COMMUNICATIONS IN INFORMATION AND SYSTEMS
Vol. 4, No. 1, pp. 103-116, September 2004

006

FIXED POINTS AND STABILITY OF DENSITY EVOLUTION*

TOM RICHARDSON† AND RÜDIGER URBANKE‡

Dedicated to Sanjoy Mitter on the occasion of his 70th birthday.

Abstract. Density evolution is a dynamic system in a space of probability distributions representing the progress of iterative decoders in the infinite block length limit. In this paper we establish some basic results concering this process. In particular we show that the decoding threshold is equivalent to to appearance of non-trivial fixed point solutions to the density evolution equations. In the case of LDPC codes we prove the sufficiency of the previously published stability condition for stability of the δ_∞ fixed point and slightly strengthen the necessity result.

Keywords: Belief propagation, LDPC, stability, threshold

1. Introduction. It is now well-known that turbo codes and low-density parity-check (LDPC) codes perform exceptionally well under *message-passing* decoding [1, 2, 3]. *Density evolution* studies the distribution of the messages exchanged in such decoders for the limiting case of infinitely long codes [4]. So far, with the exception of the binary erasure channel (BEC), density evolution for belief propagation has largely resisted analysis and one has to be content with numerical investigations. In this paper we present some analysis of density evolution for belief propagation for (irregular) LDPC codes which reduces the gap between our understanding of the BEC and the general case.

The first result shows that the *decoding threshold* [4] of the coding system can be characterized by the existence or non-existence of non-trivial fixed-point solutions to the density evolution equations. One fixed point always exists. This is the one that we call the delta-function-at-infinity and denote by δ_∞ that corresponds to perfect information. We show that the decoding threshold of the coding system is precisely the point at which other fixed point solutions come into existence. Furthermore, all fixed points correspond to channels that are strictly better than that associated to the received distribution: The latter is always *physically degraded* with respect to that associated to the fixed point. This is in complete agreement with the BEC where fixed point probabilities of erasure are always strictly smaller than the channel probability of erasure.

In general, we would like to be able to analyze the fixed point solutions, in the hope of being able to show that irregular LDPC codes can approach channel capacity on channels other than the BEC. Our ability to analyze these fixed points, however,

*Accepted for publication on June 29, 2004.

†Tom Richardson is with Flarion Technologies, Bedminster, NJ 07921, E-mail: tjr@flarion.com

‡Rüdiger Urbanke is with EPFL, LTHC-DSC, CH-1015 Lausanne, E-mail: Rudiger.Urbanke @epfl.ch

has so far been quite limited. The fixed point δ_∞ is an exception. This brings us to the second and main result. When the evolution of the densities has reached a point where the indicated probability of error is sufficiently small, it might be expected that the density will actually converge to δ_∞ but it is not guaranteed and may in fact not occur. As we will show, there is a precise analytic condition, the stability condition, that differentiates the two possibilities. In many cases of interest the stability condition turns out to have stronger than expected implications, sometimes determining the decoding threshold.[1] Examples where this occurs include specific turbo codes for the BEC, several low complexity (non-belief-propagation) decoders applied to LDPC codes [5], and, as we show in Section 6, circuit codes under belief propagation decoding.

2. Density Evolution and the Stability Condition. Consider an irregular LDPC degree distribution pair (λ, ρ) (see [6] for definitions), together with a one parameter family of channels ordered by physical degradation. Let σ denote the channel parameter with increasing σ indicating noisier (lower capacity) channels. A typical example is BPSK signaling over the AWGNC with noise variance σ^2. Our main result shows that there exists a *stability threshold* σ^{stab} such that if $\sigma < \sigma^{\text{stab}}$, then the fixed point δ_∞ is asymptotically stable and, if $\sigma > \sigma^{\text{stab}}$, then the fixed point is unstable. If, under density evolution, the probability of error reaches a sufficiently small value and δ_∞ is stable, then the probability of error is guaranteed to converge to zero. (Convergence to zero error probability is equivalent to convergence of the message density to δ_∞.) If δ_∞ is unstable then, under the same conditions, the probability of error will be bounded away from zero and will diverge to a positive limiting value for initial message densities with sufficiently small probability of error.

Density evolution is an iterative process on message densities defined by an update equation.

$$\mathsf{P}^{\ell}(\mathsf{Q}) = \mathsf{R}\lambda(\rho(\mathsf{P}^{\ell-1}(\mathsf{Q}))) \quad \text{and} \quad \mathsf{P}^{0}(\mathsf{Q}) = \mathsf{Q},$$

where multiplication under ρ is a certain convolution operation and multiplication under λ and by R is convolution over $\mathbb{R}$, see [6]. Usually it is used to study decoding and is initialized with the message density $\mathsf{Q} = \delta_0$, indicating complete absence of information. In this paper we will often consider other initializations. We will refer to the case initialized with δ_0 as decoding initiated density evolution. Under decoding initiated density evolution the probability of error associated to the message density is monotonically decreasing so the stability threshold is an upper bound on the decoding threshold, which we denote by σ^*. We recall that the decoding threshold σ^* is defined as the largest value such that $\sigma < \sigma^*$ implies that the probability of error tends to zero under decoding initiated density evolution.

[1] It can also be used to support efficient approximations to density evolution.

2.1. Distributions and Densities. The primary objects of interest in the analysis of decoding LDPC codes are probability distributions of log-likelihood ratios, and symmetric distributions (see the next section) in particular. The space of these distributions and its basic properties were presented in [6]. We recall the basic definitions here for sake of completeness.

Let $\mathcal{F}$ denote the space of right continuous, non-decreasing functions F defined on $\mathbb{R}$ satisfying $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) \leq 1$. To each $F \in \mathcal{F}$ we associate a random variable z over $(-\infty, +\infty]$. The random variable z has *law* or *distribution* F, i.e., $\Pr\{z \in (-\infty, x]\} = F(x)$. The reason we allow $\lim_{x\to\infty} F(x) \leq 1$ rather than $\lim_{x\to\infty} F(x) = 1$ is to permit z to have some probability mass at $+\infty$, indeed $\Pr\{z = +\infty\} = 1 - \lim_{x\to\infty} F(x)$. A random variable z over $(-\infty, +\infty]$ is completely specified by its distribution $F_z \in \mathcal{F}$.

We will work with "densities" over $(-\infty, +\infty]$ which, formally, can be treated as (Radon-Nikodyn) derivatives of elements of $\mathcal{F}$. Certain densities appear often in the analysis. The density δ_0 represents the derivative of the distribution $F = 1_{x\geq 0}$ and corresponds to a channel that erases with probability 1. The density δ_∞ represents the derivative of the distribution $F = 0$ and corresponds to a channel that delivers bits perfectly.

We say that a sequence of densities $f_1, f_2, \ldots$ converges to the limit f if $F_1(x)$, $F_2(x), \ldots$ converges to $F(x)$ at all points of continuity of F. We will be most interested in convergence of densities to δ_∞ or, equivalently, convergence of distributions to 0. A sequence of densities converges to δ_∞ if and only if their associated probability of error (see below for a definition) converges to 0. Thus, stability of δ_∞ can be understood in terms of convergence of probability of error to 0.

The main properties of $\mathcal{F}$ we use in this paper are sequential compactness, that any sequence of densities has a convergent subsequence, and monotonicity under physical degradation, that any sequence of densities ordered by physical degradation converges to a limit density. For a discussion of physical degradation in this context see [6].

2.2. Symmetry. To facilitate the analysis we make the standard assumption that the all '1' codeword was transmitted (we assume BPSK ± 1 signaling), this is done without loss of generality: For channels of interest the density of a received log-likelihood value R satisfies a certain condition referred to in [6] as the *symmetry condition* $\mathsf{R}(-x) = e^{-x}\mathsf{R}(x)$. Here, and in general, messages and received values are conditional log-likelihoods on an associated bit (assumed to have been transmitted as a '1'). When the channel satisfies the symmetry condition all message densities determined by density evolution are symmetric [6]. To any symmetric density P one can associate a memoryless binary input real output symmetric channel $p(\cdot|\cdot)$ defined by $p(y|x = 1) = \mathsf{P}(y)$ and $p(y|x = -1) = \mathsf{P}(-y)$. This association plays a central role

in our analysis. If R and Q are symmetric, then $\mathsf{P}^{\ell}(\mathsf{Q})$ is symmetric for each ℓ [6].

Given two symmetric densitys Q and P we say P is physically degraded with respect to Q if the symmetric channel $p(\cdot \mid \cdot)$ determined by $p(y \mid x = 1) = \mathsf{P}(y)$ is physically degraded with respect to the symmetric channel determined by $p(y \mid x = 1) = \mathsf{Q}(y)$. Equivalently, we will say that Q is physically upgraded with respect to P.

2.3. Stability. Let f be any density (not necessarily a probability density, we extend the definition to distributions whose left limit may be positive and whose right limit may be larger than 1). Two important functionals that appear in our analysis are the following:

$$\mathcal{B}(\mathsf{f}) := \int_{-\infty}^{\infty} \mathsf{f}(x)\, \mathrm{e}^{-x/2}\, \mathrm{d}x$$
$$\mathcal{P}(\mathsf{f}) := \int_{-\infty}^{0-} \mathsf{f}(x)\, dx \; + \; \frac{1}{2} \int_{0-}^{0+} \mathsf{f}(x)\, dx\,.$$

If f is the density associated to a symmetric binary channel, then $\mathcal{B}(\mathsf{f})$ is the *Bhattacharyya constant* associated to the channel and $\mathcal{P}(\mathsf{f})$ is the hard decision bit error rate of the channel. Furthermore, we have the inequalities [6]

$$2\mathcal{P}(\mathsf{f}) \le \mathcal{B}(\mathsf{f}) \le 2\sqrt{\mathcal{P}(\mathsf{f})(1-\mathcal{P}(\mathsf{f}))}\,. \tag{1}$$

In [6] the following theorem concerning the stability condition for irregular LDPC codes under belief propagation was stated, although in somewhat different form.

THEOREM 1. *If*

$$\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) > 1,$$

then the probability of error under density evolution is bounded away from zero. More generally, there exists a constant $\gamma > 0$ such that for any symmetric $\mathsf{Q} \neq \delta_{\infty}$, we have

$$\liminf_{\ell\to\infty} \mathcal{P}(\mathsf{P}^{\ell}(\mathsf{Q})) \ge \gamma.$$

Conversely, if

$$\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1,$$

then the probability of error converges to zero once it becomes sufficiently small. More generally, there exists $\epsilon > 0$ such that if $\mathcal{P}(\mathsf{Q}) < \epsilon$ then

$$\lim_{\ell\to\infty} \mathcal{P}(\mathsf{P}^{\ell}(\mathsf{Q})) = 0.$$

In [6] the necessity of the condition (the first part of the theorem) was proved with the slightly weaker result

$$\liminf_{\ell\to\infty} \mathcal{P}(\mathsf{P}^{\ell}(\mathsf{Q})) > 0.$$

In this paper we present the first proof of the sufficiency of the condition. We now present a brief outline of this part.

The density $\mathsf{P}^\ell(\mathsf{Q})$ can be interpreted as follows. Consider a random support tree associated to (λ, ρ) (see [6], and Section 3) with variable node root and variable node leaves. The tree has $\ell + 1$ levels of variable nodes and ℓ levels of constraint nodes; we say the tree has height 2ℓ. Let the received density associated to the leaves be Q and let the received density associated to the interior nodes be R. Then $\mathsf{P}^\ell(\mathsf{Q})$ is the posterior density of the log-likelihood of the bit associated to the root node of the tree, averaged over the randomization inherent in the choice of the tree. (Usually, one is interested in $\mathsf{P}^\ell(\delta_0)$, the density associated to decoding for ℓ iterations with the channel modeled by R.)

The goal is to prove that if $\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1$, then $\lim_{\ell\to\infty} \mathcal{P}(\mathsf{P}^\ell(\mathsf{Q})) = 0$ whenever $\mathcal{P}(\mathsf{Q}) \leq \epsilon$ for some sufficiently small $\epsilon > 0$. The method of proof is to consider sub-optimal alternative decodings for the tree. A hard decision based on the root node's posterior distribution minimizes probability of error, so the probability of error associated to any other decoding provides an upper bound.

We first consider optimal sequence decoding for the tree. In this decoding we find the most likely codeword and extract the root bit. Next, maintaining the assumption that the all-1 codeword was transmitted we show that the probability of error associated to this decoding is bounded above by the probability of error associated to the decoding of a certain non-linear sub-code. The sub-code consists of the all-1 codeword and those codewords with -1 in the root that are, in a well-defined sense, minimal. The performance of the decoding of this non-linear sub-code satisfies a simple recursion in the height of the tree which, when analyzed, yields the desired result.

3. Codes and Decoding on Trees . A (λ, ρ)-random (support) tree is defined as follows. Start with one edge and choose the degree of the attached variable node randomly according to λ. This node is the *root* node of the tree. For each of the other edges emanating from the root node, choose the degree of the attached constraint node randomly according to ρ. For each of the new edges emanating from the constraint nodes choose the degree of the attached variable node randomly according to λ. If we stop at this point we have a (λ, ρ)-randomly chosen tree of height 2, see Fig. 1 for an example of height 4. The variable nodes added to the tree in the last step are the *leaves* of the tree.

Let T be a height 2ℓ tree. The *codewords* of the tree $\mathcal{C}(T)$ are the $\{+1, -1\}$ bit assignments to the variable nodes such that the parity of the bits associated to the variable nodes neighboring a constraint node is even (the bits multiplied together give $+1$.) Note that $\mathcal{C}(T)$ is a linear code: If $w, v \in \mathcal{C}(T)$ then $wv \in \mathcal{C}(T)$.

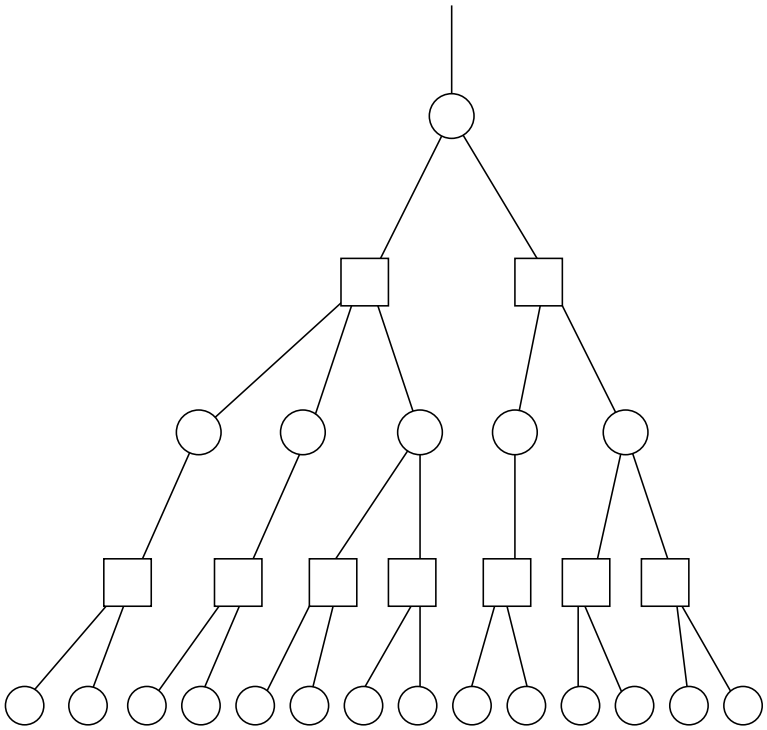

FIG. 1. *A (Tanner) graphical representation of an LDPC tree code of height 4.*

3.1. The Tree Code. Let $\mathtt{v}$ denote a variable node in T and let $w \in \mathcal{C}(T)$ be a codeword associated to T. Let $w_{\mathtt{v}} \in \{\pm 1\}$ denote the value of the bit in w associated to the variable node $\mathtt{v}$ in T. We will often refer to the value of a variable node meaning the value of the associated bit.

We shall be interested in a certain non-linear sub-code of $\mathcal{C}(T)$ which we shall call the set of *primitive codewords* of T. These codewords were identified by Wiberg [7]. A codeword $w \in \mathcal{C}(T)$ is a primitive codeword if and only if each constraint node has at most one variable node child that takes the value -1. Note that if a constraint node has one variable node child $\mathtt{v}$ with $w_{\mathtt{v}} = -1$, then the parent of that constraint node $\mathtt{v}'$ has $w_{\mathtt{v}'} = -1$. See Fig. 2 for an example. It follows that the only primitive codeword with a 1 in the root is the all-1 codeword. Let $\mathcal{C}^{-1}_{\text{prim}}(T)$ denote the set of primitive codewords with a -1 at the root.

Note that if w is a primitive codeword then the set of nodes taking the value -1 in w comprises a sub-tree of T. Let $\Psi(T)$ denote the set of *primitive sub-trees* of T defined by $t \in \Psi(T)$ if and only if $\{\mathtt{v} \in t\} = \{\mathtt{v} \in T : w_{\mathtt{v}} = -1\}$ for some primitive codeword $w \in \mathcal{C}^{-1}_{\text{prim}}(T)$.

3.2. The Tree Channel. The following construction is conceptually very useful. We refer to it as the (λ, ρ) $\ell-$tree channel. Let (λ, ρ) be a degree distribution pair.

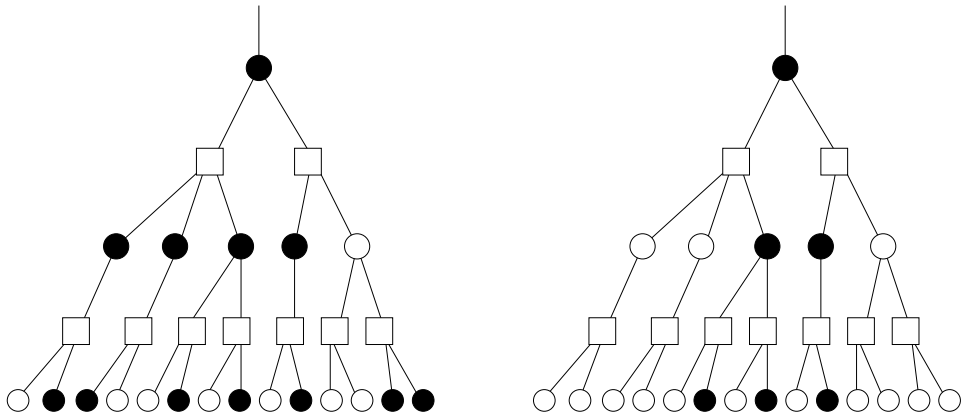

FIG. 2. *A non-primitive codeword and a primitive codeword. Note that the primitive codeword determines a primitive sub-tree.*

Consider the following binary channel. The channel takes the transmitted bit and generates a (λ, ρ)- random tree of height 2ℓ. A tree codeword is chosen uniformly at random from those codewords whose root bit matches the transmitted bit. The receiver is provided with a complete description of the tree and the observation of each bit in the codeword after passing it through some channel. Bits associated to leaf nodes (including the root if the tree has height 1) are passed through a symmetric channel with associated density Q and bits associated to internal nodes are passed through a symmetric channel with associated density R. The posterior log-likelihood of the transmitted bit can be obtained by performing belief propagation in the tree. It follows that the (log-likelihood symmetric) density associated to this channel is precisely $\mathsf{P}^{\ell}(\mathsf{Q})$.

LEMMA 1. *If, for some $k > 0$, $\mathsf{P}^k(\mathsf{Q})$ is physically degraded (upgraded) with respect to Q, then $\mathsf{P}^{\ell}(\mathsf{Q})$ converges to a fixed point F of density evolution that is physically degraded (upgraded) with respect to Q.*

Proof. Assume $\mathsf{P}^k(\mathsf{Q})$ is physically degraded (upgraded) with respect to Q. Consider the (λ, ρ) ℓ-tree channel with density R on interior nodes and density S on leaf nodes. The channel density is $\mathsf{P}^k(\mathsf{Q})$ if $\mathsf{S} = \mathsf{Q}$ and the channel density is $\mathsf{P}^{2k}(\mathsf{Q})$ if $\mathsf{S} = \mathsf{P}^k(\mathsf{Q})$.

Assume $\mathsf{P}^k(\mathsf{Q})$ is physically degraded with respect to Q and set $\mathsf{S} = \mathsf{Q}$. We can degrade the tree channel to set $\mathsf{S} = \mathsf{P}^k(\mathsf{Q})$. It follows that $\mathsf{P}^{2k}(\mathsf{Q})$ is degraded with respect to $\mathsf{P}^k(\mathsf{Q})$.

Assume $\mathsf{P}^k(\mathsf{Q})$ is physically upgraded with respect to Q and set $\mathsf{S} = \mathsf{P}^k(\mathsf{Q})$. We can degrade the tree channel to set $\mathsf{S} = \mathsf{Q}$. It follows that $\mathsf{P}^k(\mathsf{Q})$ is degraded with respect to $\mathsf{P}^{2k}(\mathsf{Q})$.

Proceeding inductively in each case we see that $\mathsf{P}^{jk}(\mathsf{Q}), j = 1, 2, ...$ is a sequence of densities monotonic with respect to physical degradation and hence converges to

a limit density, which is physically degraded (upgraded) with respect to Q. We can conclude that $\mathsf{P}^{\ell}(\mathsf{Q}), \ell = 1, 2, ...$ converges because density evolution is continuous in the space of distributions. □

3.3. Decoding the root of a tree. Consider again the (λ, ρ) tree channel. If we decode the root bit using the sum-product, or belief propagation algorithm, then, since the graph is a tree, the decoding of the root bit is actually a MAP decoding and the up-going message from the root represents the conditional distribution of the root bit conditioned on all observations associated to the tree. A hard decision for the bit based on this message has minimum probability of error over all possible decodings.

A sub-optimal alternative decoding is to decode the root bit by performing sequence decoding (find for the most likely codeword in the tree-code) and extracting the root bit. We will bound the probability of error of this decoding by relating it to the performance of the code consisting of all primitive codewords.

For each $\mathtt{v}$ in a tree T let $L_{\mathtt{v}}$ denote a random variable distributed according to the channel density associated to $\mathtt{v}$. More formally, if b denotes the bit associated to $\mathtt{v}$ and $p_{\mathtt{v}}$ denotes the associated channel, then

$$L_{\mathtt{v}} = \log \frac{p_{\mathtt{v}}(y|b=1)}{p_{\mathtt{v}}(y|b=-1)}$$

where y has density to $p_{\mathtt{v}}(y|b=1)$. (In the above scenarios this density is either R or Q.) From the receiver's perspective, the conditional log-likelihood that the root bit is 1, conditioned on all observations from the tree is given by

$$\frac{1}{2} \log \frac{\sum_{w \in \mathcal{C}^1(T)} \exp(\sum_{\mathtt{v}} w_{\mathtt{v}} L_{\mathtt{v}})}{\sum_{w \in \mathcal{C}^{-1}(T)} \exp(\sum_{\mathtt{v}} w_{\mathtt{v}} L_{\mathtt{v}})},$$

where $\mathcal{C}^{-1}(T)$ denotes the set of codewords with a -1 at the root and $\mathcal{C}^1(T)$ denotes the set of codewords with a 1 at the root. Under belief propagation, assuming log-likelihood representations, this quantity is the up-going message from the root node of T. Let $p_{\mathrm{bp}}(T)$ denote the associated probability of error.

Under maximum likelihood *sequence* decoding we find the most likely codeword $w \in \mathcal{C}(T)$, i.e., the codeword w^{seq} define by

$$w^{\mathrm{seq}} := \arg\max_{w \in \mathcal{C}(T)} \sum_{\mathtt{v} \in T} w_{\mathtt{v}} L_{\mathtt{v}},$$

and extract the root bit, i.e., decode the root as $w^{\mathrm{seq}}_{\mathrm{root}}$. If $w^{\mathrm{seq}}_{\mathrm{root}}$ is not uniquely determined, i.e., there is a tie for most likely codeword with at least one having a 1 at the root and at least one having a -1 at the root, then the decoder picks ± 1 each with probability $1/2$.

Given a tree T and a channel assignment to the nodes let $p_{\mathrm{seq}}(T)$ denote the probability of error associated to this decoding method. For any tree T we have

$$p_{\mathrm{bp}}(T) \leq p_{\mathrm{seq}}(T).$$

Let us also consider consider the maximum likelihood sequence decoder for the code consisting only of primitive codewords. Let $p_{\text{prim}}(T)$ denote the probability of error for this decoder under the assumption that the all-1 codeword was transmitted.

PROPOSITION 1. *For any tree T we have*

$$p_{seq}(T) \leq p_{prim}(T)\,.$$

Proof. Assume the all-1 codeword is transmitted. Assume $w^{\text{seq}}_{\text{root}}$ is uniquely determined as -1. An optimal codeword w^{seq} (there may be a tie) can always be written as $w^{\text{seq}} = w' \cdot w''$ where w' is a primitive codeword, $w'' \in \mathcal{C}^1(T)$ is a codeword with a 1 in the root, and

$$\{\mathtt{v} \in T : w'_{\mathtt{v}} = -1\} \cap \{\mathtt{v} \in T : {w''}_{\mathtt{v}} = -1\} = \emptyset\,.$$

Since $w^{\text{seq}}_{\text{root}} = -1$ we have $w'_{\text{root}} = -1$. Let t be the primitive sub-tree of T associated with w'. Since w^{seq} is a most likely codeword and $w_{\text{root}} = -1$ for all most likely codewords, it follows that $w' \cdot w''$ is strictly more likely than w''. This can be written as

$$\sum_{\mathtt{v}\in T} w^{\text{seq}}_{\mathtt{v}} L_{\mathtt{v}} = \sum_{\mathtt{v}\in T} w'_{\mathtt{v}} w''_{\mathtt{v}} L_{\mathtt{v}} > \sum_{\mathtt{v}\in T} w''_{\mathtt{v}} L_{\mathtt{v}}$$

which implies

$$-2\sum_{\mathtt{v}\in t} w''_{\mathtt{v}} L_{\mathtt{v}} = \sum_{\mathtt{v}\in T} (w'_{\mathtt{v}} - 1) w''_{\mathtt{v}} L_{\mathtt{v}} > 0\,.$$

For $\mathtt{v} \in t$ we have $w''_{\mathtt{v}} = 1$, hence we obtain

$$\sum_{\mathtt{v}\in t} L_{\mathtt{v}} < 0,$$

which implies

$$\sum_{\mathtt{v}\in T} w'_{\mathtt{v}} L_{\mathtt{v}} > \sum_{\mathtt{v}\in T} L_{\mathtt{v}},$$

and therefore w' is more likely than the all-1 codeword.

In the event that $w^{\text{seq}}_{\text{root}}$ is not uniquely determined, i.e., there is a tie among codewords with different root bits, then the above argument gives

$$\sum_{\mathtt{v}\in T} w'_{\mathtt{v}} L_{\mathtt{v}} \geq \sum_{\mathtt{v}\in T} L_{\mathtt{v}},$$

so, at best, there is an ambiguous tie among primitive codewords. □

The above proposition gives us the bound

$$\begin{aligned} p_{\text{seq}}(T) &\leq \Pr\left\{\sum_{\mathtt{v}\in T} L_{\mathtt{v}} \leq \max_{w\in\mathcal{C}^{-1}_{\text{prim}}(T)} \sum_{\mathtt{v}\in T} w_{\mathtt{v}} L_{\mathtt{v}}\right\} + \frac{1}{2}\Pr\left\{\sum_{\mathtt{v}\in T} L_{\mathtt{v}} = \max_{w\in\mathcal{C}^{-1}_{\text{prim}}(T)} \sum_{\mathtt{v}\in T} w_{\mathtt{v}} L_{\mathtt{v}}\right\} \\ &= \Pr\left\{\min_{t\in\Psi(T)} \sum_{\mathtt{v}\in t} L_{\mathtt{v}} < 0\right\} + \frac{1}{2}\Pr\left\{\min_{t\in\Psi(T)} \sum_{\mathtt{v}\in t} L_{\mathtt{v}} = 0\right\}, \end{aligned}$$

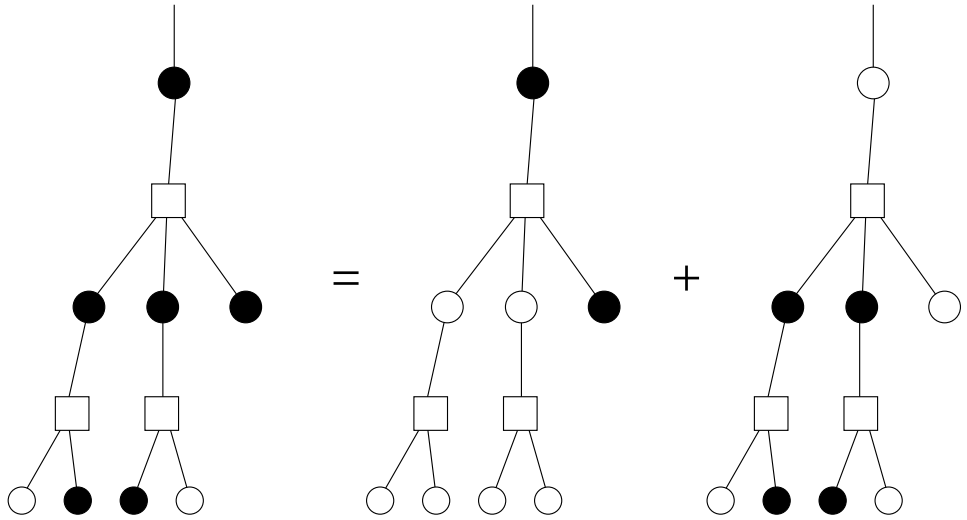

FIG. 3. *A 1-codeword decomposed as a sub-primitive codeword and a 0-codeword.*

so we now have

$$p_{\mathrm{bp}}(T) \leq \Pr\left\{\min_{t\in\Psi(T)} \sum_{\mathrm{v}\in t} L_{\mathrm{v}} < 0\right\} + \frac{1}{2}\Pr\left\{\min_{t\in\Psi(T)} \sum_{\mathrm{v}\in t} L_{\mathrm{v}} = 0\right\}. \tag{2}$$

4. Fixed points of Density Evolution. In [4] it was shown that density evolution for belief propagation always converges to a fixed point. The *decoding threshold* is defined as the (parameter determining) the worst channel below which this fixed point is δ_∞. A priori, it is possible that even below threshold other fixed point solutions to density evolution might exist but that the iterative process does not converge to them. This is in fact not possible. We will show in this section that σ is below σ^* if and only if density evolution has no fixed point solution other than δ_∞.

First, however, we will show that any fixed point is physically upgraded with respect to R.

THEOREM 2. *Let* R *be the received density for a symmetric channel and let* (λ, ρ) *be a degree distribution pair. If* Q *is a fixed point of the associated density evolution for belief propagation, then* R *is physically degraded with respect to* Q.

Proof. Consider the (λ, ρ) 1-tree channel with leaf density Q, a fixed point of density evolution, and internal density R. The symmetric density associated to this channel is Q itself, since it is a fixed point of density evolution. We can degrade this channel by erasing the information provided by the leaf nodes. For the degraded channel the associated density is R. Hence, R is physically degraded with respect to Q. □

LEMMA 2. *If* $\mathsf{Q} \neq \delta_\infty$ *is a fixed point of density evolution for* (λ, ρ) *with channel parameter* σ, *then* $\sigma \geq \sigma^*(\lambda, \rho)$.

Proof. Let Q be any symmetric density not δ_∞. Assume $\sigma < \sigma^*$, consider the associated density evolution, and let ℓ be large enough so that $\mathcal{P}(\mathsf{P}^\ell(\delta_0)) < \mathcal{P}(\mathsf{Q})$. Consider the (λ, ρ) ℓ-tree channel with leaf density Q. The probability of error associated to this channel is $\mathcal{P}(\mathsf{P}^\ell(\mathsf{Q}))$. We can degrade this channel by erasing the information at the leaf nodes. For this degraded channel the associated probability of error is $\mathcal{P}(\mathsf{P}^\ell(\delta_0))$. Since $\mathcal{P}(\mathsf{P}^\ell(\mathsf{Q})) \leq \mathcal{P}(\mathsf{P}^\ell(\delta_0)) < \mathcal{P}(\mathsf{Q})$ it follows that Q cannot be a fixed point of density evolution for (λ, ρ) with channel parameter σ. □

We remark that this theorem applies to density evolution for belief propagation decoding of turbo codes.

5. Stability. In this section we prove Theorem 1. We shall prove the second part of the theorem (sufficiency) first.

For a given tree T let X_T denote the random variable defined by

$$X_T := \min_{t \in \Psi(T)} \sum_{\mathtt{v} \in t} L_{\mathtt{v}},$$

where $L_{\mathtt{v}}$ denotes the posterior log-likelihood, i.e., a random variable distributed according to R for interior variable nodes and Q for leaf (variable) nodes. Let X^ℓ denote the average of X_T over the distribution of trees of height 2ℓ determined by (λ, ρ). Let G_ℓ denote the cumulative distribution of X^ℓ,

$$G_\ell(x) := \Pr\{X^\ell \leq x\}.$$

We will now develop a recursion relating $G_{\ell+1}$ to G_ℓ.

Let $\mathtt{v}_{\text{root}}$ denote the root node of a tree of height $2(\ell+1)$ and assume it has degree d_{root}. Let $\mathtt{c}_1, ..., \mathtt{c}_{d_{\text{root}}-1}$ denote the children of $\mathtt{v}_{\text{root}}$ and let $d_1, ..., d_{d_{\text{root}}-1}$ denote their degrees. Note that from each grandchild of $\mathtt{v}_{\text{root}}$ a (λ, ρ)-random tree of depth 2ℓ is extended.

For a node $\mathtt{v}$ in the tree let $x_{\mathtt{v}}$ denote the minimum value of the sum of log-likelihoods over primitive sub-trees rooted at $\mathtt{v}$. For a node $\mathtt{u}$ in the tree let $\mathtt{p}(\mathtt{u})$ denote the parent of $\mathtt{u}$ in the tree. It is easy to see that

$$x_{\mathtt{v}} = L_{\mathtt{v}} + \sum_{\mathtt{c}:\mathtt{p}(\mathtt{c})=\mathtt{v}} \min\{x_{\mathtt{v}'} : \mathtt{p}(\mathtt{v}') = \mathtt{c}\},$$

where $L_{\mathtt{v}}$ is the received value associated to $\mathtt{v}$. It is clear if $\mathtt{p}(\mathtt{p}(\mathtt{v}')) = \mathtt{v}_{\text{root}}$ then $x_{\mathtt{v}'}$ has cumulative distribution G_ℓ so it follows that if $\mathtt{p}(\mathtt{c}) = \mathtt{v}_{\text{root}}$ then the quantity $\min\{x_{\mathtt{v}'} : \mathtt{p}(\mathtt{v}') = \mathtt{c}\}$ has cumulative distribution $1 - (1 - G_\ell)^{d_\mathtt{c}-1}$. If we temporarily assume that $d_1 = ... = d_{d_{\text{root}}-1} = d_c$ then $x_{\mathtt{v}_{\text{root}}}$ has distribution $\mathsf{R} \otimes ((d_\mathtt{c} - 1)(1 - G_\ell)^{d_\mathtt{c}-2} g_\ell)^{\otimes(d_{\text{root}}-1)}$ where $\otimes$ denotes convolution and we have introduced the notation $g_\ell(x) := \frac{d}{dx} G_\ell(x)$. In general, where d_{root} is distributed according to λ and $d_\mathtt{c}$ is distributed according to ρ, we have,

$$g_{\ell+1} = \mathsf{R} \otimes \lambda(\rho'(1 - G_\ell) g_\ell), \quad g_0 = \mathsf{Q}, \tag{3}$$

where multiplication under λ is convolution and multiplication under ρ is point-wise.

Our aim is to show that if $\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1$ and $\mathcal{P}(\mathsf{Q})$ is sufficiently small then g_ℓ will converge to δ_∞. Let us define a new sequence of densities $\tilde{g}_\ell$ by the recursion

$$\tilde{g}_{\ell+1} = \mathsf{R} \otimes \lambda(\rho'(1)\tilde{g}_\ell), \quad \tilde{g}_0 = \mathsf{Q}.$$

Noting $\tilde{g}_0 = g_0$, it is easy to see that $\tilde{g}_\ell \geq g_\ell$ point-wise for each l by induction since λ and ρ are polynomials with positive coefficients. We remark that $\tilde{g}_\ell$ will normally not be a probability density, i.e., it will not integrate to 1.

We then have

$$\mathcal{B}(\tilde{g}_{\ell+1}) = \mathcal{B}(\mathsf{R})\, \lambda(\rho'(1)\mathcal{B}(\tilde{g}_\ell)), \ \mathcal{B}(\tilde{g}_0) = \mathcal{B}(\mathsf{Q}).$$

LEMMA 3. *Assume $\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1$. Then there exists $\eta > 0$ such that if $\mathcal{B}(\mathsf{Q}) < \eta$ then $\mathcal{B}(\tilde{g}_\ell) \to 0$ as $l \to \infty$.*

Proof. The proof proceeds by induction. The inductive hypothesis consists of the following two inequalities.

$$\text{A.} \xi_\ell := \mathcal{B}(\mathsf{R}) \left(\lambda_2 + (1-\lambda_2)\, \rho'(1)\mathcal{B}(\tilde{g}_\ell)\right) \rho'(1) < 1,$$
$$\text{B.} \rho'(1)\, \mathcal{B}(\tilde{g}_\ell) < 1.$$

Note that since $\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1$ there exists $\eta > 0$ such that $\mathcal{B}(\mathsf{R}) \left(\lambda_2 + (1-\lambda_2)\, \rho'(1)\eta\right) \rho'(1) < 1$ and $\rho'(1)\, \eta < 1$. Thus, if $\mathcal{B}(\mathsf{Q}) < \eta$ then A and B hold for $\ell = 0$. Now, assume A and B hold for some ℓ. Since B holds and $\lambda(1) = 1$ it follows that

$$\lambda(\rho'(1)\mathcal{B}(\tilde{g}_\ell)) \leq \left(\lambda_2 + (1-\lambda_2)\, \rho'(1)\, \mathcal{B}(\tilde{g}_\ell)\right) \rho'(1)\, \mathcal{B}(\tilde{g}_\ell).$$

Noting $\mathcal{B}(\tilde{g}_{\ell+1}) = \mathcal{B}(\mathsf{R})\lambda(\rho'(1)\mathcal{B}(\tilde{g}_\ell))$ we now have $\mathcal{B}(\tilde{g}_{\ell+1}) \leq \xi_\ell \mathcal{B}(\tilde{g}_\ell)$. Since $\xi_\ell < 1$ we easily conclude that A and B hold for $\ell + 1$. It follows that $\mathcal{B}(\tilde{g}_\ell) \leq (\xi_0)^\ell \mathcal{B}(\mathsf{Q})$. □

Let us assume that $\mathcal{B}(\mathsf{R})\lambda'(0)\rho'(1) < 1$. Let η be the constant from Lemma 3. By (1) we have $\mathcal{B}(\mathsf{R}) \leq 2\sqrt{\mathcal{P}(\mathsf{Q})\,(1-\mathcal{P}(\mathsf{Q}))}$. Thus, there exists $\epsilon > 0$ such that $\mathcal{P}(\mathsf{Q}) < \epsilon$ implies $\mathcal{B}(\mathsf{R}) < \eta$. It then follows from Lemma 3 that $\mathcal{B}(\tilde{g}_\ell) \to 0$ as $l \to \infty$. But since $\mathcal{B}(\tilde{g}_\ell) \geq \mathcal{B}(g_\ell) \geq \mathcal{P}(g_\ell) \geq \mathcal{P}(\mathsf{P}^\ell(\mathsf{Q}))$, we have $\mathcal{P}(\mathsf{P}^\ell(\mathsf{Q})) \to 0$ completing the sufficiency part of the Theorem 1.

The necessity part of the proof follows closely that in [6] so we will be brief. Consider the density $\eta\delta_0 + (1-\eta)\delta_\infty$ for $\eta \in (0,1)$. This is the BEC with erasure probability η. It is fairly easy to see that

$$\mathsf{P}^\ell(\eta\delta_0 + (1-\eta)\delta_\infty) = \eta(\lambda_2\rho'(1))^\ell \mathsf{R}^{\otimes\ell} + O(\eta^2),$$

where $\mathsf{R}^{\otimes\ell}$ denotes R convolved with itself (over $\mathbb{R}$) ℓ times. Let Q satisfy $\mathcal{P}(\mathsf{Q}) = \eta/2$. Then Q is physically degraded with respect to $\eta\delta_0 + (1-\eta)\delta_\infty$ [6]. For large ℓ, $\mathcal{P}(\mathsf{R}^{\otimes\ell})$ behaves like $\mathcal{B}(\mathsf{R})^\ell$, in particular one can show [8]

$$\mathcal{P}(\mathsf{R}^{\otimes\ell}) \geq \mathcal{B}(\mathsf{R})^\ell \left(\frac{1}{1+\frac{e^2\mathcal{B}(\mathsf{R})}{4(\ell+1)}}\right)\left(\frac{e}{3\pi}\sqrt{\frac{\mathcal{B}(\mathsf{R})}{\ell+1}}\right).$$

Thus,

$$\mathcal{P}(\mathsf{P}^\ell(\eta\delta_0 + (1-\eta)\delta_\infty)) \geq c\eta(\lambda_2\rho'(1)\mathcal{B}(\mathsf{R}))^\ell + O(\eta^2)$$

for some constant c independent of ℓ. Assume $\lambda_2\rho'(1)\mathcal{B}(\mathsf{R}) > 1$, then we can find k large enough so that $c(\lambda_2\rho'(1)\mathcal{B}(\mathsf{R}))^k > 1$. Thus there exists a constant $\gamma > 0$ such that if $\eta \leq \gamma$ then

$$\mathcal{P}(\mathsf{P}^k(\eta\delta_0 + (1-\eta)\delta_\infty)) \geq \eta\,.$$

It follows now that $\mathsf{P}^k(\eta\delta_0 + (1-\eta)\delta_\infty)$ is physically degraded with respect to $2\eta\delta_0 + (1-2\eta)\delta_\infty$ (which is physically degraded with respect to $\eta\delta_0 + (1-\eta)\delta_\infty$). By Lemma 1 and induction we conclude that $\mathsf{P}^\ell(\eta\delta_0 + (1-\eta)\delta_\infty)$ converges to a limit density that is physically degraded with respect to $2\gamma\delta_0 + (1-2\gamma)\delta_\infty$. Hence, for any density Q with $\mathcal{P}(\mathsf{Q}) \leq \gamma$ we have

$$\liminf_{\ell\to\infty} \mathcal{P}(\mathsf{P}^\ell(\mathsf{Q})) \geq \gamma\,,$$

and the same clearly holds for any $Q \neq \delta_\infty$. This completes the proof of Theorem 1.

6. Circuit Codes. In this section we prove that the stability condition determines the threshold for circuit codes. This was proved for the binary symmetric channel in [9] where it is also proved that the same threshold applies to maximum likelihood decoding. Circuit codes, as we consider them, are identical to regular LDPC codes with left degree two, i.e., $\lambda(x) = x$. As usual, we are interested in the asymptotic loop-free limit.

Thus, consider an LDPC degree distribution (λ, ρ) with $\lambda_2 = 1$. For any (λ, ρ) random tree T of height 2ℓ, the number of primitive codewords in $\mathcal{C}_{\text{prim}}^{-1}(T)$ equals the number of leaves. The expected number of leaves is $\rho'(1)^\ell$.

Let $M = M(T)$ denote the number of leaves in a (λ, ρ) random tree T. The probability that an element of $\mathcal{C}_{\text{prim}}^{-1}(T)$ is preferred to the all-1 codeword is less than or equal to (union bound) $M(T)$ times $\mathcal{P}(\mathsf{R}^{\otimes(\ell+1)})$. By Proposition 1 this provides a bound on the probability of error for T.

For large ℓ, $\mathcal{P}(\mathsf{R}^{\otimes(\ell+1)})$ behaves like $\mathcal{B}(\mathsf{R})^{\ell+1}$, in particular we have

$$\mathcal{P}(\mathsf{R}^{\otimes(\ell+1)}) \leq \mathcal{B}(\mathsf{R}^{\otimes(\ell+1)}) = \mathcal{B}(\mathsf{R})^{\ell+1}.$$

Now,

$$\mathrm{E}(M(T)\mathcal{B}(\mathsf{R})^{\ell+1}) = \rho'(1)^{\ell+1}\mathcal{B}(\mathsf{R})^{\ell+1}\,.$$

If $\mathcal{B}(\mathsf{R})\rho'(1) < 1$ then the probability of error converges to zero as $\ell \to \infty$. Thus, in this case, the stability threshold and the decoding threshold coincide.

7. Conclusions. The main result of the paper is the proof of Theorem 1. It has also been shown that the decoding threshold of an iterative coding system coincides with the appearance of non-trivial fixed point solutions to the density evolution equations. Both of these results are relatively easy to prove once the concept of tree channel has been developed and its performance analyzed.

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, *Near Shannon limit error-correcting coding and decoding*, in: Proceedings of ICC'93, (Geneve, Switzerland), pp. 1064–1070, May 1993.

[2] R. Gallager, *Low-density parity-check codes*, IRE Transactions on Information Theory, January 1962.

[3] D. J. C. MacKay and R. M. Neal, *Good codes based on very sparse matrices*, in: Cryptography and Coding. 5th IMA Conference (C. Boyd, ed.), no. 1025 in Lecture Notes in Computer Science, pp. 100–111, Berlin: Springer, 1995.

[4] T. Richardson and R. Urbanke, *The capacity of low-density parity check codes under message-passing decoding*, IEEE Trans. on Information Theory, 47(2001), pp. 599–619.

[5] L. Bazzi, T. Richardson, and R. Urbanke, *Exact thresholds and optimal codes for the binary symmetric channel and Gallager's decoding algorithm A*, submitted IEEE IT.

[6] T. Richardson, A. Shokrollahi, and R. Urbanke, *Design of capacity approaching low-density parity-check codes*, IEEE Trans. on Information Theory, 47(2001), pp. 599–619.

[7] N. Wiberg, *Codes and Decoding on General Graphs*. PhD thesis, Linköping University, S-581 83, Linköping, Sweden, 1996.

[8] T. Richardson and R. Urbanke, *Multi-edge type ldpc codes*, in preparation.

[9] L. Decreusefond and G. Zémor, *On the error-correcting capabilities of cycle codes of graphs*, Combinatorics, Probability and Computing, :6(1997), pp. 27–38.