mathcal{O} \left(\left(\frac{1}{((d_v - 1)(d_c - 1))^{l_n}} \right)^{d_c - 2} \right) \\ &= \mathcal{O} \left(n^{-\frac{4}{5}(d_c - 2)} \right). \end{aligned} \quad (34)$$

The last inequality follows from the assumption that \mathbf{s} is neither all-zero nor all-one. The reason why we can exclude the case that \mathbf{s} is all-one is that, if d_v is odd, then there is an even number

¹⁰Unfortunately, $\left(\mathbf{A}' \right)$ is not of full row rank. We can only show that with sufficiently large n , the row rank of $\left(\mathbf{A}' \right)$ converges to the number of rows minus one by methods similar to those in [39]. A simple constraint propagation argument is still necessary for this approach.

of 1's in each column of \mathbf{T}_{l_n} . Since there is only one 1 in each column of \mathbf{U}_{l_n+1} , by (30), an all-one \mathbf{s} can only generate an all-zero \mathbf{r} , which puts no constraints on $\mathcal{N}_{(i_0, j_0)}^{2l_n}$. If d_v is even, by the same reasoning, an all-one \mathbf{s} will generate \mathbf{r} of the form

$$(00 \cdots 0 \overbrace{11 \cdots 1}^{5 \cdot 8^{l_n-1}}).$$

Nevertheless, when d_v is even, this specific type of \mathbf{r} is in the row space of \mathbf{A}_{l_n} , which does not fulfill the requirement in (31). From the above reasoning, we can exclude the all-one \mathbf{s} .

Let m_r denote the number of rows of $\begin{pmatrix} \mathbf{A}' \\ \mathbf{A}'' \end{pmatrix}$ minus $\text{Rank}(\begin{pmatrix} \mathbf{A}' \\ \mathbf{A}'' \end{pmatrix})$. The number of vectors \mathbf{s} satisfying (32) is upper-bounded by 2^{m_r} . By (34), Proposition 4 (which will be formally stated and proved later), and the union bound, we have

$$\begin{aligned} & \text{P} \left(\mathcal{N}_{(i_0, j_0)}^{2l_n} \text{ is not perfectly projected} \mid \mathcal{N}_{(i_0, j_0)}^{2(l_n+1)} \text{ is cycle-free} \right) \\ &= \text{P}(\exists \mathbf{r} \text{ satisfying (30) and (31)}) \\ &= \text{P}(\exists \mathbf{s}, \text{ which satisfies (32) and (33) but is not all-one}) \\ &\leq n^{1.1} \cdot \text{P}(\mathbf{s} \text{ satisfies (33)} \mid \exists \mathbf{s} \text{ satisfies (32) and is not } \mathbf{1}) \\ &\quad \cdot \text{P}(\# \text{ of } \mathbf{s} \text{ is smaller than } n^{1.1}) \\ &\quad + \text{P}(\# \text{ of } \mathbf{s} \text{ is larger than } n^{1.1}) \\ &= n^{1.1} \mathcal{O} \left(n^{-\frac{4}{3}(d_c-2)} \right) + \text{P}(2^{m_r} > n^{1.1}) \\ &= \mathcal{O}(n^{-0.1}), \quad \forall d_c \geq 5. \end{aligned} \quad (35)$$

To prove the case $d_c < 5$, we focus on the probability that the constraints propagate two levels rather than just one level, i.e., instead of (28), we focus on proving the following statement:

$$\begin{aligned} & \text{P} \left(\mathcal{N}_{(i_0, j_0)}^{2l_n} \text{ is perfectly projected} \mid \mathcal{N}_{(i_0, j_0)}^{2(l_n+2)} \text{ is cycle-free} \right) \\ &= 1 - \mathcal{O}(n^{-0.1}). \end{aligned}$$

Most of the analysis remains the same. The conditional probability in (34) will be replaced by

$$\begin{aligned} & \text{P}((\mathbf{0}|\mathbf{s}|\mathbf{0}) \text{ is able to propagate two levels} \mid \exists \mathbf{s} \text{ satisfying (32)}) \\ &= \text{P}((\mathbf{0}|\mathbf{s}|\mathbf{0}) \text{ propagates the second level} \mid \\ &\quad (\mathbf{0}|\mathbf{s}|\mathbf{0}) \text{ propagates the first level, } \exists \mathbf{s} \text{ satisfying (32)}) \\ &\quad \cdot \text{P}((\mathbf{0}|\mathbf{s}|\mathbf{0}) \text{ propagates the first level} \mid \exists \mathbf{s} \text{ satisfying (32)}) \\ &= \sum_{a,b} \frac{\binom{10 \cdot 8^{l_n-1}}{a}}{\binom{5 \cdot 8^{l_n}}{4a}} \frac{\binom{10 \cdot 8^{l_n}}{b}}{\binom{5 \cdot 8^{l_n+1}}{4b}} \\ &\quad \cdot \text{P}(4a \text{ and } 4b \text{ 1's to propagate the second and first levels}) \\ &\stackrel{(a)}{\leq} \frac{\binom{10 \cdot 8^{l_n-1}}{1}}{\binom{5 \cdot 8^{l_n}}{4}} \frac{\binom{10 \cdot 8^{l_n}}{4}}{\binom{5 \cdot 8^{l_n+1}}{4 \cdot 4}} \\ &= \mathcal{O} \left(n^{-\frac{4}{3}(d_c^2 - 2d_c)} \right) \end{aligned}$$

where the inequality marked (a) follows from an analysis of the minimum number of bits required for the constraint propagation similar to that for the single-level case. By this stronger inequality and a bounding inequality similar to that in (35), we thus complete the proof of the case $d_c \geq 3$ for all regular codes of practical interest. \square

Note: This constraint propagation argument shows that the convergence to a perfectly projected tree is very strong. Even

for codes with redundant check node equations (not of full row rank), it is probabilistically hard for the external constraints to propagate inside and impose on the variable nodes within \mathcal{N}^{2l} . This property is helpful when we consider BP decoding on the alternative graph representation as in [40].

We close this appendix by stating the proposition regarding m_r , the number of linearly dependent rows in $\begin{pmatrix} \mathbf{A}' \\ \mathbf{A}'' \end{pmatrix}$. The proof is left to Appendix II.

Proposition 4: Consider the semiregular code ensemble $\mathcal{C}_{m', m''}^n(d_v, d_c)$ generated by equiprobable edge permutation on a bipartite graph with n variable nodes of degree d_v , and m' and m'' check nodes with respective degrees $(d_c - 1)$ and d_c . The corresponding parity-check matrix is $\mathbf{A} = \begin{pmatrix} \mathbf{A}' \\ \mathbf{A}'' \end{pmatrix}$. With m_r denoting the number of linearly dependent rows in \mathbf{A} , i.e., $m_r := m' + m'' - \text{Rank}(\mathbf{A})$, we have

$$\mathbb{E}\{2^{m_r}\} = \mathcal{O}(n)$$

which automatically implies

$$\text{P}(2^{m_r} > n^{1+\alpha}) = \text{P}(m_r > ((1+\alpha) \ln n)/(\ln 2)) = \mathcal{O}(n^{-\alpha})$$

for any $\alpha > 0$.

Corollary 7: Let R denote the rate of a regular LDPC code ensemble $\mathcal{C}^n(d_v, d_c)$, i.e., $R = (n - \text{Rank}(\mathbf{A}))/n$, where \mathbf{A} is the corresponding parity-check matrix. Then R converges to $(n - m)/n$ in L^1 , i.e.,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \left| R - \frac{n - m}{n} \right| \right\} = 0.$$

Proof: It is obvious that $R \geq (n - m)/n$. To show that

$$\limsup_{n \rightarrow \infty} \mathbb{E}\{R - (n - m)/n\} = 0$$

we let $m_1 = 0$ and rewrite

$$R = (n - \text{Rank}(\mathbf{A}))/n = (n - m)/n + m_r/n.$$

By Proposition 4 and the fact that $m_r/n \leq 1$, we have $\lim_{n \rightarrow \infty} \mathbb{E}\{m_r/n\} = 0$. This completes the proof. \square

A stronger version of the convergence of R with respect to the block length n can be found in [39].

APPENDIX II PROOF OF PROPOSITION 4

We finish the proof of Proposition 4 by first stating the following lemma.

Lemma 2: For all $0 < k \in \mathbb{N}, 0 < i < n \in \mathbb{N}$, we have

$$\frac{\binom{n}{i}}{\binom{kn}{ki}} \leq \sqrt{k} e^{\frac{1}{6}} 2^{-(k-1)nH_2(i/n)}.$$

Proof: By Stirling's double inequality

$$\sqrt{2\pi n}^{(n+\frac{1}{2})} e^{-(n+\frac{1}{12n+1})} < n! < \sqrt{2\pi n}^{(n+\frac{1}{2})} e^{-(n+\frac{1}{12n})}$$

we can prove

$$e^{-\frac{1}{6}} < \frac{\binom{n}{\theta n}}{\frac{1}{\sqrt{2\pi}} 2^{nH_2(i/n)} \sqrt{\frac{n}{i(n-i)}}} < 1,$$

which immediately leads to the desired inequality. \square

Proof of Proposition 4: By the definition of m_r , we have $2^{m_r} = (\text{total \# of codewords})/2^{n-m}$, where $m = m' + m''$. Then

$$\mathbb{E} \left\{ \frac{2^{m_r}}{n} \right\} \leq \frac{2}{n2^{n-m}} + \sum_{i=1}^{n-1} \frac{\mathbb{E}\{\# \text{ of codewords of weight } i\}}{n2^{n-m}}.$$

Using the enumerating function as in [41], [39] and define $g(x)$ as

$$g(i, x) := \frac{\left(\frac{(1+x)^{d_c-1} + (1-x)^{d_c-1}}{2} \right)^{m'} \left(\frac{(1+x)^{d_c} + (1-x)^{d_c}}{2} \right)^{m''}}{x^{i d_v}}$$

the above quantity can be further upper bounded as follows.

$$\begin{aligned} & \mathbb{E} \left\{ \frac{2^{m_r}}{n} \right\} \\ & \leq \frac{2}{n2^{n-m}} + \sum_{i=1}^{n-1} \frac{\binom{n}{i} \inf_{x>0} g(i, x)}{n2^{n-m}} \\ & \leq \frac{2}{n2^{n-m}} + \sqrt{d_v} e^{1/6} \\ & \quad \times \sum_{i=1}^{n-1} \frac{1}{n} 2^{-(d_v-1)nH_2(i/n)} \frac{\inf_{x>0} g(i, x)}{2^{n-m}} \\ & \leq \frac{2}{n2^{n-m}} + \sqrt{d_v} e^{1/6} \sum_{i=1}^{n-1} \frac{1}{n} 2^{-d_v n H_2(i/n)} \frac{\inf_{x>0} g(i, x)}{2^{n-m}} \end{aligned} \quad (36)$$

where the second inequality follows from Lemma 2 and the third inequality follows from the fact that the binary entropy function $H_2(\cdot)$ is upper-bounded by 1.

By defining

$$f_n(i, x) := 2^{m-d_v n H_2(i/n)} g(i, x)$$

the summation in (36) is upper-bounded¹¹ by

$$\max_{i \in [0, n]} \inf_{x>0} f_n(i, x) \leq \inf_{x>0} \max_{i \in [0, n]} f_n(i, x) \leq \max_{i \in [0, n]} f_n(i, 1).$$

By simple calculus, $\max_{i \in [0, n]} f_n(i, 1)$ is attained when $i = n/2$. Since $f_n(n/2, 1) = 1$, the summation in (36) is bounded by 1 for all n , and therefore,

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left\{ \frac{2^{m_r}}{n} \right\} \leq \sqrt{d_v} e^{1/6}.$$

The proof is complete. \square

APPENDIX III

PROOF OF COROLLARY 5

We prove one direction that

$$\begin{aligned} p_{1 \rightarrow 0, \text{linear}}^* & := \sup \left\{ p_{1 \rightarrow 0} > 0 : \lim_{l \rightarrow \infty} p_{e, \text{linear}}^{(l)} = 0 \right\} \\ & > \sup \left\{ p_{1 \rightarrow 0} > 0 : \lim_{l \rightarrow \infty} p_{e, \text{coset}}^{(l)} = 0 \right\} - \epsilon \\ & := p_{1 \rightarrow 0, \text{coset}}^* - \epsilon. \end{aligned}$$

¹¹The range of i is expanded here from a discrete integer set to a continuous interval.

The other direction that $p_{1 \rightarrow 0, \text{coset}}^* > p_{1 \rightarrow 0, \text{linear}}^* - \epsilon$ can be easily obtained by symmetry.

By definition, for any $\epsilon > 0$, we can find a sufficiently large $l_0 < \infty$ such that for a z -channel with one-way crossover probability $p_{1 \rightarrow 0} := p_{1 \rightarrow 0, \text{coset}}^* - \epsilon$, $P_{\text{coset}}^{(l_0)}$ is in the interior of the stability region. We note that the stability region depends only on the Bhattacharyya noise parameter of $P_{\text{coset}}^{(l_0)}$, which is a continuous function with respect to convergence in distribution. Therefore, by Theorem 7, there exists a $\Delta \in \mathbb{N}$ such that $\langle P^{(l_0)} \rangle$ is also in the stability region. By the definition of the stability region, we have $\lim_{l \rightarrow \infty} p_{e, \text{linear}}^{(l)} = 0$, which implies $p_{1 \rightarrow 0, \text{linear}}^* \geq p_{1 \rightarrow 0}$. The proof is thus complete.

APPENDIX IV

THE CONVERGENCE RATES OF (26) AND (27)

For (26), we will consider the cases that $k = 0$ and $k = 1$ separately. By the BASC decomposition argument, namely, all asymmetric channels can be decomposed as the probabilistic combination of many BASCs, we can limit our attention to simple BASCs rather than general memoryless asymmetric channels. Suppose $P_{\text{a.p.}}^{(l-1)}(0)$ and $P_{\text{a.p.}}^{(l-1)}(1)$ correspond to a BASC with crossover probabilities ϵ_0 and ϵ_1 . Without loss of generality, we may assume $\epsilon_0 + \epsilon_1 < 1$ because of the previous assumption that $\forall x \in \text{GF}(2)$, $P_{\text{a.p.}}^{(l-1)}(x)(m=0) = 0$. We then have

$$\begin{aligned} \Phi_{P_0'} \left(k, \frac{r}{\Delta} \right) & = (1 - \epsilon_0) e^{i \frac{r}{\Delta} \ln \frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1}} \\ & \quad + (-1)^k \epsilon_0 e^{i \frac{r}{\Delta} \ln \frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}} \end{aligned}$$

and

$$\begin{aligned} \Phi_{P_1'} \left(k, \frac{r}{\Delta} \right) & = (1 - \epsilon_1) e^{i \frac{r}{\Delta} \ln \frac{1+\epsilon_0-\epsilon_1}{1-\epsilon_0-\epsilon_1}} \\ & \quad + (-1)^k \epsilon_1 e^{i \frac{r}{\Delta} \ln \frac{1-\epsilon_0+\epsilon_1}{1-\epsilon_0-\epsilon_1}}. \end{aligned}$$

By Taylor's expansion, for $k = 0$, (26) becomes

$$\begin{aligned} & 2 \left(\frac{\Phi_{P_0'} \left(0, \frac{r}{\Delta} \right) - \Phi_{P_1'} \left(0, \frac{r}{\Delta} \right)}{2} \right)^\Delta \\ & = 2 \left(i \left(\frac{1-\epsilon_0-\epsilon_1}{2} \right) \left(\frac{r}{\Delta} \right) \ln \left(\frac{1-\epsilon_0+\epsilon_1}{1+\epsilon_0-\epsilon_1} \right) \right. \\ & \quad \left. + \mathcal{O} \left(\left(\frac{r}{\Delta} \right)^2 \right) \right)^\Delta \end{aligned}$$

which converges to zero with convergence rate $\mathcal{O}(\mathcal{O}(\Delta)^{-\Delta})$. For $k = 1$, we have

$$\begin{aligned} & 2 \left(\frac{\Phi_{P_0'} \left(1, \frac{r}{\Delta} \right) - \Phi_{P_1'} \left(1, \frac{r}{\Delta} \right)}{2} \right)^\Delta \\ & = 2 \left((\epsilon_1 - \epsilon_0) + \frac{i}{2} \left(\frac{r}{\Delta} \right) \left((1 - \epsilon_0 + \epsilon_1) \ln \frac{1 - \epsilon_0 + \epsilon_1}{1 - \epsilon_0 - \epsilon_1} \right. \right. \\ & \quad \left. \left. - (1 + \epsilon_0 - \epsilon_1) \ln \frac{1 + \epsilon_0 - \epsilon_1}{1 - \epsilon_0 - \epsilon_1} \right) + \mathcal{O} \left(\left(\frac{r}{\Delta} \right)^2 \right) \right)^\Delta \end{aligned}$$

which converges to zero with convergence rate $\mathcal{O}(\text{const}^\Delta)$, where const satisfies $|\epsilon_1 - \epsilon_0| < \text{const} < 1$. Since the convergence rate is determined by the slower of the above two, we

have proven that (26) converges to zero with rate $\mathcal{O}(\text{const}^\Delta)$ for some $\text{const} < 1$.

Consider (27). Since we assume that the input is not perfect, we have $\max(\epsilon_0, \epsilon_1) > 0$. For $k = 0$, by Taylor's expansion, we have

$$\begin{aligned} & \left(\frac{\Phi_{P'_0} \left(0, \frac{r}{\Delta} \right) + \Phi_{P'_1} \left(0, \frac{r}{\Delta} \right)}{2} \right)^\Delta \\ &= \left(1 + \frac{i}{2} \left(\frac{r}{\Delta} \right) \left((1 - \epsilon_0 + \epsilon_1) \ln \frac{1 - \epsilon_0 + \epsilon_1}{1 - \epsilon_0 - \epsilon_1} \right. \right. \\ & \quad \left. \left. + (1 + \epsilon_0 - \epsilon_1) \ln \frac{1 + \epsilon_0 - \epsilon_1}{1 - \epsilon_0 - \epsilon_1} \right) + \mathcal{O} \left(\left(\frac{r}{\Delta} \right)^2 \right) \right)^\Delta \end{aligned}$$

which converges to

$$e^{i \left(\frac{r}{\Delta} \right) \left((1 - \epsilon_0 + \epsilon_1) \ln \frac{1 - \epsilon_0 + \epsilon_1}{1 - \epsilon_0 - \epsilon_1} + (1 + \epsilon_0 - \epsilon_1) \ln \frac{1 + \epsilon_0 - \epsilon_1}{1 - \epsilon_0 - \epsilon_1} \right)}$$

with rate $\mathcal{O}(\Delta^{-1})$. For $k = 1$, we have

$$\begin{aligned} & \left(\frac{\Phi_{P'_0} \left(1, \frac{r}{\Delta} \right) + \Phi_{P'_1} \left(1, \frac{r}{\Delta} \right)}{2} \right)^\Delta \\ &= \left((1 - \epsilon_0 - \epsilon_1) \left(\frac{e^{i \frac{r}{\Delta} \ln \frac{1 - \epsilon_0 + \epsilon_1}{1 - \epsilon_0 - \epsilon_1}} + e^{i \frac{r}{\Delta} \ln \frac{1 + \epsilon_0 - \epsilon_1}{1 - \epsilon_0 - \epsilon_1}}}{2} \right) \right)^\Delta \end{aligned}$$

which converges to zero with rate $\mathcal{O}((1 - \epsilon_0 - \epsilon_1)^\Delta)$. Since the overall convergence rate is the slower of the above two, we have proven that the convergence rate is $\mathcal{O}(\Delta^{-1})$.

REFERENCES

- [1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo codes," *IEEE Trans. Inf. Theory*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
- [2] R. G. Gallager, *Low-Density Parity-Check Codes*, ser. Research Monograph Series. Cambridge, MA: MIT Press, 1963.
- [3] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [4] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1710–1722, Nov. 1996.
- [5] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.
- [6] R. J. McEliece, D. J. C. MacKay, and J. F. Cheng, "Turbo decoding as an instance of Pearl's "Belief Propagation" algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb. 1998.
- [7] S. Y. Chung, G. D. Forney Jr., T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.
- [8] T. J. Richardson and R. L. Urbanke, "Efficient encoding of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 638–656, Feb. 2001.
- [9] D. A. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1723–1731, Nov. 1996.
- [10] J. W. Byers, M. Luby, and M. Mitzenmacher, "A digital fountain approach to asynchronous reliable multicast," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 8, pp. 1528–1540, Oct. 2002.
- [11] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Analysis of low-density codes and improved designs using irregular graphs," in *Proc. 30th Annu. ACM Symp. Theory of Computing*, Dallas, TX, May 1998, pp. 249–258.
- [12] —, "Efficient erasure correcting codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.
- [13] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.
- [14] J. Hou, P. H. Siegel, and L. B. Milstein, "Performance analysis and code optimization of low density parity-check codes on Rayleigh fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 924–934, May 2001.
- [15] J. Garcia-Frias, "Decoding of low-density parity check codes over finite-state binary Markov channels," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun./Jul. 2001, p. 72.
- [16] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution and code performance bound," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1636–1652, Jul. 2003.
- [17] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, "Joint message-passing decoding of LDPC codes and partial-response channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1410–1422, Jun. 2002.
- [18] J. Li, K. R. Narayanan, E. Kurtas, and C. N. Georghiades, "On the performance of high-rate TPC/SPC codes and LDPC codes over partial response channels," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 723–734, May 2002.
- [19] A. Thangaraj and S. W. McLaughlin, "Thresholds and scheduling for LDPC-coded partial response channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2307–2309, Sep. 2002.
- [20] G. Caire, D. Burshtein, and S. Shamai, "LDPC coding for interference mitigation at the transmitter," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002.
- [21] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, "Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2141–2155, Sep. 2003.
- [22] C. Di, D. Proietti, E. Í. Telatar, T. J. Richardson, and R. L. Urbanke, "Finite-length analysis of low-density parity-check codes on the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1570–1579, Jun. 2002.
- [23] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [24] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: Asymptotic distance distributions," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 887–908, Apr. 2002.
- [25] H. Jin and R. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1451–1461, Jun. 2002.
- [26] F. Lehmann, "Distance properties of irregular LDPC codes," *Proc. IEEE Int. Symp. Information Theory*, p. 85, Jun./Jul. 2003.
- [27] S. Shamai (Shitz) and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.
- [28] A. Bannatan and D. Burshtein, "Iterative decoding of LDPC codes over arbitrary discrete-memoryless channels," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1416–1425.
- [29] A. Bannatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 417–438, Mar. 2004.
- [30] C. C. Wang, S. R. Kulkarni, and H. V. Poor, "On finite-dimensional bounds for LDPC-like codes with iterative decoding," in *Proc. Int. Symp. Information Theory and Its Applications*, Parma, Italy, Oct. 2004.
- [31] A. Khandekar, "Graph-based codes and iterative decoding," Ph.D. dissertation, Calif. Inst. Technol., Pasadena, CA, 2002.
- [32] A. Orlitsky, K. Viswanathan, and J. Zhang, "Stopping set distribution of LDPC code ensembles," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 123.
- [33] E. E. Majani and H. Rumsey Jr., "Two results on binary-input discrete memoryless channels," in *Proc. IEEE Int. Symp. Information Theory*, Budapest, Hungary, Jun. 1991, p. 104.
- [34] N. Shulman and M. Feder, "The uniform distribution as a universal prior," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1356–1362, Jun. 2004.
- [35] R. J. McEliece, "Are turbo-like codes effective on nonstandard channels?," *IEEE Information Theory Soc. Newslett.*, vol. 51, no. 4, Dec. 2001.
- [36] M. Yang and W. E. Ryan, "Lowering the error-rate floors of moderate-length high-rate irregular LDPC codes," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 237.
- [37] E. Ratzner and D. MacKay, "Sparse low-density parity-check codes for channels with cross-talk," in *Proc. IEEE Information Theory Workshop*, Paris, France, Mar./Apr., 2003, pp. 127–130.
- [38] H. Jin and T. J. Richardson, "Fast density evolution," in *Proc. 38th Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 2004.
- [39] G. Miller and G. Cohen, "The rate of regular LDPC codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2989–2992, Nov. 2003.
- [40] V. Kumar, O. Milenkovic, and K. Prakash, "On graphical representations of algebraic codes suitable for iterative decoding," in *Proc. 39th Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 2005.
- [41] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1115–1131, Jun. 2004.