Bounds on the Performance of Belief Propagation Decoding

David Burshtein, *Senior Member, IEEE*, and Gadi Miller

***Abstract*—We consider Gallager's soft-decoding (belief propagation) algorithm for decoding low-density parity-check (LDPC) codes, when applied to an arbitrary binary-input symmetric-output channel. By considering the expected values of the messages, we derive both lower and upper bounds on the performance of the algorithm. We also derive various properties of the decoding algorithm, such as a certain robustness to the details of the channel noise. Our results apply both to regular and irregular LDPC codes.**

***Index Terms*—Belief propagation, iterative decoding, low-density parity-check (LDPC) codes, sum product algorithm.**

I. Introduction

LOW-density parity-check (LDPC) codes were introduced by Gallager [5] in 1963, but were relatively ignored for more than 30 years. Recently, following the introduction of turbo codes by Berrou *et al.* [2], LDPC codes have attracted a great deal of interest. For various channels, it was demonstrated [12] that when properly designed, these codes can be used to transmit information reliably at rates which may be higher than those achievable with turbo codes. In fact, for a considerable number of examples, the maximum transmission rate at which it is possible to transmit information reliably, using these codes, is very close to channel capacity.

By considering the performance of these codes under optimal (maximum-likelihood, ML) decoding, LDPC codes have been shown to possess some very desirable properties [5], [8], [9]. In fact, it was shown [9] that for properly chosen ensembles of LDPC codes, which are based on regular bipartite graphs, these codes have an error exponent arbitrarily close to the random coding error exponent.

ML decoding of LDPC codes is in general not feasible. Instead, Gallager proposed an iterative soft-decoding algorithm, which is also called belief propagation [10]. Unfortunately, the performance of this algorithm is difficult to analyze. As an alternative, Gallager proposed analyzing a hard-decision decoding algorithm, using that as a lower bound on the performance of belief propagation. In his proof, Gallager assumed a tree-like graph structure and showed how to construct such graphs. Luby *et al.* [7] generalized this argument to random graphs. Gallager also noted that for any given channel conditions, it is possible to evaluate the performance of belief propagation by evolving the distribution of the messages. Richardson and Urbanke [11] extended this idea, and showed how to apply density evolution efficiently.

One practical obstacle encountered when using density evolution is the continuous nature of the messages. This problem may be partially overcome by quantizing the messages using a sufficiently large number of levels. A shortcoming of density evolution is that it is hard to analyze. As an alternative, for the additive white Gaussian noise channel, Chung *et al.* [4] proposed using a Gaussian approximation for the message distribution. The evolution of the infinite-dimensional density space is then reduced to the evolution of a single parameter.

In this paper, we consider Gallager's soft-decoding algorithm, when applied to an arbitrary binary-input symmetric-output channel. Similar to [4], we reduce the evolution of the infinite-dimensional space to one dimension. To this end, we use a rigorous functional evolution approach. By considering the conditional expectation of the messages given some known (e.g., the all-zero) codeword, we derive both lower and upper bounds on the performance of the algorithm. We also derive various properties of the decoding algorithm. These properties include the fact that the algorithm possesses a certain robustness to the details of the channel noise. Another result applies to LDPC codes which are based on regular graphs with large enough connectivity. In that case, we show that the decoding error probability after a finite number of iterations is bounded away from zero for a sufficiently large block size. This result is interesting since in that case the error probability of a typical code in the ensemble, when using optimal decoding, can be shown to approach the random coding error exponent [9]. Hence, as the connectivity of the graph increases, the gap in performance between belief propagation and optimal decoding increases. In this paper, we analyze both regular and irregular LDPC codes. For the case of irregular LDPC codes, our results may be used for deriving simple methods to design the distribution of edge-degrees in the graph.

The paper is organized as follows. In Section II, we provide some background information on regular and irregular LDPC codes. We also briefly describe Gallager's soft-decoding algorithm. In Sections III and IV, we analyze the soft-decoding algorithm for regular and irregular LDPC codes, respectively. Section V concludes the paper.

II. LDPC Codes and the Soft-Decoding Algorithm

Throughout the paper, we assume a binary-input ($\{0, 1\}$), symmetric-output, memoryless channel. We first describe the

Manuscript received September 6, 2000; revised June 19, 2001. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Washington, DC, June 2001.

The authors are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: burstyn@eng.tau.ac.il; gmiller@eng.tau.ac.il).

Communicated by F. R. Kschischang, Associate Editor for Coding Theory.
Publisher Item Identifier S 0018-9448(02)00006-8.

ensemble of irregular LDPC codes that we consider in this paper. The regular code ensemble is a special case of the irregular one. The irregular code ensemble is based on an ensemble of irregular bipartite graphs [6]. We first specify two probability vectors

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_c) \qquad \boldsymbol{\rho} = (\rho_1, \ldots, \rho_d).$$

λ_l is the fraction of edges with left degree l. ρ_l is the fraction of edges with right degree l. Let $\mathcal{E}$ denote the total number of edges. Then there are $\lambda_l \mathcal{E}/l$ left nodes with degree l, and $\rho_l \mathcal{E}/l$ right nodes with degree l. Let N denote the number of left nodes. Similarly, M denotes the number of right nodes. Then

$$N = \mathcal{E} \sum_{l=1}^{c} \frac{\lambda_l}{l} \qquad M = \mathcal{E} \sum_{l=1}^{d} \frac{\rho_l}{l}.$$

The $\mathcal{E}$ edges originating from left nodes are labeled from 1 to $\mathcal{E}$. The same procedure is applied to the $\mathcal{E}$ edges originating from right nodes. The ensemble of bipartite graphs is obtained by choosing a permutation π with uniform probability from the space of all permutations of $\{1, 2, \ldots, \mathcal{E}\}$. For each i, the edge labeled i on the left side is associated with the edge labeled π_i on the right side. Note that in this way multiple edges may link a pair of nodes.

The nodes on the left side are associated with the codeword bits (variable nodes) and the nodes on the right are associated with the parity-check equations (constraints or check nodes). The mapping from the bipartite graph space to the parity-check matrix space is such that an element $A_{i,j}$ in the matrix, corresponding to the ith node on the right and jth node on the left, is set to "1" if there is an odd number of edges between the two nodes, and to "0" otherwise.

The rate R' of each code in the ensemble satisfies $R' \geq R$, where

$$R \triangleq 1 - \frac{M}{N} = 1 - \frac{\sum_{l=1}^{d} \rho_l / l}{\sum_{l=1}^{c} \lambda_l / l} \tag{1}$$

(the inequality is due to a possible degeneracy in the M parity-check equations).

A special case of the irregular code ensemble that was described above is obtained when all edges have left degree c and right degree d. In that case, the ensemble is $c - d$ regular and $R = 1 - c/d$. For various channels it was demonstrated [12] that by setting $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ appropriately, the performance of irregular LDPC codes can be made superior to the performance of both regular LDPC codes and turbo codes.

We now describe the subgraph spanned from some edge e to depth l. For $l = 0$, this subgraph is e. If $l = 1$, this subgraph has two levels. Level 0 comprises e. Level 1 comprises all other edges (excluding e) originating from v, where v is the left vertex of e. If $l = 2$ there are three levels. The first two coincide with those of the depth 1 subgraph. Level 2 comprises all the edges originating from the leaf vertices of the depth 1 subgraph, excluding level 1 edges. This process may be repeated for an arbitrary l.

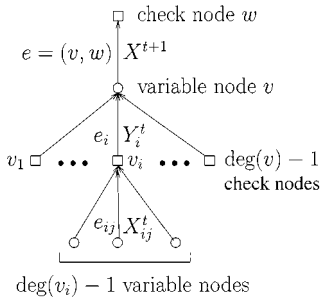

Fig. 1. A depth two tree spanned from $e = (v, w)$.

Our ensemble of bipartite graphs has the following property [5], [7], [11]. Let $\mathcal{T}_l$ denote the event that the subgraph spanned from some edge e to depth l is tree-like. For fixed values of $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$, and l we have

$$\lim_{N \to \infty} P(\mathcal{T}_l) = 1. \tag{2}$$

This property also holds for other ensembles of bipartite graphs that define LDPC codes such as Gallager's original ensemble [5] and the ensemble described in [8]. Consequently, our results in this paper hold for these ensembles as well.

Gallager's soft-decoding (belief propagation) algorithm utilizes *leftbound* and *rightbound* messages. A leftbound message is a message transmitted from a check node to a variable node. A rightbound message is a messages transmitted from a variable node to a check node.

Let s_v $v = 1, \ldots, N$ be the channel output for the vth input bit. $\eta_v = P(1|s_v)$ denotes the conditional probability that the vth transmitted bit is one given s_v. The algorithm is initialized by assigning rightbound message values to each edge in the graph, such that for an edge e with left vertex v this value is η_v. The algorithm then proceeds as described in [5].

Belief propagation is a "message-passing algorithm" in the sense that information is transmitted back and forth between variable and check nodes along the edges. The transmitted message along an edge is a function of all received messages at the node except for the said edge. This property of the algorithm ensures that the incoming messages are independent when the subgraph spanned from e is tree-like.

Consider the subgraph of depth two spanned from an edge e, with left vertex v, shown in Fig. 1. We assume that this subgraph is tree-like. Denote the edges of this tree by $e = (v, w)$, $e_i = (v, v_i)$, and $e_{i,j}$ $(i = 1, \ldots, \deg(v) - 1; j = 1, \ldots, \deg(v_i) - 1)$ for levels 0, 1, and 2 of the tree, respectively. Denote by X^{t+1} the random variable (r.v.) corresponding to the rightbound value passed in the decoding algorithm along e at time $t+1$. Similarly, denote by Y_i^t the leftbound message along the edge e_i at time t, and by $X_{i,j}^t$ the rightbound message along the edge $e_{i,j}$ at time t. Let $\eta = \eta_v$ denote the conditional probability that the transmitted bit corresponding to v was 1, given the vth-channel output. Then the following relations hold:

$$Y_i^t = \tfrac{1}{2}\left(1 - \prod_{j=1}^{d-1} (1 - 2X_{i,j}^t)\right), \qquad \text{where } d = \deg(v_i) \tag{3}$$

$$X^{t+1} = G(Y_1^t, Y_2^t, \ldots, Y_{c-1}^t, \eta), \qquad \text{where } c = \deg(v) \tag{4}$$

where

$$G(a_1, a_2, \ldots, a_n) \triangleq \frac{\prod_{i=1}^{n} a_i}{\prod_{i=1}^{n} a_i + \prod_{i=1}^{n} (1-a_i)}.$$

The reasoning behind belief propagation is as follows. Suppose that the tree assumption holds. X^t is the probability that $v = 1$ given the channel outputs corresponding to the nodes of the tree spanned by c to depth $2t$. Given this data, ML decoding may be realized by decoding v as zero if $X^t < 1/2$, and as one if $X^t > 1/2$. If $X^t = 1/2$ we choose between zero and one with equal probabilities.

It is easy to verify the following properties of G.

1) Symmetry:

$$G(a_1, a_2, \ldots, a_n) = G(a_{\pi_1}, a_{\pi_2}, \ldots, a_{\pi_n}) \tag{5}$$

where $(\pi_1, \ldots, \pi_n)$ is some permutation of $(1, \ldots, n)$.

2) Chain rule:

$$G(a_1, a_2, \ldots, a_n) = G(a_1, G(a_2, \ldots, a_n)). \tag{6}$$

We now discuss the symmetry property of the messages.

Definition 1: An r.v. (discrete, continuous, or mixed) U is said to be **symmetric** if $0 \le U \le 1$ and

$$P(U = u | U \in \{u, 1-u\}) = 1 - u, \quad \text{for all } 0 \le u \le 1 \text{ and } u \ne 1/2.$$

Suppose that the all-zero codeword has been transmitted. It was shown in [12] that the η_v's are symmetric and that the messages (both rightbound and leftbound) remain symmetric throughout the evolution of the decoding algorithm (as long as the tree assumption holds). Thus, we may assume that the $X_{i,j}^t$'s are symmetric, and hence also the Y_i^t's.

Note: In [12], symmetry is defined for log-likelihood ratio messages. For a continuous r.v. with probability density $f(u)$, symmetry may be defined by $uf(u) = (1-u)f(1-u)$. If the plain likelihood messages satisfy this condition, then their log-likelihood ratio satisfies the symmetry definition in [12] and *vice versa*. To see that, let U denote the plain likelihood message, and let V denote the log-likelihood ratio message. Then, $V = g(U)$ where $g(x) = \log[(1-x)/x]$. Now, since $f_V(v) = f_U(u)/|g'(u)|$, we have $f_V(v) = u(1-u)f_U(u)$. Thus, the symmetry definition for log-likelihood ratios $f_V(v) = e^v f_V(-v)$ implies $uf_U(u) = (1-u)f_U(1-u)$.

Definition 1 utilizes conditional expectations in order to generalize the notion of symmetry to an arbitrary r.v. For convenience we also define the following.

Definition 2: We say that an r.v. U has a **binary-symmetric** distribution with parameter $0 \le x \le 1$, $x \ne 1/2$, and denote this as $U \sim \mathrm{BS}(x)$, if U equals x with probability $1-x$, and equals $1-x$ with probability x. $U \sim \mathrm{BS}(1/2)$ denotes $U \equiv 1/2$ (i.e., $U = 1/2$ with probability 1).

Definition 3: For any symmetric r.v. U we say that $\tilde{U}$ is the **binary-symmetric r.v. corresponding to** U if $\tilde{U}$ is a binary-symmetric r.v. such that $\mathrm{E}\tilde{U} = \mathrm{E}U$.

By this definition

$$\tilde{U} \sim \mathrm{BS}\left(\frac{1}{2}\left(1 - \sqrt{1 - 2\mathrm{E}U}\right)\right). \tag{7}$$

III. Bounds on the Performance—Regular Graphs

In order to analyze the algorithm we assume throughout the paper that the all-zero codeword was transmitted. We then obtain bounds on the conditional expectation of the messages given this assumption.

A. A Lower Bound on the Performance

Recall that $\mathcal{T}_l$ denotes the event that the subgraph spanned from some edge e to depth l is tree-like. Let $l = 2t+2$ and suppose that $\mathcal{T}_l$ is satisfied. The $X_{i,j}^t$'s are then independent and identically distributed (i.i.d.) and are also independent of η. Let X^t have the same (conditional) distribution as each of the $X_{i,j}^t$'s. Then $\mathrm{E}(X_{i,j}^t | \mathcal{T}_l) = \mathrm{E}(X^t | \mathcal{T}_l)$ for all i, j. In fact, the expectation is also conditional on an all-zero transmitted codeword. However, in order to avoid complicated notation, we do not explicitly indicate this conditioning throughout the paper. Similarly, $\mathrm{E}(Y_i^t | \mathcal{T}_l) = \mathrm{E}(Y^t | \mathcal{T}_l)$, where Y^t has the same (conditional) distribution as each of the Y_i^t's. From (3) we have

$$\mathrm{E}(Y^t | \mathcal{T}_l) = \frac{1}{2}(1 - [1 - 2\mathrm{E}(X^t | \mathcal{T}_l)]^{d-1}). \tag{8}$$

We make use of the following lemma, proved in Appendix A.

Lemma 1: Let Y_i, $1 \le i \le n$ be n symmetric statistically independent r.v.'s, and let $\tilde{Y}_i$ be the binary-symmetric r.v. corresponding to Y_i. Define

$$Z = G(Y_1, Y_2, \ldots, Y_n)$$
$$Z_1 = G(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_{n-1}, Y_n)$$

and

$$Z_0 = G(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_{n-1}, \tilde{Y}_n).$$

Then $\mathrm{E}Z \le \mathrm{E}Z_1 \le \mathrm{E}Z_0$.

The lemma asserts that of all symmetric Y_i's with given expectations, the maximum of $\mathrm{E}Z$ is obtained when the Y_i's are binary-symmetric.

Lemma 1 implies the following.

Theorem 1: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the regular bipartite graph ensemble with parameters c and d. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} < (1+\epsilon) f(\mathrm{E}X^t, c, d)$$

where

$$f(x, c, d) \triangleq \sum_{i=0}^{c-1} \binom{c-1}{i} (1-q)^i q^{c-1-i} \cdot \mathrm{E}_\eta \frac{1}{1 + \frac{1-\eta}{\eta}\left(\frac{1-q}{q}\right)^{2i-(c-1)}} \tag{9}$$

and

$$q \triangleq \frac{1}{2}\Big(1-(1-2x)^{\frac{d-1}{2}}\Big). \tag{10}$$

Note: Recall that η is the conditional probability that the transmitted bit is 1 given the channel output. Our assumption that the all-zero codeword was transmitted determines the distribution of η.

Proof: Recall that $\mathcal{T}_l$ is the event that the subgraph spanned from some edge to depth l is a tree (with $l = 2t+2$). Note that

$$\mathrm{E}X^t = \mathrm{E}(X^t|\mathcal{T}_l)P(\mathcal{T}_l) + \mathrm{E}(X^t|\mathcal{T}_l^c)P(\mathcal{T}_l^c)$$

($(\cdot)^c$ denotes complementary event). Recalling (2) and using the fact that $0 \le X^t \le 1$ we thus have for any fixed t

$$\mathrm{E}X^t - \mathrm{E}(X^t|\mathcal{T}_l) \to 0, \qquad \text{as } N \to \infty. \tag{11}$$

To conclude the proof we show that

$$\mathrm{E}(X^{t+1}|\mathcal{T}_l) \le f(\mathrm{E}(X^t|\mathcal{T}_l),\, c,\, d).$$

Now, under the tree assumption η and the Y_i^t's are symmetric and statistically independent r.v.'s. Set $n = c$, $Y_i = Y_i^t$ (for $i = 1, 2, \ldots, c-1$), and $Y_c = \eta$ in Lemma 1. Using (7), $\tilde{Y}_i \sim \mathrm{BS}(q)$ where

$$q = \frac{1}{2}\Big(1-\sqrt{1-2\mathrm{E}(Y^t|\mathcal{T}_l)}\Big). \tag{12}$$

Thus, by Lemma 1, $\mathrm{E}(X^{t+1}|\mathcal{T}_l) = \mathrm{E}Z \le \mathrm{E}Z_1$. Now

$$\begin{aligned}\mathrm{E}Z_1 &= \mathrm{E}_{\tilde{Y}_1,\ldots,\tilde{Y}_{c-1},\eta}\frac{1}{1+\frac{1-\tilde{Y}_1}{\tilde{Y}_1}\cdots\frac{1-\tilde{Y}_{c-1}}{\tilde{Y}_{c-1}}\frac{1-\eta}{\eta}}\\ &= \sum_{i=0}^{c-1}\binom{c-1}{i}(1-q)^i q^{c-1-i}\\ &\quad\cdot \mathrm{E}_\eta\frac{1}{1+\frac{1-\eta}{\eta}\left(\frac{1-q}{q}\right)^{2i-(c-1)}}.\end{aligned}$$

Finally, (10) follows from (8) and (12). □

Let $E_t \triangleq \mathrm{E}X^t$. Under the all-zero codeword assumption, $E_0 = \mathrm{E}\eta$. Theorem 1 is now written as

$$E_{t+1} < (1+\epsilon)f(E_t,\, c,\, d), \qquad \forall \epsilon > 0. \tag{13}$$

Suppose that $f(x, c, d)$ satisfies $f(x, c, d) < \gamma x$, where $0 < \gamma < 1$. Then by (13) $E_t < E_0\gamma^t$. Thus, E_t approaches 0 exponentially with t. By Markov's inequality, $P(X^t \ge 1/2) \le 2E_t$. But $P(X^t \ge 1/2)$ is just the decoding error probability of the message at the tth iteration. Thus, if $\gamma < 1$ then for t and N large enough (first $N \to \infty$ then $t \to \infty$) the bit error probability can be made arbitrarily small. To show that the block error probability also approaches zero, expander graph arguments [13] may be used. In particular, it was shown in [3] that for $c > 5$ a vanishing bit error probability is a sufficient condition to ensure that for N (code block length) large enough, the algorithm (slightly modified to include appropriate clipping) successfully decodes all N bits with probability arbitrarily close to one.

Let us define

$$\Gamma(c,\, d) \triangleq \sup\left\{\frac{f(x,\, c,\, d)}{x},\, x \in (0,\, \mathrm{E}\eta]\right\}.$$

$\Gamma(c, d) < 1$ ensures that with the c and d under consideration, the decoding algorithm succeeds for large enough N with probability arbitrarily close to 1.

Unfortunately, $f(x, c, d)$ is too complex for $\Gamma(c, d)$ to be analytically solved. In order to estimate $\Gamma(c, d)$ we use the following procedure. We first consider the ratio $f(x, c, d)/x$ for $x \ll 1$ and c odd. For this case, it can be shown that $f(x, c, d)/x$ reduces to

$$\frac{f(x,\, c,\, d)}{x} = \binom{c-1}{\frac{c-1}{2}}\left(\frac{d-1}{2}\right)^{(c-1)/2} x^{(c-3)/2}\,\mathrm{E}\eta + o\Big(x^{(c-3)/2}\Big) \tag{14}$$

where $y = o(x)$ denotes $y/x \to 0$ as $x \to 0^+$ (a similar expression can be obtained when c is even). To derive (14), we consider the sum in (9). By examining the ratio of consecutive terms, we show that the middle term is the dominant term in the sum. Hence, for $c > 3$

$$\lim_{x\to 0^+}\frac{f(x,\, c,\, d)}{x} = 0.$$

Thus, if ϵ is sufficiently small then for $c > 3$

$$\Gamma(c,\, d) = \sup\left\{\frac{f(x,\, c,\, d)}{x},\, x \in (\epsilon,\, \mathrm{E}\eta]\right\}.$$

(For $c = 3$, $f(x, c, d)/x = (d-1)\mathrm{E}\eta$ as $x \to 0^+$. Hence, in this case, we require $\mathrm{E}\eta < 1/(d-1)$).

In addition to that, in Appendix B we show that $f(x, c, d)$ is monotonically increasing in x. Thus, it is sufficient to partition the interval $[\epsilon, \mathrm{E}\eta]$ into intervals $[x_i, x_{i+1}]$ of length $\hat{\epsilon}$, and verify that $f(x_{i+1}, c, d)/x_i \le \Gamma^*$, where $x_i = \epsilon + \hat{\epsilon}i$ for an upper bound Γ^* on Γ. Since $f(x, c, d)$ is continuous in x, setting $\hat{\epsilon}$ arbitrarily small yields a bound arbitrarily close to $\Gamma(c, d)$.

A channel W' is *physically degraded* with respect to some other channel W if it can be represented as a concatenation of W and some auxiliary channel Q. For example, a binary-symmetric channel (BSC) with some crossover parameter is degraded with respect to a BSC with a smaller crossover. Consider two binary-input symmetric-output channels W and W', such that W' is degraded with respect to W. Suppose that belief propagation is applied to decode both channels. Also, suppose that the tree assumption holds. Consider the decoding error probability after some number of iterations in both cases. Then by [11, Theorem 1], the decoding error probability of W cannot be larger than the error probability of W'. Essentially this is due to the fact that under the tree assumption, belief propagation is the ML decoder given the channel outputs corresponding to the nodes of the tree. Hence, if using our method above we can show that belief propagation when applied to W' satisfies $P(X^t \ge 1/2) \to 0$ as $t, N \to \infty$ (first $N \to \infty$, then $t \to \infty$), then the same applies to W.

B. A Simplified Lower Bound

The bound in Theorem 1 may be simplified to obtain the following weaker bound.

Theorem 2: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the regular

bipartite graph ensemble with parameters c and d. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} < (1+\epsilon)\hat{f}(\mathrm{E}X^t, c, d, p)$$

where

$$\hat{f}(x, c, d, p) \triangleq \sum_{i=0}^{c-1} \binom{c-1}{i} \cdot \left[\frac{(1-p)(1-q)^i q^{c-1-i}}{1+\frac{1-p}{p}\left(\frac{1-q}{q}\right)^{2i-(c-1)}} + \frac{p(1-q)^i q^{c-1-i}}{1+\frac{p}{1-p}\left(\frac{1-q}{q}\right)^{2i-(c-1)}} \right]. \quad (15)$$

$p = \frac{1}{2}(1-\sqrt{1-2\mathrm{E}\eta})$ and q is given by (10).

Proof: The proof is similar to the proof of Theorem 1, except that when applying Lemma 1 we use the assertion $\mathrm{E}Z \le \mathrm{E}Z_0$ instead of $\mathrm{E}Z \le \mathrm{E}Z_1$. □

Note that the bound in Theorem 2 depends on the channel only through $\mathrm{E}\eta$. Hence, $\mathrm{E}\eta$ may be viewed as a quality measure of the channel. This shows that the algorithm possesses a certain robustness to the details of the channel noise.

Let

$$\hat{\Gamma}(c, d, p) \triangleq \sup\left\{ \frac{\hat{f}(x, c, d, p)}{x}, x \in (0, \mathrm{E}\eta] \right\}.$$

We compute $\hat{\Gamma}(c, d, p)$ using the same method that was outlined above for $\Gamma(c, d)$. For that purpose, we utilize the monotonicity of $\hat{f}(x, c, d, p)$ in x (Appendix B). $\hat{\Gamma}(c, d, p) < 1$ ensures that with the c, d, and p under consideration, the decoding algorithm succeeds for large enough N with probability arbitrarily close to 1. Since $\hat{f}(x, c, d, p)$ is monotonically increasing in p, so is $\hat{\Gamma}(c, d, p)$.

There is another way to simplify Theorem 1. Using hard decision, we can turn the channel into a BSC, such that given the received symbol corresponding to some bit v, the output of the BSC is set to one if $\eta_v > 1/2$ and to zero otherwise. Let η' be the r.v. corresponding to the probability that the transmitted symbol was one given the output of the (combined) BSC. Denoting the crossover probability of this BSC by p' and assuming that the all-zero codeword was transmitted, $\eta' \sim \mathrm{BS}(p')$. Consider the belief propagation algorithm when applied to the BSC outputs instead of the original channel outputs. Let X'^t denote the corresponding message value at time t. By Theorem 1

$$\mathrm{E}X'^{t+1} < (1+\epsilon)\hat{f}(\mathrm{E}X'^t, c, d, p') \quad (16)$$

where $\hat{f}(\cdot)$ is defined in (15). Now, under the tree assumption, $P(X^{t+1} \ge 1/2)$ is the error probability when using optimal (ML) decoding given the original channel outputs corresponding to the nodes of the tree. $P(X'^{t+1} \ge 1/2)$ is the error probability when using optimal decoding given the (combined) BSC outputs corresponding to the nodes of the tree. Since the BSC output is determined from the original channel output, we have

$$P\left(X^{t+1} \ge \frac{1}{2}\right) \le P\left(X'^{t+1} \ge \frac{1}{2}\right) \le 2\mathrm{E}X'^{t+1}$$

(the second transition is Markov's inequality). Hence, $\mathrm{E}X'^{t+1} \to 0$ implies $P(X^{t+1} \ge \frac{1}{2}) \to 0$. However, in Appendix D we show that $p \le p'$. Hence, by the monotonicity of $\hat{f}(x, c, d, p)$ in the last argument (Appendix B)

$$\hat{f}(x, c, d, p) \le \hat{f}(x, c, d, p').$$

Hence, the first method to simplify Theorem 1, summarized in Theorem 2, provides a tighter bound compared to the second method, summarized in (16).

Example—The BSC: Let p denote the crossover parameter of the BSC. The achievable crossover probability $p^*(c, d)$ of the decoding algorithm is defined such that for any crossover probability $p < p^*(c, d)$, $\lim_{t\to\infty} P(X^t \ge 1/2) = 0$. To lower-bound $p^*(c, d)$ we define

$$\pi(c, d) \triangleq \sup\{p | \hat{\Gamma}(c, d, p) < 1\}.$$

In view of the monotonicity of $\hat{f}(x, c, d, p)$ in p (Appendix B), for any $p < \pi(c, d)$, $\hat{\Gamma}(c, d, p) < 1$. Thus, $\pi(c, d)$ is a lower bound on $p^*(c, d)$. Note that, by Theorem 2, the decoding algorithm succeeds for any binary-input symmetric-output channel with $\mathrm{E}\eta < 2\pi(c, d)(1-\pi(c, d))$.

Consider the ensemble of LDPC codes with $c = 3$ and $d = 6$. In Fig. 2, we plot $f(x, c, d)$ for a BSC with crossover parameter $p = 0.0708$. In this case, $f(x, c, d) < x$ for $x \in (0, 1/2]$. Hence, $p^*(3, 6) \ge 0.0708$. Moreover, for $p \le 0.0708$ the algorithm succeeds for any initial symmetric message distribution. Using density evolution it can be verified that $p^*(3, 6) = 0.084$ [11]. Our bound may also be compared to the bound obtained by using Gallager's hard-decoding Algorithm A, $p^*(3, 6) \ge 0.0395$ [1].

C. An Upper Bound on the Performance

In Appendix C, we prove the following Lemma, which is analogous to Lemma 1.

Lemma 2: Let Y_i, $1 \le i \le n$, be n symmetric statistically independent r.v.'s. Let $\hat{Y}_i$ be an r.v. taking the values $\frac{1}{2}$ and 0 with probabilities $2\mathrm{E}Y_i$ and $1 - 2\mathrm{E}Y_i$, respectively. Define

$$Z = G(Y_1, Y_2, \ldots, Y_n)$$

and

$$\hat{Z} = G(\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_{n-1}, Y_n).$$

Then $\mathrm{E}Z \ge \mathrm{E}\hat{Z}$.

Note that $\hat{Y}_i$ is a symmetric r.v. taking the values 0 and $1/2$. Lemma 2 implies the following.

Theorem 3: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the regular bipartite graph ensemble with parameters c and d. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} > (1-\epsilon)[1-(1-2\mathrm{E}X^t)^{d-1}]^{c-1}\mathrm{E}\eta.$$

Proof: We follow the proof of Theorem 1. Recalling (11), it is sufficient to show that

$$\mathrm{E}(X^{t+1}|\mathcal{I}_l) \ge (1-[1-2\mathrm{E}(X^t|\mathcal{I}_l)]^{d-1})^{c-1}\mathrm{E}\eta.$$

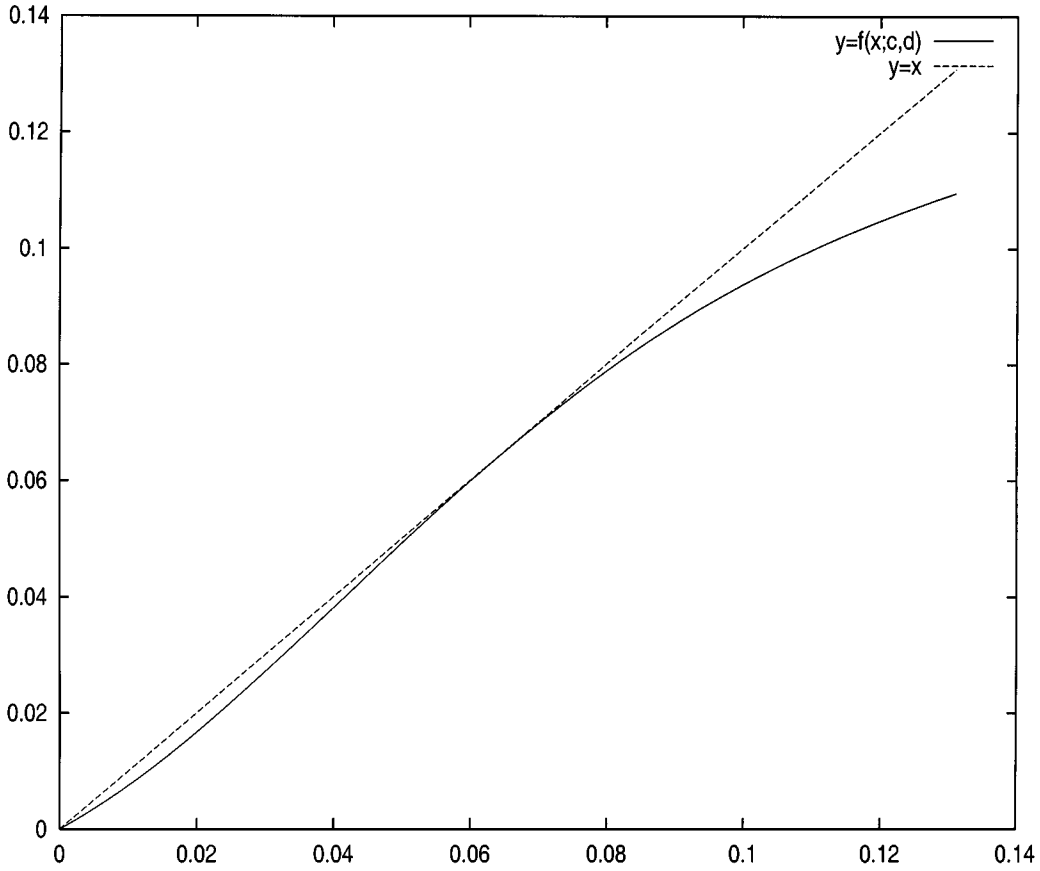

Fig. 2. A plot of $f(x, c, d)$ for $c = 3$, $d = 6$, and crossover parameter $p = 0.0708$.

Under the tree assumption, η and the Y_i^t's are symmetric and statistically independent r.v.'s. Set $n = c$, $Y_i = Y_i^t$ (for $i = 1, 2, \ldots, c-1$), and $Y_c = \eta$ in Lemma 2. By Lemma 2, $\mathrm{E}(X^{t+1}|\mathcal{I}_l) = \mathrm{E}Z \geq \mathrm{E}\hat{Z}$. Now

$$\mathrm{E}\hat{Z} = \mathrm{E}_{\hat{Y}_1, \ldots, \hat{Y}_{c-1}, \eta} \frac{\eta \prod_{i=1}^{c-1} \hat{Y}_i}{\eta \prod_{i=1}^{c-1} \hat{Y}_i + (1-\eta) \prod_{i=1}^{c-1} (1-\hat{Y}_i)}$$

$$= (2\mathrm{E}(Y^t|\mathcal{I}_l))^{c-1} \mathrm{E}\eta.$$

Using (8) we obtain the required bound. □

Theorem 3 may be used to obtain a lower bound on $\mathrm{E}X^t$. In order to obtain a lower bound on the bit error probability $P(X^t \geq 1/2)$, the following lemma may be used.

Lemma 3: Let X be a symmetric r.v. such that $EX \geq \alpha$. Then $P(X \geq 1/2) \geq 1/2(1 - \sqrt{1-2\alpha})$.

The proof of Lemma 3 is provided in Appendix E.

Now suppose that $\mathrm{E}X^t \geq \alpha$. Then by Lemma 3 and (11), for any $\epsilon > 0$ and N sufficiently large, we have

$$P\left(X^t \geq \frac{1}{2}\right) > \frac{1-\epsilon}{2}\left(1 - \sqrt{1-2\alpha}\right). \tag{17}$$

Theorem 3 and Lemma 3 imply the following.

Corollary 1: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the regular bipartite graph ensemble with parameters c and d. For any $\epsilon > 0$, any integer t, and any $R_0 < 1$ the following holds: If c and d are sufficiently large and satisfy $1 - c/d \geq R_0$, then for N sufficiently large

$$P\left(X^t \geq \frac{1}{2}\right) > \frac{1-\epsilon}{2}\left(1 - \sqrt{1-2\mathrm{E}\eta}\right).$$

Proof: Fix some $0 < \delta < 1/2$. If $\mathrm{E}X^t > \delta$ then by Theorem 3 for any $\hat{\epsilon} > 0$

$$\mathrm{E}X^{t+1} > (1-\hat{\epsilon})[1 - (1-2\delta)^{d-1}]^{c-1}\mathrm{E}\eta.$$

Now $(1-2\delta) < e^{-2\delta}$. Hence

$$\mathrm{E}X^{t+1} > (1-\hat{\epsilon})\left[1 - e^{-2(d-1)\delta}\right]^{(1-R_0)d-1}\mathrm{E}\eta.$$

Finally,

$$\left[1 - e^{-2(d-1)\delta}\right]^{(1-R_0)d-1} \to 1, \qquad \text{as } d \to \infty.$$

Hence if $\mathrm{E}X^t > \delta$ then, under the conditions of the corollary, $\mathrm{E}X^{t+1} > (1-\hat{\epsilon})\mathrm{E}\eta$. Thus, $\mathrm{E}X^t > (1-\hat{\epsilon})\mathrm{E}\eta$ for any integer t (since $\mathrm{E}X^0 = \mathrm{E}\eta > 0$). The required result follows by (17). □

Example—The BSC: Consider a BSC with a crossover parameter p. We now have $\mathrm{E}\eta = 2p(1-p)$. Thus, Corollary 1 now reads

$$P\left(X^t \geq \frac{1}{2}\right) > (1-\epsilon)p.$$

Hence, in this case, the bit error probability after t iterations approaches the uncoded bit error probability, provided that N is

sufficiently large (under the tree assumption the bit error probability is monotonically nonincreasing in t).

It follows from [5, Theorem 3.3] that a necessary condition for a capacity-achieving sequence of codes is that the average right degree approaches infinity. Hence, the average left degree also approaches infinity (to keep the rate constant). For regular codes this implies $c, d \to \infty$. Combining this with Corollary 1 we see that regular codes cannot approach capacity. This result was shown for the binary erasure channel in [6].

IV. Bounds on the Performance—Irregular Graphs

We now generalize the results of the previous section to the irregular case. Equations (3) and (4) are still valid, but now c and d are r.v.'s. Thus, in place of (8) we now have

$$\mathrm{E}(Y^t|\mathcal{I}_l) = \frac{1}{2}\left(1 - \sum_i \rho_i (1 - 2\mathrm{E}(X^t|\mathcal{I}_l))^{i-1}\right). \tag{18}$$

Message symmetry still holds [12]. From (4) and Lemma 1 we have

$$\begin{aligned}\mathrm{E}(X^{t+1}|\mathcal{I}_l) &= \sum_i \lambda_i \mathrm{E}(G(Y_1^t, Y_2^t, \ldots, Y_{i-1}^t, \eta)|\mathcal{I}_l) \\ &\le \sum_i \lambda_i \mathrm{E}G(\tilde{Y}_1^t, \tilde{Y}_2^t, \ldots, \tilde{Y}_{i-1}^t, \eta) \\ &= \sum_i \lambda_i \sum_{j=0}^{i-1} \binom{i-1}{j} (1-q)^j q^{i-1-j} \\ &\quad \cdot \mathrm{E}_\eta \frac{1}{1 + \frac{1-\eta}{\eta}\left(\frac{1-q}{q}\right)^{2j-(i-1)}}\end{aligned}$$

where $\tilde{Y}_i^t$ is the binary symmetric r.v. corresponding to Y_i^t and where q is given by

$$q = \frac{1}{2}\left(1 - \sqrt{1 - 2\mathrm{E}(Y^t|\mathcal{I}_l)}\right).$$

Recalling (18) and following the proof of Theorem 1 we have the following.

Theorem 4: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the irregular bipartite graph ensemble with parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} < (1+\epsilon) f(\mathrm{E}X^t, \boldsymbol{\lambda}, \boldsymbol{\rho})$$

where

$$\begin{aligned}f(x, \boldsymbol{\lambda}, \boldsymbol{\rho}) &\triangleq \sum_i \lambda_i \sum_{j=0}^{i-1} \binom{i-1}{j} (1-q)^j q^{i-1-j} \\ &\quad \cdot \mathrm{E}_\eta \frac{1}{1 + \frac{1-\eta}{\eta}\left(\frac{1-q}{q}\right)^{2j-(i-1)}}\end{aligned}$$

and

$$q \triangleq \frac{1}{2}\left(1 - \sqrt{\sum_i \rho_i (1-2x)^{i-1}}\right). \tag{19}$$

If $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ have only one nonzero component, $\lambda_c = \rho_d = 1$, then $f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})$ reduces to $f(x, c, d)$ introduced in Section III. Let

$$\Gamma(R) \triangleq \min_{\boldsymbol{\lambda}, \boldsymbol{\rho}} \sup\left\{\frac{f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})}{x},\ x \in (0, \mathrm{E}\eta]\right\} \tag{20}$$

where the minimum is taken over all probability vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$, subject to the rate constraint (1). $\Gamma(R)$ is monotonically increasing in R. To see that suppose that $\boldsymbol{\lambda}_1$ and $\boldsymbol{\rho}_1$ achieve the minimum for $R = R_1$ in (20). Now consider $\Gamma(R_2)$ for some $R_2 < R_1$. Set $\boldsymbol{\lambda}_2 = \boldsymbol{\lambda}_1$. It is easy to verify that there exists $\boldsymbol{\rho}_2$ such that (1) holds for R_2 and such that $q(\boldsymbol{\rho}_2) < q(\boldsymbol{\rho}_1)$ for all $0 < x < 1/2$ (by shifting weight toward lower degrees). The assertion follows by the monotonicity of f in q (which follows immediately from the regular case in Appendix B, since $f(\cdot)$ is the weighted sum of functions that are each monotonically increasing in q). Hence the value R_0 such that $\Gamma(R_0) = 1$ provides a lower bound on the achievable rate of belief propagation for the given channel. The term

$$\sup\{f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})/x,\ x \in (0, \mathrm{E}\eta]\}$$

may be estimated as in the previous section, by sampling $f(\cdot)$ at sufficiently small intervals. Note that we utilize the monotonicity of $f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})$ in x.

In order to obtain R_0 and the corresponding values of $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ in (20) we may use some general optimization method. As an alternative, linear programming may be employed using a technique similar to the one proposed in [7] in the context of Gallager's hard-decoding algorithm. Given $\boldsymbol{\rho}$ and R, we seek for a probability vector $\boldsymbol{\lambda}$ that satisfies $f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})/x < 1$ in the interval $(0, \mathrm{E}\eta]$, as well as the rate constraint (1). If such $\boldsymbol{\lambda}$ exists then $R_0 \ge R$. As in [7], we sample the interval $(0, \mathrm{E}\eta]$ and search for a feasible solution to a linear programing problem. It was found empirically that it is sufficient to consider only right degree sequences with either one nonzero component or two consecutive ones. This observation is in accordance with the results in [1] that apply to Gallager's hard-decoding algorithm.

Example—The BSC: Consider a BSC and an LDPC code with $R = 1/2$. Using the linear programming method discussed above, we designed an irregular code with $\rho_{15} = 1$ and with a maximal left degree 200. The resulting code achieves reliable communication at least for crossover parameter less than 0.092. In Fig. 3, we show $f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})$ for this code for a crossover parameter $p = 0.092$. Note that by the converse to the coding theorem, the maximal crossover for any rate–$1/2$ code is 0.110.

In a similar way to the generalization in Theorem 4, Theorems 2 and 3 may be generalized as follows.

Theorem 5: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the irregular bipartite graph ensemble with parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} < (1+\epsilon) \hat{f}(\mathrm{E}X^t, \boldsymbol{\lambda}, \boldsymbol{\rho}, p)$$

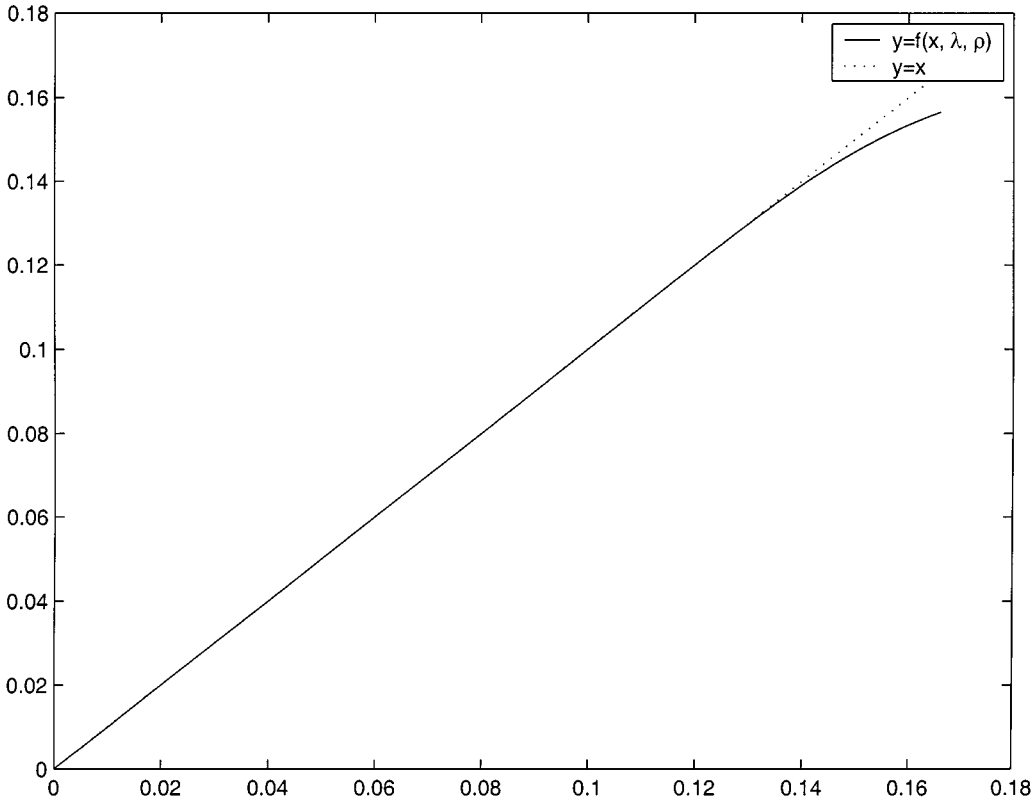

Fig. 3. A plot of $f(x, \boldsymbol{\lambda}, \boldsymbol{\rho})$ for rate–$1/2$ irregular LDPC code. The crossover parameter is $p = 0.092$.

where

$$\hat{f}(x, \boldsymbol{\lambda}, \boldsymbol{\rho}, p) \triangleq \sum_i \lambda_i \sum_{j=0}^{i-1} \binom{i-1}{j} \cdot \left[\frac{(1-p)(1-q)^j q^{i-1-j}}{1 + \frac{1-p}{p}\left(\frac{1-q}{q}\right)^{2j-(i-1)}} + \frac{p(1-q)^j q^{i-1-j}}{1 + \frac{p}{1-p}\left(\frac{1-q}{q}\right)^{2j-(i-1)}} \right].$$

$p = \frac{1}{2}(1 - \sqrt{1 - 2\mathrm{E}\eta})$ and q is given by (19).

Note again that the bound in Theorem 5 depends on the channel only through $\mathrm{E}\eta$. Hence, if the bound is effective for some channel, it will be equally effective for any other channel with the same quality measure ($\mathrm{E}\eta$).

Theorem 6: Consider a binary-input symmetric-output channel and consider the belief propagation algorithm when applied to decode an LDPC code chosen from the irregular bipartite graph ensemble with parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. For any $\epsilon > 0$, any integer t and N sufficiently large

$$\mathrm{E}X^{t+1} > (1-\epsilon)\mathrm{E}\eta \sum_i \lambda_i \left[1 - \sum_j \rho_j (1 - 2\mathrm{E}X^t)^{j-1} \right]^{i-1}.$$

V. Conclusion

We obtained bounds on the performance of Gallager's soft-decoding algorithm, and derived various properties of the algorithm. In particular, our bounds indicate that the algorithm possesses a certain robustness to the details of the channel noise. For the case of LDPC codes based on regular bipartite graphs with graph connectivity and block length sufficiently large, we showed that the decoding algorithm cannot be very effective after any fixed number of iterations (it is completely useless in the BSC case).

In order to obtain a lower bound on the performance of belief propagation, Gallager [5] proposed analyzing a hard-decoding algorithm. Our approach for lower-bounding the performance is different. It utilizes properties of the iterative procedure and results in improved bounds.

In order to derive our results we used the expected value of the messages conditioned on an all-zero transmitted codeword assumption. It is possible that our results may be improved by using the same technique with functionals other than simple expectation.

Appendix A
Proof of Lemma 1

We begin by proving the following lemma.

Lemma 4: Let Y_1 and Y_2 be two symmetric statistically independent r.v.'s and let $\tilde{Y}_1$ be the binary-symmetric r.v. corresponding to Y_1. Further define $Z = G(Y_1, Y_2)$ and $Z' = G(\tilde{Y}_1, Y_2)$. Then $\mathrm{E}Z \leq \mathrm{E}Z'$.

Proof: We introduce the notation

$$\overline{x} \triangleq x(1-x). \tag{21}$$

The following two properties hold for a symmetric r.v. X.

1) For all $0 \le x \le 1$, $x \ne 1/2$

$$P\big(X=x|\overline{X}=\overline{x}\big)=P(X=x|X\in\{x,\,1-x\})=1-x. \tag{22}$$

2) Using

$$\mathrm{E}_{X|\overline{X}}\big(X|\overline{X}=\overline{u}\big)=\sum_{x=u,\,1-u} xP\big(X=x|\overline{X}=\overline{x}\big) = \sum_{x=u,\,1-u} x(1-x)=2\overline{u}$$

we have

$$\mathrm{E}X=\mathrm{E}_{\overline{X}}\mathrm{E}_{X|\overline{X}}\big(X|\overline{X}\big)=2\mathrm{E}\overline{X}. \tag{23}$$

Now

$$\mathrm{E}Z=\mathrm{E}_{Y_1,\,Y_2}Z=\mathrm{E}_{\overline{Y_1},\,\overline{Y_2}}\mathrm{E}_{Y_1,\,Y_2|\overline{Y_1},\,\overline{Y_2}}\big(Z|\overline{Y_1},\,\overline{Y_2}\big). \tag{24}$$

Using (22) we can write

$$\begin{aligned}&\mathrm{E}_{Y_1,\,Y_2|\overline{Y_1},\,\overline{Y_2}}\big(Z|\overline{Y_1}=\overline{u_1},\,\overline{Y_2}=\overline{u_2}\big)\\ &\quad=\sum_{y_1=u_1,\,1-u_1}\ \sum_{y_2=u_2,\,1-u_2} P\big(Y_1=y_1|\overline{Y_1}=\overline{y_1}\big)\\ &\qquad\cdot P\big(Y_2=y_2|\overline{Y_2}=\overline{y_2}\big)G(y_1,\,y_2)\\ &\quad=(1-u_1)(1-u_2)G(u_1,\,u_2)+u_1(1-u_2)G(1-u_1,\,u_2)\\ &\qquad+(1-u_1)u_2G(u_1,\,1-u_2)+u_1u_2G(1-u_1,\,1-u_2)\\ &\quad=g(\overline{u_1},\,\overline{u_2})\end{aligned} \tag{25}$$

where

$$g(\overline{u_1},\,\overline{u_2})\triangleq\frac{2\overline{u_1}\,\overline{u_2}}{\overline{u_1}+\overline{u_2}-4\overline{u_1}\,\overline{u_2}}.$$

By (24) and (25)

$$\mathrm{E}Z=\mathrm{E}G(Y_1,\,Y_2)=\mathrm{E}g\big(\overline{Y_1},\,\overline{Y_2}\big). \tag{26}$$

Now, $g(x_1,\,x_2)$ is convex $\cap$ in each argument separately (when the other argument is kept fixed) in the region $0 \le x_1,\,x_2 \le 1/4$. Applying Jensen's inequality we thus obtain

$$\begin{aligned}&\mathrm{E}_{\overline{Y_1},\,\overline{Y_2}}g\big(\overline{Y_1},\,\overline{Y_2}\big)\\ &\quad=\mathrm{E}_{\overline{Y_2}}\mathrm{E}_{\overline{Y_1}|\overline{Y_2}}\big(g\big(\overline{Y_1},\,\overline{Y_2}\big)|\overline{Y_2}\big)\\ &\quad\le\mathrm{E}_{\overline{Y_2}}g\big(\mathrm{E}\big(\overline{Y_1}|\overline{Y_2}\big)\,,\,\overline{Y_2}\big)=\mathrm{E}_{\overline{Y_2}}g\big(\mathrm{E}\overline{Y_1},\,\overline{Y_2}\big)\end{aligned} \tag{27}$$

(in the last transition we used the fact that $\overline{Y_1}$ and $\overline{Y_2}$ are statistically independent). Equations (26) and (27) may be summarized as

$$\mathrm{E}Z=\mathrm{E}g\big(\overline{Y_1},\,\overline{Y_2}\big)\le\mathrm{E}_{\overline{Y_2}}g\big(\mathrm{E}\overline{Y_1},\,\overline{Y_2}\big). \tag{28}$$

Now, recalling (21) and the definition of $\tilde{Y}_1$, we have $\overline{\tilde{Y}_1}\equiv\mathrm{E}\overline{\tilde{Y}_1}$ (i.e., $\overline{\tilde{Y}_1}=\mathrm{E}\overline{\tilde{Y}_1}$ with probability 1). In addition to that $\mathrm{E}\tilde{Y}_1=\mathrm{E}Y_1$. Hence, by property (23), $\mathrm{E}\overline{\tilde{Y}_1}=\mathrm{E}\overline{Y_1}$. Thus, $\overline{\tilde{Y}_1}\equiv\mathrm{E}\overline{Y_1}$. Therefore,

$$\mathrm{E}Z'=\mathrm{E}g\Big(\overline{\tilde{Y}_1},\,\overline{Y_2}\Big)=\mathrm{E}_{\overline{Y_2}}g\big(\mathrm{E}\overline{Y_1},\,\overline{Y_2}\big) \tag{29}$$

(the first equality follows from (26) by replacing Y_1 with $\tilde{Y}_1$). The claim of the lemma follows from (28) and (29). □

We are now ready to prove Lemma 1.

Denote

$$\zeta_k\triangleq G(\tilde{Y}_1,\,\ldots,\,\tilde{Y}_k,\,Y_{k+1},\,\ldots,\,Y_n)$$

and

$$\zeta_k'\triangleq G(\tilde{Y}_1,\,\ldots,\,\tilde{Y}_{k-1},\,Y_{k+1},\,\ldots,\,Y_n).$$

Then ζ_k' is symmetric. Furthermore, by using (5) and (6) it may be verified that $\zeta_k=G(\tilde{Y}_k,\,\zeta_k')$ and $\zeta_{k-1}=G(Y_k,\,\zeta_k')$. Employing Lemma 4 with Y_k and ζ_k' in place of Y_1 and Y_2, we thus obtain $\mathrm{E}\zeta_{k-1}\le\mathrm{E}\zeta_k$ for $1\le k\le n$. Therefore, we have

$$\mathrm{E}Z=\mathrm{E}\zeta_0\le\mathrm{E}\zeta_1\le\cdots\le\mathrm{E}\zeta_{n-1}=\mathrm{E}Z_1\le\mathrm{E}\zeta_n=\mathrm{E}Z_0.\ \square$$

Appendix B
Proof of the Monotonicity of $f(x,\,c,\,d)$ and $\hat{f}(x,\,c,\,d,\,p)$

We prove that $f(x,\,c,\,d)$ is monotonically increasing in x. We also prove that $\hat{f}(x,\,c,\,d,\,p)$ is monotonically increasing both in x and in p.

It is evident from (10) that q is monotonically increasing in x. Showing that f is monotonically increasing in q will thus establish the monotonicity of $f(x,\,c,\,d)$ in x.

Define

$$h(q_1,\,q_2,\,\ldots,\,q_{c-1})\triangleq\mathrm{E}G(V_1,\,V_2,\,\ldots,\,V_{c-1},\,\eta)$$

where $V_i\sim\mathrm{BS}(q_i)$. Further define

$$U_i\triangleq G(V_1,\,\ldots,\,V_{i-2},\,V_{i-1},V_{i+1},V_{i+2},\,\ldots,\,V_{c-1},\,\eta).$$

Then, using (5), (6), and (26) we have

$$h(q_1,\,\ldots,\,q_{c-1})=\mathrm{E}G(V_i,\,U_i)=\mathrm{E}g\big(\overline{V_i},\,\overline{U_i}\big)=\mathrm{E}g\big(\overline{q_i},\,\overline{U_i}\big)\,,\quad i=1,\,2,\,\ldots,\,c-1$$

(in the last transition we used the fact that $\overline{V_i}\equiv\overline{q_i}$) Now, $g(\cdot,\,\cdot)$ is monotonically increasing in the first argument in $[0,\,1/4]$, and $\overline{x}\colon[0,\,1/2]\to[0,\,1/4]$ is monotonically increasing in x. Thus, $h(q_1,\,q_2,\,\ldots,\,q_{c-1})$ is monotonically increasing in q_i.

Finally, the proof of Theorem 1 shows that $f(x,\,c,\,d)=h(q,\,\ldots,\,q)$ (there are $c-1$ q's. q is given by (10)). The monotonicity of f in q follows immediately.

The monotonicity of $\hat{f}(x,\,c,\,d,\,p)$ both in x and in p is proved in a very similar way. In this case, we use the function

$$\hat{h}(q_1,\,\ldots,\,q_{c-1},\,p)=\mathrm{E}G(V_1,\,\ldots,\,V_{c-1},\,W)$$

where $V_i\sim\mathrm{BS}(q_i)$ and $W\sim\mathrm{BS}(p)$. □

Appendix C
Proof of Lemma 2

The proof is mostly analogous to that of Lemma 1. The main difference is that instead of using Jensen's inequality (in (27)), we now need the following lemma.

Lemma 5: Let $h(x)$ be a convex $\cap$ function on some interval $[a,\,b]$. Let X be an r.v. satisfying $a\le X\le b$, and let $\hat{X}$ be an r.v. satisfying $\hat{X}\in\{a,\,b\}$. Suppose that $\mathrm{E}X=\mathrm{E}\hat{X}$. Then

$$\mathrm{E}h(X)\ge\mathrm{E}h(\hat{X}).$$

Proof: Let $f(x)$ be a function defined by

$$x = f(x)a + (1 - f(x))b. \tag{30}$$

Since $a \leq X \leq b$, we have $0 \leq f(X) \leq 1$. Hence, we may define a binary $\{a, b\}$ r.v. Y as follows:

$$P(Y = a|X = x) = f(x) \quad P(Y = b|X = x) = 1 - f(x).$$

By (30) we have

$$\mathrm{E}Y = \mathrm{E}_X \mathrm{E}_{Y|X} Y = \mathrm{E}_X(f(X)a + (1 - f(X))b) = \mathrm{E}X = \mathrm{E}\hat{X}.$$

Y and $\hat{X}$ are both binary $\{a, b\}$ r.v.'s with the same expectation. Hence they are identically distributed. In particular, $\mathrm{E}h(\hat{X}) = \mathrm{E}h(Y)$. To conclude the proof we show that $\mathrm{E}h(X) \geq \mathrm{E}h(Y)$. Now

$$\begin{aligned}&\mathrm{E}(h(X) - h(Y))\\ &\quad = \mathrm{E}_X(h(X) - \mathrm{E}_{Y|X} h(Y))\\ &\quad = \mathrm{E}_X(h(X) - [f(X)h(a) + (1 - f(X))h(b)]).\end{aligned} \tag{31}$$

Using the convexity of $h(\cdot)$ we have

$$\begin{aligned}&f(X)h(a) + (1 - f(X))h(b)\\ &\qquad \leq h(f(X)a + [1 - f(X)]b) = h(X).\end{aligned} \tag{32}$$

The required result follows from (31) and (32). □

Next we prove the following.

Lemma 6: Let Y_1 and Y_2 be two symmetric statistically independent r.v.'s, and define an r.v. $\hat{Y}_1$ as in Lemma 2. Further define $Z = G(Y_1, Y_2)$ and $Z' = G(\hat{Y}_1, Y_2)$. Then $\mathrm{E}Z \geq \mathrm{E}Z'$.

Proof: By definition of $\hat{Y}_1$, $\mathrm{E}\hat{Y}_1 = \mathrm{E}Y_1$. Hence, by (23), $\mathrm{E}\overline{\hat{Y}_1} = \mathrm{E}\overline{Y_1}$. Thus, setting $h(x) = g(x, K)$, $a = 0$, and $b = 1/4$ in Lemma 5 (recall that $g(x, K)$ is convex $\cap$ in x) yields

$$\mathrm{E}_{\overline{Y_1}} g\left(\overline{Y_1}, K\right) \geq \mathrm{E}_{\overline{\hat{Y}_1}} g\left(\overline{\hat{Y}_1}, K\right)$$

for any $0 \leq K \leq 1/4$. Thus,

$$\begin{aligned}&\mathrm{E}_{\overline{Y_1}, \overline{Y_2}} g\left(\overline{Y_1}, \overline{Y_2}\right)\\ &\quad = \mathrm{E}_{\overline{Y_2}} \mathrm{E}_{\overline{Y_1}|\overline{Y_2}} \left(g\left(\overline{Y_1}, \overline{Y_2}\right)|\overline{Y_2}\right) = \mathrm{E}_{\overline{Y_2}} \mathrm{E}_{\overline{Y_1}} g\left(\overline{Y_1}, \overline{Y_2}\right)\\ &\quad \geq \mathrm{E}_{\overline{Y_2}} \mathrm{E}_{\overline{\hat{Y}_1}} g\left(\overline{\hat{Y}_1}, \overline{Y_2}\right) = \mathrm{E}_{\overline{Y_2}} \mathrm{E}_{\overline{\hat{Y}_1}|\overline{Y_2}} \left(g\left(\overline{\hat{Y}_1}, \overline{Y_2}\right)|\overline{Y_2}\right)\\ &\quad = \mathrm{E}_{\overline{\hat{Y}_1}, \overline{Y_2}} g\left(\overline{\hat{Y}_1}, \overline{Y_2}\right)\end{aligned} \tag{33}$$

(the second transition is due to the statistical independence of $\overline{Y_1}$ and $\overline{Y_2}$; the fourth transition is due to the statistical independence of $\overline{\hat{Y}_1}$ and $\overline{Y_2}$). From (26) and (33) we have

$$\mathrm{E}Z = \mathrm{E}g\left(\overline{Y_1}, \overline{Y_2}\right) \geq \mathrm{E}g\left(\overline{\hat{Y}_1}, \overline{Y_2}\right) = \mathrm{E}Z'. \quad \square$$

Lemma 2 is now proved, using Lemma 6, just as Lemma 1 was proved using Lemma 4. □

Appendix D
Proof That $p \leq p'$

We shall need the following auxiliary lemma.

Lemma 7: Let $f(x)$ and $g(x)$ be two nonnegative functions, and let $A \subseteq (-\infty, \infty)$. Then

$$\left(\frac{1}{\int_A f(x)\,dx} + \frac{1}{\int_A g(x)\,dx}\right) \int_A \frac{dx}{\frac{1}{f(x)} + \frac{1}{g(x)}} \leq 1.$$

Proof: Let

$$a = \int_A f(x)\,dx \qquad b = \int_A g(x)\,dx,$$

$$\hat{f}(x) = \frac{f(x)}{a} \qquad \hat{g}(x) = \frac{g(x)}{b}.$$

Also, let $\alpha = a/(a + b)$. We need to show that

$$\int_A \frac{\hat{f}(x)\hat{g}(x)\,dx}{\alpha\hat{f}(x) + (1 - \alpha)\hat{g}(x)} \leq 1.$$

It can easily be verified that the expression on the left-hand side is convex $\cup$ with respect to α (by differentiating it twice with respect to α). Hence, the maximum value of the left-hand side is obtained on the boundary of α, i.e., either when $\alpha = 0$ or when $\alpha = 1$. However, in both cases this expression is 1. □

We now show that $p \leq p'$. By the definition of p in Theorem 2, $\mathrm{E}\eta = 2p(1 - p)$. In addition to that, by definition of η', $\mathrm{E}\eta' = 2p'(1 - p')$. Hence, the claim will follow by showing that $\mathrm{E}\eta \leq \mathrm{E}\eta'$.

We now calculate $P(1|x)$, the probability that the transmitted symbol was 1 given that the channel output was x. Let $f(x) = P(x|1)$. Then $P(x|0) = f(-x)$ and

$$\begin{aligned}P(1|x) &= \frac{P(x|1) \cdot 1/2}{P(x)} = \frac{P(x|1) \cdot 1/2}{P(x|1) \cdot 1/2 + P(x|0) \cdot 1/2}\\ &= \frac{f(x)}{f(x) + f(-x)}.\end{aligned} \tag{34}$$

Under the all zero codeword assumption, and using (34)

$$\begin{aligned}\mathrm{E}\eta &= \int_{-\infty}^{\infty} f(-x)P(1|x)\,dx = \int_{-\infty}^{\infty} \frac{f(x)f(-x)\,dx}{f(x) + f(-x)}\\ &= \int_0^{\infty} \frac{2\,dx}{\frac{1}{f(x)} + \frac{1}{f(-x)}}\\ &= \int_0^{\infty} \frac{2\,dx}{\frac{1}{\min(f(x), f(-x))} + \frac{1}{\max(f(x), f(-x))}}.\end{aligned} \tag{35}$$

By the definition of p'

$$\begin{aligned}p' &= \frac{1}{2}\int_{-\infty}^{\infty} \min(f(x), f(-x))\,dx\\ &= \int_0^{\infty} \min(f(x), f(-x))\,dx.\end{aligned}$$

Using

$$\begin{aligned}&\int_0^{\infty} \min(f(x), f(-x))\,dx\\ &\qquad + \int_0^{\infty} \max(f(x), f(-x))\,dx = 1\end{aligned} \tag{36}$$

we have

$$\mathrm{E}\eta' = 2p'(1-p') = 2\left(\int_0^\infty \min(f(x),\, f(-x))\,dx\right) \cdot \left(\int_0^\infty \max(f(x),\, f(-x))\,dx\right). \quad (37)$$

Using Lemma 7, (35)–(37) imply $\mathrm{E}\eta \le \mathrm{E}\eta'$. □

Appendix E
Proof of Lemma 3

Since X is symmetric, for any $0 \le x \le 1/2$ we have

$$P(X \ge \tfrac{1}{2} \,|\, \overline{X} = \overline{x}) = P(X = 1-x | \overline{X} = \overline{x}) \ge x$$

where the inequality, rather than equality, is due to the case $x = 1/2$. Let

$$\psi(x) \triangleq \frac{1}{2}\left(1 - \sqrt{1-4x}\right).$$

Then $x = \psi(\overline{x})$, where $\overline{x}$ is defined in (21). Hence for all $0 \le \overline{X} \le 1/4$

$$P\left(X \ge \frac{1}{2} \middle| \overline{X}\right) \ge \psi(\overline{X}).$$

Therefore,

$$P\left(X \ge \frac{1}{2}\right) = \mathrm{E}_{\overline{X}} P\left(X \ge \frac{1}{2} \middle| \overline{X}\right) \ge \mathrm{E}\psi(\overline{X}). \quad (38)$$

Differentiating $\psi(x)$ twice we obtain $\psi''(x) = 2(1-4x)^{-3/2}$. Thus, for $0 \le x \le 1/4$, $\psi''(x) \ge 0$, indicating that $\psi(x)$ is convex $\cup$ on $[0,\, 1/4]$. Jensen's inequality therefore yields

$$\mathrm{E}\psi(\overline{X}) \ge \psi(\mathrm{E}\overline{X}). \quad (39)$$

Now since X is symmetric, (23) applies. This fact together with (38) and (39) imply $P(X \ge 1/2) \ge \psi(\mathrm{E}X/2)$. Finally, noting that $\psi(x)$ is monotonically increasing yields the required result. □

References

[1] L. Bazzi, T. Richardson, and R. Urbanke, "Exact thresholds and optimal codes for the binary symmetric channel and Gallager's decoding algorithm A," *IEEE Trans. Inform. Theory*, submitted for publication.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," in *Proc. 1993 IEEE Int. Conf. Communications*, Geneva, Switzerland, 1993, pp. 1064–1070.

[3] D. Burshtein and G. Miller, "Expander graph arguments for message passing algorithms," *IEEE Trans. Inform. Theory*, vol. 47, pp. 782–790, Feb. 2001.

[4] S.-Y. Chung, T. J. Richardson, and R. L. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 657–670, Feb. 2001.

[5] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

[6] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," in *Proc. 29th Annu. ACM Symp. Theory of Computing*, 1997, pp. 150–159.

[7] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 585–598, Feb. 2001.

[8] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399–431, Mar. 1999.

[9] G. Miller and D. Burshtein, "Bounds on the maximum-likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2696–2710, Nov. 2001.

[10] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann, 1988.

[11] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 599–618, Feb. 2001.

[12] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 619–637, Feb. 2001.

[13] M. Sipser and D. Spielman, "Expander codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1710–1723, Nov. 1996.