The asymptotic error floor of LDPC ensembles under BP decoding

Andrea Montanari

***Abstract*— We consider communication over binary memoryless symmetric channels using random elements from irregular low density parity check code (LDPC) ensembles, and belief propagation (BP) decoding. Under the assumption that the corresponding Tanner graph includes a non-vanishing fraction of degree 2 variable nodes, we determine the large blocklength behavior of the bit error rate for noise levels below threshold. More precisely, we show that the BP bit error rate is, asymptotically, $\mathfrak{F}/n$, where n is the blocklength and $\mathfrak{F}$ an explicit constant.**

Surprisingly, this is the same behavior found for maximum likelihood (ML) decoding, implying that the BP and ML error floor are asymptotically equal.

I. Introduction and main result

Density evolution analysis of LDPC ensembles [1] implies the existence of a threshold noise level $\epsilon_{\rm BP}$ such that the bit error rate vanishes (asymptotically in the blocklength and number of iterations) for $\epsilon < \epsilon_{\rm BP}$ and remains bounded away from 0 for $\epsilon > \epsilon_{\rm BP}$. In applications one is of course limited to use a finite blocklength n. For 'reasonable' channel models, the bit error rate is then a smooth and strictly positive function of ϵ. For n large enough, two regimes can be distinguished in the performance curves: the 'error floor' at low noise, and the 'waterfall' close to threshold. Mathematically, the first one can be studied by taking $n \to \infty$ at $\epsilon < \epsilon_{\rm BP}$ fixed, while the second corresponds to the singular limit $n \to \infty$, $\epsilon \to \epsilon_{\rm BP}$. A third, extreme low noise regime (corresponding to $\epsilon \to 0$ at n fixed) has also been investigated extensively in the literature [4], [5].

This papers deals with the error floor for communication over binary memoryless symmetric channels $\mathrm{BMS}(\epsilon)$, for sparse graph code ensembles. Surprisingly little is known about this regime. Basic questions remain unanswered: How does the error probability decreases if the blocklength is doubled? Is the error floor under BP decoding much worse than the ML one?

To be definite, we focus here on the standard $\mathrm{LDPC}(n, \Lambda, P)$ ensemble [6]. Let us recall that an element from this ensemble is generated by constructing a Tanner graph on n variable nodes (of which Λ_2 of degree 2, Λ_3 of degree 3, ..., $\Lambda_{l_{\max}}$ of degree $l_{\max}$) and m check nodes (of which P_3 of degree 3, P_4 of degree 4, ..., $\Lambda_{r_{\max}}$ of degree $r_{\max}$) as follows. To each node we associate a number of sockets equal to its degree, and then match check and variable sockets according to a uniformly random permutation.

A. Montanari is with Departments of Electrical Engineering and Statistics, Stanford University, Stanford CA-94305, USA (on leave from Laboratoire de Physique Théorique de l'Ecole Normale Superieure, Paris) montanari@stanford.edu

We let $\hat{\Lambda}_l = \Lambda_l / \sum_{l'} \Lambda_{l'}$, $\hat{P}_l = P_r / \sum_{r'} P_{r'}$. Further, we denote by (λ, ρ) the edge-perspective distribution: $\lambda_l = l\Lambda_l / \sum_{l'} l'\Lambda_{l'}$, $\rho_r = rP_r / \sum_{r'} r'P_{r'}$. Finally, the corresponding generating functions are $\lambda(x) = \sum_l \lambda_l x^{l-1}$, $\rho(x) = \sum_r \rho_r x^{r-1}$.

As a performance measure, we shall consider the expected bit error rate after t BP iterations, to be denoted as $\overline{\mathrm{P}}_{\rm b}^{(t,n)}$. For finite blocklengths the $t \to \infty$ limit does not necessarily exists. We thus set the iteration number by defining

$$\overline{\mathrm{P}}_{\rm b}^{(n)} \equiv \inf_{t \geq 0} \overline{\mathrm{P}}_{\rm b}^{(t,n)} \,. \tag{1}$$

Taking the inf over the iterations number is a reasonable choice because this is done before the ensemble average. We may think of estimating the optimal $t_*(n)$ once and for all and then using it in implementations As we shall see, it turns out that $t_*(n) = \Theta(\log n)$. Below we shall compare $\overline{\mathrm{P}}_{\rm b}^{(n)}$ with the expected bit error rare under symbol MAP decoding, to be denoted as $\overline{\mathrm{P}}_{\rm b}^{(\mathrm{MAP},n)}$.

The channel $\mathrm{BMS}(\epsilon)$ can be uniquely characterized through its log-likelihood distribution [1]. We denote the corresponding random variable as Z. Di and Urbanke [3], [1] determined the asymptotic error floor under MAP decoding: $\overline{\mathrm{P}}_{\rm b}^{(\mathrm{MAP},n)} = \mathfrak{F}(\epsilon)\, n^{-1} + o(n^{-1})$. The constant $\mathfrak{F}(\epsilon)$ is given either in terms of the quantities $\mathfrak{p}_n \equiv \mathbb{P}_\epsilon\{Z^{(n)} < 0\} + \frac{1}{2}\mathbb{P}_\epsilon\{Z^{(n)} = 0\}$ ($Z^{(n)}$'s being the sum of n i.i.d. copies of Z), or of the normalized generating function $f(\omega) = \lambda'(0)\rho'(1)\mathbb{E}\exp\{i\omega Z\}$:

$$\mathfrak{F}(\epsilon) = \frac{1}{2}\sum_{n=1}^{\infty} \left(\lambda'(0)\rho'(1)\right)^n \mathfrak{p}_n \tag{2}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{f(\omega)}{1-f(\omega)}\, \widehat{\theta}(\omega)\, \mathrm{d}\omega \,, \tag{3}$$

where $\widehat{\theta}(\omega) \equiv \pi\delta(\omega) + i\mathsf{P}\frac{1}{\omega}$.

Our main result is stated below. In order to keep the statement as simple as possible, we consider a channel family $\{\mathrm{BMS}(\epsilon)\}$, ordered by physical degradation and continuous with respect to the noise level ϵ. By continuous we mean that expectations of continuous bounded functionals of the log-likelihood Z are continuous in ϵ. Finally, we assume $\mathrm{BMS}(\epsilon = 0)$ to be the noiseless channel. In fact the proof is more general but it would be hard to find interesting examples outside this class.

Theorem 1. *Consider communication over a continuous channel family $\{\mathrm{BMS}(\epsilon)\}$ using random elements from the ensemble $\mathrm{LDPC}(n, \Lambda, P)$, and BP decoding. Then there*

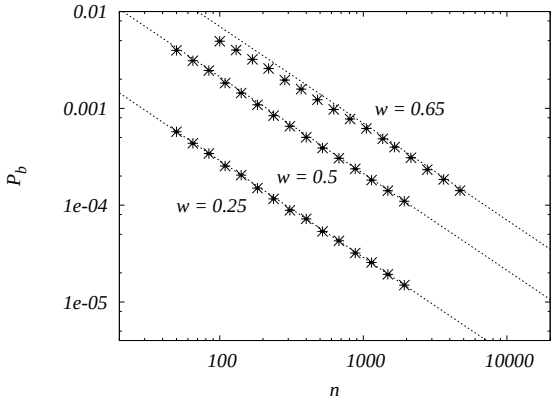

Fig. 1. Belief propagation error floor (bit error rate) for the $(2,3)$ regular ensemble ($\lambda(x) = x$, $\rho(x) = x^3$), used over the AWGN channel. Continuous lines correspond to the asymptotic prediction of Theorem 1, and symbols to numerical simulations (error bars are smaller that the symbols in most cases).

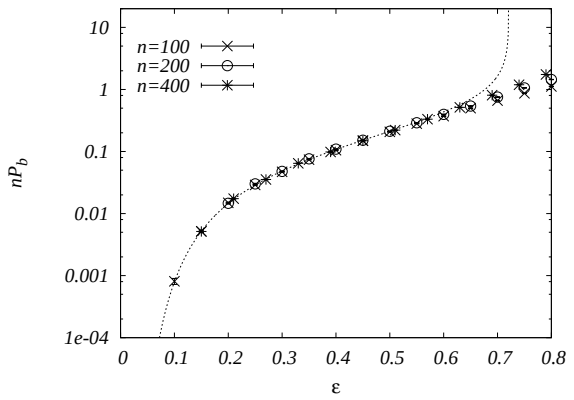

Fig. 2. More data as in figure 1, plotted as a function of the noise parameter w. Here we consider the rescaled bit error rate $n\overline{\mathrm{P}}_{\mathrm{b}}^{(n)}$, converging to the limit theoretical curve (continuous line).

exists $\epsilon_* \in (0, \epsilon_{\mathrm{BP}}]$ *such that, for any* $\epsilon < \epsilon_*$, $\overline{\mathrm{P}}_{\mathrm{b}}^{(n)} = \mathfrak{F}(\epsilon)\, n^{-1} + o(n^{-1})$.

Explicit estimates for ϵ_* can be obtained by carefully reconsidering the proofs of Propositions 1, 2 below below. We think however that these are of limited interest because of the following

Conjecture 1. *Theorem 1 holds true with* $\epsilon_* = \epsilon_{\mathrm{BP}}$.

It is immediate to see that $\mathfrak{F}(\epsilon)$ is strictly positive at any non-vanishing noise value if and only if the Tanner graph contains a finite fraction of degree 2 variable nodes (i.e. $\lambda'(0) > 0$). As a straightforward consequence, the BP and MAP error floors are asymptotically equal for large blocklengths.

Corollary 1. *Assume* $\lambda_2 > 0$, *and* $\epsilon < \epsilon_*$. *Then, for any* $\delta > 0$, $\mathrm{P}_{\mathrm{b}}^{(n)} \leq (1+\delta)\mathrm{P}_{\mathrm{b}}^{(\mathrm{MAP},n)}$ *with probability approaching* 1 *in the large blocklength limit.*

Notice that the relation holds for any given code with high probability thanks to the fact that $\mathrm{P}_{\mathrm{b}}^{(n)} \geq \mathrm{P}_{\mathrm{b}}^{(\mathrm{MAP},n)}$ by definition and noticing that $\mathrm{P}_{\mathrm{b}}^{(\mathrm{MAP},n)} = \Theta(1/n)$ with high probability (this is a consequence of the proof in [3]).

II. Numerical simulations

It is interesting to compare Theorem 1 with numerical simulations for at least two reasons. First, as our statement is only asymptotic[1], it is important to understand whether it provides a good description for finite values of n. Second, we conjectured such an asymptotic form holds for all $\epsilon < \epsilon_{\mathrm{BP}}$. Simulations may provide some support to this claim.

In Figures 1 to 4, we present the outcome of numerical simulations for two code ensembles, respectively a regular one and an irregular optimized one (see caption for the precise degree distribution). In both cases we considered communication over the additive white noise gaussian

[1]In principle upper and lower bounds on the expected bit error rate can be constructed by carefully reconsidering the estimates in Section III-A.

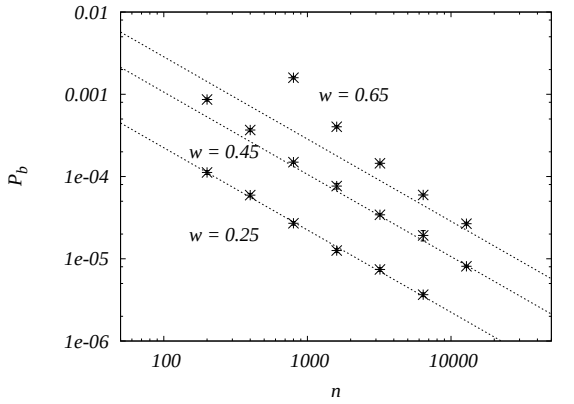

Fig. 3. As in Fig. 1 for the irregular ensemble defined by $\lambda(x) = 0.266191x + 0.256412x^2 + 0.04605470x^3 + 0.431342x^9$ and $\rho(x) = 0.65x^6 + 0.35x^7$.

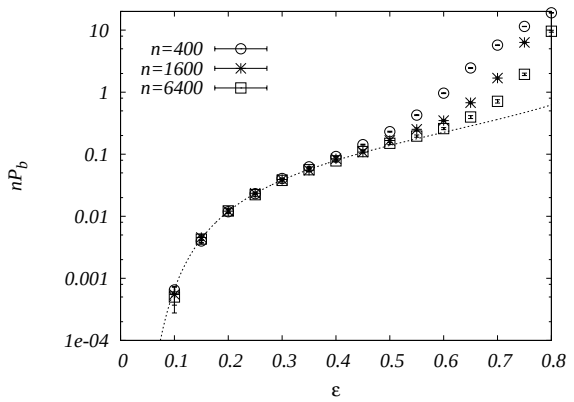

Fig. 4. As in Fig. 2 for the irregular ensemble defined by $\lambda(x) = 0.266191x + 0.256412x^2 + 0.04605470x^3 + 0.431342x^9$ and $\rho(x) = 0.65x^6 + 0.35x^7$.

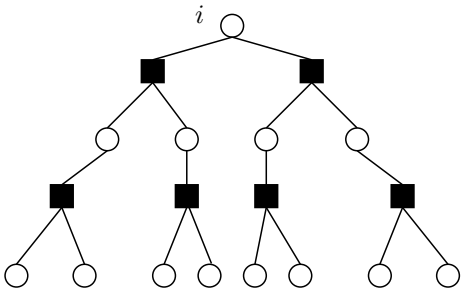

Fig. 5. An example of depth $t = 2$ neighborhood of node i in a Tanner graph from the LDPC(n, λ, ρ) ensemble. More precisely, this is an example of tree neighborhood T_2.

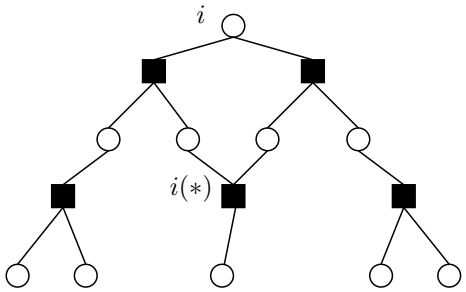

Fig. 6. Example of single-loop neighborhood $\mathsf{R}_2^{(1)}$.

(AWGN) channel with noise variance ϵ (explicitly, on input $X \in \{+1, -1\}$, the channel output is $Y = X + \sqrt{\epsilon}U$, U being a standard normal random variable). In all cases ran BP for a 50 iterations and averaged over 10^5 channel/code realizations.

For the regular ensemble, the asymptotic prediction is very accurate already at small blocklengths: already at $n = 100$ the relative error is only a few percent. For the optimized ensemble, the approach is slower. Nevertheless the agreement is reasonably good from moderate blocklengths ($n \gtrsim 1000$ or $n \gtrsim 2000$ depending on the noise level).

As expected, the agreement worsens close to the BP threshold. However, in both cases considered above, the data suggest that the asymptotic behavior in Theorem 1 is indeed correct for any $\epsilon < \epsilon_{\rm BP}$.

III. Outline of the proof

This Section contains a sketch of the proof of Theorem 1. This is based on two propositions stated in Section III-A whose proof is deferred to a longer publication. The argument for Theorem 1 is provided in Section III-B.

A. Propositions

The proof is based on a careful analysis of the local structure of the Tanner G graph associated to a random element from the LDPC(n, Λ, P) ensemble. To this end we introduce some notations. Unless specified otherwise, we shall denote by $i, j, k, \dots$ variable nodes and by $a, b, c, \dots$ check nodes. The distance $d(i, j)$ between two variable nodes is the length of the shortest path joining them on the Tanner graph (the path length being the number of check nodes encountered along the path). The distance between a variable and a check node $d(i, a)$ is defined analogously (the path length will not include counting node a: thus if x_i is involved in the a-th check, $d(i, a) = 0$).

Given an integer $t \geq 0$, we let $\mathsf{B}(i, t)$ be the subgraph of G induced by the variable nodes j such that $d(i, j) \leq t$ and by the check nodes a such that $d(i, a) \leq t-1$ (as well as all the edges joining them). Analogously $\widehat{\mathsf{B}}(i, t)$ is the subgraph induced by variable nodes with $d(i, j) \leq t$ and check nodes with $d(i, a) \leq t$. Examples are shown in Figures 5, 6. In both cases, it is understood that the degree of 'boundary' nodes is included in the description of $\mathsf{B}(i, t)$, $\widehat{\mathsf{B}}(i, t)$ (in other words, degrees are 'attached', or 'written in' the nodes).

We also use generically the symbols T_t, $\widehat{\mathsf{T}}_t$ to denote rooted tree-Tanner graphs of 'depth' t, the root being denoted as 0. For T_t, this means that all variable nodes $j \in \mathsf{T}_t$ have distance $d(0, j) \leq t$ and all check nodes $a \in \mathsf{T}_t$ distance $d(0, a) \leq t - 1$. For $\widehat{\mathsf{T}}_t$, $d(0, j), d(0, a) \leq t$.

We want to determine the asymptotic $n \to \infty$ behavior of the distribution of $\mathsf{B}(i, t)$ when G is drawn from the LDPC(Λ, P, n) ensemble. The basic underpinning of density evolution analysis is the observation that for any fixed t, $\mathsf{B}(i, t)$ is a tree with probability approaching 1 as $n \to \infty$ and that its distribution converges to

$$\mathbb{P}_0 \{\mathsf{B}(i, t) = \mathsf{T}_t\} = \hat{\Lambda}_{l_0} \prod_{j \in \mathsf{T}_t \setminus 0} \lambda_{l_j} \prod_{a \in \mathsf{T}_t} \rho_{r_a} \,, \tag{4}$$

where we denoted by l_j (respectively r_a) the degree of variable node j (check node a). In words, a tree from the distribution $\mathbb{P}_0 \{ \cdot \}$ is generated by drawing the root degree from $\hat{\Lambda}_l$, connect to l check nodes and proceed recursively. For each newly added node, the corresponding degree is drawn from λ_l (if it is a variable node) or ρ_r (if it is a check node), independently from the others.

The first ingredient in out proof is an estimate for the error probability at the root of a random code from the tree ensemble $\mathbb{P}_0\{ \cdot \}$.

Proposition 1. *Let the degree distributions (λ, ρ) be given and T_t be a random rooted tree Tanner graph drawn from the corresponding ensemble $\mathbb{P}_0\{ \cdot \}$. Let $\mathcal{C}_t$ be the low density parity check code associated to T_t and assume a codeword from the $\mathcal{C}_t$ is transmitted through the channel family $\{\mathrm{BMS}(\epsilon)\}$. Finally let $\overline{\mathrm{P}}_{\mathrm{b}}^{(t,\mathrm{tree})}$ be the expected error probability for the root bit under symbol MAP decoding. Then for any ϵ small enough there exists a positive constant $\kappa(\epsilon)$ such that*

$$\overline{\mathrm{P}}_{\mathrm{b}}^{(t,\mathrm{tree})} \leq e^{-\kappa(\epsilon) t} \,, \tag{5}$$

Further $\kappa(\epsilon) \uparrow \infty$ as $\epsilon \downarrow 0$.

Similar statements (or proofs implying this one) have been proved several times in the literature, see for instance [7].

In this paper we improve over the tree description of the neighborhood of $\mathsf{B}(i, t)$ in two ways: first we include sub-dominant contributions that yield leading order corrections to density evolution; second we establish error bounds for the new refined model. In the following we shall denote by

$\beta = (l_{\max}-1)(k_{\max}-1)$ the maximum growth rate of the neighborhood of i (more precisely $|\mathsf{B}(i,t)| \le \beta^t$)

The refined model differs in two ways with respect to the limiting one, cf. Eq. (4). First of all the probability that $\mathsf{B}(i,t) = \mathsf{T}_t$ for any particular tree T_t changes due to correlations of degrees at different nodes. This change will be denoted by $f(\mathsf{T}_t)$.

Next, with some finite probability $\mathsf{B}(i,t)$ includes one cycle (the probability that its cyclic number is larger than one being a sub-leading correction). We can distinguish two cases. We denote by $\mathsf{S}_t^{(s)}$ a unicyclic graph of radius t centered at i, such that the smallest neighborhood of i not being a tree is $\mathsf{B}(i,s)$ (with an abuse of notation we refer here to the neighborhood of i in a specific graph, namely $\mathsf{S}_t^{(s)}$). Analogously, we let $\mathsf{R}_t^{(s)}$ be a unicyclic graph of radius t centered at i, such that the smallest neighborhood of i not being a tree is $\widehat{\mathsf{B}}(i,s)$. Denote by $i(*)$ (respectively, by $a(*)$) the node on the loop farthest away from i. Then we introduce the following distributions of such classes of neighborhoods:

$$\mathbb{P}_{1,s}(\mathsf{S}_t^{(s)}) = \frac{l_{i(*)}-1}{\lambda'(1)} \binom{q(s)}{2}^{-1} \hat{\Lambda}_{l_0} \prod_{i\in\mathsf{S}_t^{(s)}} \lambda_{l_i} \prod_{a\in\mathsf{S}_t^{(s)}} \rho_{r_a}\,, \tag{6}$$

$$\widehat{\mathbb{P}}_{1,s}(\mathsf{R}_t^{(s)}) = \frac{r_{a(*)}-1}{\rho'(1)} \binom{\hat{q}(s)}{2}^{-1} \hat{\Lambda}_{l_0} \prod_{i\in\mathsf{R}_t^{(s)}} \lambda_{l_i} \prod_{a\in\mathsf{R}_t^{(s)}} \rho_{r_a}\,. \tag{7}$$

In words, the distribution $\mathbb{P}_{1,s}(\,\cdot\,)$ is described as follows: generate the first s levels of $\mathsf{S}_t^{(s)}$ according to the model $\mathbb{P}_0(\,\cdot\,)$ (i.e. drawing the root degree from $\hat{\Lambda}$. the other node degrees independently from λ (for variable nodes) and ρ (check nodes). Then pick two nodes uniformly at random among the ones at generation s and identify them. Denote the resulting node as $i(*)$. Finally, generate the last $(t-s)$ levels once again using λ, ρ i.i.d. degrees, except for the node $i(*)$ whose degree has distribution $(l-1)\lambda_l/\lambda'(1)$.

The probability that the actually neighborhood of $\mathsf{B}(i,t)$ is of the type $\mathsf{S}_t^{(s)}$ (respectively $\mathsf{R}_t^{(s)}$) will be approximated by δ_s/n ($\widehat{\delta}_s/n$), where

$$\delta_s = \frac{\lambda'\int\lambda}{2}[\rho''\lambda' + \lambda''\rho'^2]\,(\lambda'\rho')^{s-1}\frac{1-(\lambda'\rho')^s}{1-\lambda'\rho'}\,, \tag{8}$$

$$\widehat{\delta}_s = \frac{\rho'\int\rho}{2}\lambda'^2[\rho''\lambda' + \lambda''\rho'^2]\,(\lambda'\rho')^{s-1}\frac{1-(\lambda'\rho')^s}{1-\lambda'\rho'} + \frac{1}{2}\lambda''(\lambda'\rho')^s\,. \tag{9}$$

All the polynomials in this expressions are understood to be evaluated at 1. Notice that $\delta_s, \widehat{\delta}_s \le C\beta^{2t}/n$ for some constant C.

Finally, we recall that the total variation distance $||\mu-\nu||_{\rm TV}$ between two distributions on a finite set $\mathcal{X}$ is defined as $||\mu-\nu||_{\rm TV} = \frac{1}{2}\sum_{x\in\mathcal{X}}|\mu(x)-\nu(x)|$.

Proposition 2. *Consider a Tanner graph from the* $\mathrm{LDPC}(n,\Lambda,P)$ *ensemble, let* $\mathsf{B}(i,t)$ *be the radius* t*-neighborhood of a uniformly random variable node* i*, and denote by* $\mathbb{P}\{\,\cdot\,\}$ *its distribution. Assume* $\Lambda_l, P_r > n\xi$*, and* $l_{\max}, r_{\max}$ *to be given. Then there exists a function* $f(\mathsf{T}_t)$ *and positive constant* C *depending uniquely on* ξ *and* K *such that, for any* t *and* n $|f(\mathsf{T}_t)| \le C\beta^{2t}/n$*, and:*

$$\Big|\Big|\mathbb{P}\{\,\cdot\,\} - (1+f(\,\cdot\,))\mathbb{P}_0\{\,\cdot\,\} - \sum_{s=1}^{t}\delta_s\,\mathbb{P}_{1,s}\{\,\cdot\,\} - \sum_{s=1}^{t-1}\widehat{\delta}_s\,\widehat{\mathbb{P}}_{1,s}\{\,\cdot\,\}\Big|\Big|_{\rm TV} \le \frac{C\,\beta^{3t}}{n^2}\,. \tag{10}$$

The proof of this statement consists in writing down explicit formulas for the actual probabilities $\mathbb{P}\{\mathsf{B}(i,t)=\mathsf{T}_t\}$, $\mathbb{P}\left\{\mathsf{B}(i,t)=\mathsf{S}_t^{(s)}\right\}$, etc., and Taylor expanding these formulae as $n\to\infty$.

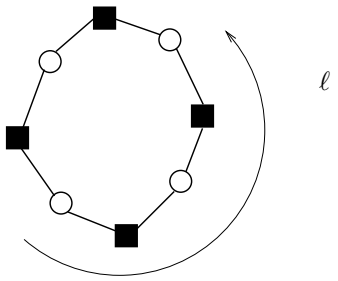

Fig. 7. An isolated loop of length ℓ is equivalent to a 'loop' or repetition code.

B. Proof of Theorem 1

First notice that, using the result of [3], and the optimality of MAP decoding it is enough to prove an upper bound of the form $\overline{\mathrm{P}}_{\rm b}^{(n)} \le \mathfrak{F}(\epsilon)/n + o(n^{-1})$.

By linearity of expectation, the average bit error rate $\overline{\mathrm{P}}_{\rm b}^{(n,t)}$ is just the expectation with respect to the graph ensemble, of the probability $\mathrm{P}_i^{(t)}(G)$ that bit i is incorrectly decoded after t iterations. This in turns depends on the graph G only through the depth t neighborhood of i, $\mathsf{B}(i,t)$. By Proposition 2 we have

$$\begin{aligned}\overline{\mathrm{P}}_{\rm b}^{(t,n)} &= \mathbb{E}\mathrm{P}_i^{(t)}(\mathsf{B}(i,t)) \\ &\le \mathbb{E}_0\mathrm{P}_i^{(t)}(\mathsf{T}_t) + \sum_{s=1}^{t}\frac{\delta_s}{n}\,\mathbb{E}_{1,s}\mathrm{P}_i^{(t)}(\mathsf{S}_t^{(s)}) \\ &\quad + \sum_{s=1}^{t-1}\frac{\widehat{\delta}_s}{n}\,\widehat{\mathbb{E}}_{1,s}\mathrm{P}_i^{(t)}(\mathsf{R}_t^{(s)}) + \frac{C\,\beta^{3t}}{n^2}\,,\end{aligned} \tag{11}$$

where $\mathbb{E}_0$, $\mathbb{E}_{1,s}$, $\widehat{\mathbb{E}}_{1,s}$ denote expectation with respect to the neighborhood ensembles defined in the previous section.

Since $\overline{\mathrm{P}}_{\rm b}^{(n)} = \inf_t \overline{\mathrm{P}}_{\rm b}^{(n,t)}$, we obtain an upper bound on $\overline{\mathrm{P}}_{\rm b}^{(n)}$ by setting $t = t_*(n) = \xi\log n$. In particular, if we chose $\xi \in (0,1/3)1$ the last term in the above expression is $o(n^{-1})$. Further, by Proposition 1 we can take ϵ_* such that $\kappa(\epsilon)\xi > 1$ for all $\epsilon \le \epsilon_*$. As a consequence for $\epsilon \le \epsilon_*$, the first term is $o(n^{-1})$ as well.

We are now left with the task of estimating the two sums in Eq. (11) that include expectations over unicyclic

neighborhoods. Given an integer $T > 0$, we distinguish in each of the two sums the terms with $s \le T$ and those with $s > T$. We claim that the last one are negligible. More precisely

$$\sum_{s=T+1}^{t_*(n)} \frac{\delta_s}{n} \, \mathbb{E}_{1,s} \mathrm{P}_i^{(t)}(\mathsf{S}_t^{(s)}) \le r_T n^{-1} \,, \qquad (12)$$

where r_T can be made arbitrarily small by taking T large enough.

Before considering this claim, let us focus on the terms with $s \le T$ which form the leading contribution. Assume for notational simplicity that the all zeros codeword has been transmitted. Consider the graph $\mathsf{S}_{t_*(n)}^{(s)}$ and the BP messages at any $t > 0$. All the messages entering the unique loop in $\mathsf{S}_{t_*(n)}^{(s)}$ are *exactly* distributed according to the tree ensemble, i.e. as predicted by density evolution. Since $t_*(n) = \xi \log n \uparrow \infty$, this means that the incoming log-likelihood ratios are, asymptotically in t and n, $+\infty$. If any of the variable nodes in the loop has degree 3 or larger, this implies that along edges inside the loop positive messages are passed at all times large enough. The root node i will be decoded correctly both if it belongs to the unique loop in $\mathsf{S}_t^{(s)}$, and if it does not.

Consider now the case in which all the variable nodes on the loop have degree 2. The BP dynamics inside the loop will converge to the one inside a 'loop code' (an LDPC code whose Tanner graph is a single loop, cf. Fig. 7. This is in turn easily understood. Each variable-to-check $j \to b$ message directed clockwise is equal to the one flowing along the previous variable-to-check edge one time step earlier, plus the log-likelihood of x_j (call it z_j). An analogous statement is valid of course for counter-clockwise messages. Asymptotically for large t all the messages in the loop are $(\sum_j z_j)t/\ell$. As a consequence the bits in the loop are decoded correctly if and only if $\sum_j z_j > 0$ (and with probability $1/2$, if $\sum_j z_j = 0$). Notice that, in any case, messages outgoing from the loop are either positive or bounded uniformly in t (if $\sum_j z_j = 0$). The root node i is incorrectly decoded only if it belongs to such a loop and $\sum_j z_j < 0$ (and with probability $1/2$, if $\sum_j z_j = 0$).

The above argument implies that, for fixed s and $t \to \infty$ $\mathrm{P}_i^{(t)}(\mathsf{S}_t^{(s)}) \to 0$ if the loop in $\mathsf{S}_t^{(s)}$ contains nodes of degree 3 or larger, while $\mathrm{P}_i^{(t)}(\mathsf{S}_t^{(s)}) \to \mathfrak{p}_{2s}$ otherwise. Further, it is well known that the probability that the root belongs to such a loop is $\frac{1}{2n(2s)}(\lambda'(0)\rho'(1))^{2s}$.

Neighborhoods of the form $\mathsf{R}_t^{(s)}$ are treated analogously. By summing on $s \le T$, we get a contribution to the form $[\mathfrak{F}(\epsilon) - r'_T]/n$, where r'_T can be made arbitrarily small by chosing T large enough.

The proof is completed by showing that the contribution of loops of large size (larger than $2T$) is negligible. This can be done by showing that either of the two following happens: (1) $s > (1-\eta)t_*(n)$ for some small η: then only a little fraction of received messages is used twice and one can reduce itself to the tree case; (2) $s < (1-\eta)t_*(n)$: then, with high probability, all the messages entering the loop are positive. One can reduce itself to the small loop case, but for the fact that $\sum_j z_j > 0$ whp.

IV. Open problems

The analysis presented in this paper admit several generalizations. Among the most immediate, it would be interesting to generalize our main result to expurgated ensembles as well to other message passing algorithms. It would be also important to improve the accuracy by considering higher order (multi-loop) structures.

References

[1] T. Richardson and R. Urbanke, *Modern Coding Theory*, draft available at `http://lthcwww.epfl.ch/index.php`

[2] A. Amraoui, A. Montanari, and R. Urbanke, "How to Find Good Finite-Length Codes: From Art towards Science," Eur. Trans. Telecom. Special issue Selected Papers from 4^{th} International Symposium on Turbo Codes and Related Topics, Munich April 3-7, 2006

[3] C. Di and R. Urbanke, unpublished (2005)

[4] C. Wang, S. R. Kulkarni, and H. V. Poor, "Exhausting Error-Prone Patterns in LDPC Codes," IEEE Trans. Inform. Theory, submitted.

[5] T. Richardson, "Error floors of LDPC codes," Proc. 41^{st} Annual Allerton Conf. on Comm., Contr., and Computing. Monticello, IL, 2003

[6] M. Luby, M. Mitzenmacher, A. Shokrollahi, D. A. Spielman, and V. Stemann, "Practical loss-resilient codes," In *Proceedings of the* 29*th annual ACM Symposium on Theory of Computing*, pages 150–159, 1997.

[7] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "The Generalized Area Theorem and Some of its Consequences," IEEE Trans. Inform. Theory, submitted.