Finite-length analysis of the TEP decoder for LDPC ensembles over the BEC

Pablo M. Olmos, Fernando Pérez-Cruz
Departamento de Teoría de la Señal y Comunicaciones.
Universidad Carlos III in Madrid.
email: {olmos,fernando}@tsc.uc3m.es

Luis Salamanca, Juan José Murillo-Fuentes
Departamento de Teoría de la Señal y Comunicaciones.
Universidad de Sevilla.
email:{salamanca,murillo}@us.es

Abstract—In this work, we analyze the finite-length performance of low-density parity check (LDPC) ensembles decoded over the binary erasure channel (BEC) using the tree-expectation propagation (TEP) algorithm. In a previous paper, we showed that the TEP improves the BP performance for decoding regular and irregular short LDPC codes, but the perspective was mainly empirical. In this work, given the degree-distribution of an LDPC ensemble, we explain and predict the range of code lengths for which the TEP improves the BP solution. In addition, for LDPC ensembles that present a single critical point, we propose a scaling law to accurately predict the performance in the waterfall region. These results are of critical importance to design practical LDPC codes for the TEP decoder.

I. INTRODUCTION

Tree-structured expectation propagation (TEP) was first proposed in [1] to construct tractable approximations to graphical models. Its application to low-density parity-check (LDPC) decoding is introduced in [2], [3], [4] for the binary erasure channel (BEC), where we show that it improves the performance of the belief propagation (BP) algorithm, thanks to a better estimation of the posterior probability for each coded bit. Contrary to alternative techniques to BP, such as generalized-BP (GBP) [5], the TEP complexity for the BEC is of the same order as BP, i.e., linear with the LDPC code length [6]. For LDPC over the BEC, the TEP can be explained as a Gaussian elimination algorithm where rows with up to two ones can be processed. For properly designed Raptor codes, extensively considered nowadays in broadcast erasure networks, inactivation decoding [7], [8] achieves the ML solution by GE of rows with no more than two ones, i.e., is equivalent to the TEP.

In [4], [3], we compute in detail the TEP decoder graph evolution, which constitutes the basic tool to analyze the TEP decoder properties over the BEC. In [4], we show the conditions that have to be fulfilled by an LDPC ensemble to improve the BP limiting performance, i.e., the decoding threshold \( \epsilon_{\text{BP}} \). These conditions are quite restrictive and no code satisfying them have been found to date. On the contrary, the TEP decoder significantly improves the BP solution when decoding finite-length LDPC codes [2], [3]. In [2], we illustrate this result for regular and irregular LDPC ensembles. Based on empirical evidence, we concluded that TEP either provides a gain in word error rate (WER) for a given code length, or a reduction of the code length needed to achieve a target WER.

In this contribution, we analyze the finite-length performance of the TEP decoder for the BEC from a theoretical perspective. For a given LDPC ensemble, we are able to explain the gain in performance with respect to BP by comparing the expected graph evolution for each decoder. In addition, we derive a scaling law (SL) that accurately predicts the TEP performance for any given LDPC ensemble in the waterfall region. The TEP decoder SL can be used for TEP-oriented finite-length LDPC optimization. These results can be also applied to estimate the finite-length ML decoder performance for Raptor codes.

II. TEP decoder

The TEP decoder for the erasure channel can be formulated as a peeling algorithm [6], which works over the Tanner graph of the code [4], [2]. We assume, without loss of generality, that the all zero codeword\(^1\) is transmitted through a BEC with erasure parameter \( \epsilon \). The Tanner graph is initialized by removing all the non-erased variables, along with all their links. Then, one of the two following steps is performed at each iteration:

1) **Variable identification:** Look for a degree-one check node in the graph and remove it along with the variable it is connected to. The parity of the check node indicates the value of the variable, which is declared as revealed.

2) **Graph reduction:** Look for a degree-two check node in the graph and remove it along with one of the variables it is connected to. The removed variable will be revealed once the remaining variable in the graph is revealed. This process is sketched in Fig. 1. The variable \( V_1 \) inherits the connections of \( V_2 \) (solid lines) in Fig. 1(b) and the check \( P_1 \) and the variable \( V_2 \) are removed.

The BP decoder as a peeling-type algorithm, usually referred to as peeling decoder (PD), only implements the first step, i.e., variable identification. Note the equivalence between the graph reduction step and the triangulation procedure in [9] for Gaussian elimination. During triangulation procedure, a variable is related to some “reference” variables or pivots, as done during the graph reduction step. Accordingly, the TEP decoder triangulates the residual matrix by only processing rows with up to two ones.

\(^1\)The full algorithm is described in [4].
The TEP operates until all variable nodes have been removed, i.e., successful decoding, or until there are no degree-one and degree-two check nodes left. In the graph reduction step, we eventually create check nodes of degree one. For example, in Fig. 2(a), we create a check node of degree one since $V_1$ and $V_2$ also share $P_3$, a check node of degree three. Once $P_1$ and $V_2$ are removed, the check node $P_3$ becomes degree one, as illustrated in Fig. 2(b). $V_3$ is now revealed.

The TEP decoder complexity is analyzed in [4], where we show that the graph reduction step is performed by a basic BP iteration followed by the addition of two columns of the parity check matrix. Along the decoding process, the matrix remains sparse and the resulting complexity is of order $O(n)$, where $n$ is the code length.

### III. Expected Graph Evolution under TEP Decoding

Both the PD and the TEP decoder sequentially reduce the LDPC Tanner graph and, as a consequence, the decoding process yields a random sequence of residual graphs. As proved in [10], this random process follows a typical path or expected evolution [6] that is computed as the solution of a set of differential equations. In [11], the authors show that individual sequences of residual graphs under BP decoding are Gaussian distributed around such expected evolution, and that the covariance decays as $O(1/n)$.

Consider an LDPC ensemble of code length $n$ and degree distribution (DD) $(\lambda(x), \rho(x))$:

$$\lambda(x) = \sum_{i=2}^{\text{max}} \lambda_i x^{i-1}, \quad \rho(x) = \sum_{j=2}^{\text{max}} \rho_j x^{j-1},$$

where $\lambda_i$ represents the fraction of edges connected to a variable node of degree $i$, i.e., edges with left degree $i$, and $\rho_j$ is the fraction of edges connected to a check node of degree $j$, i.e., edges of right degree $j$ [10]. In [4], we compute the expected graph evolution during the TEP decoding, also with a set of non-linear differential equations. These equations track down the expected progression of the fraction of edges with left degree $i$, $l_i(\tau)$ for $i = 1, \ldots, l_{\text{max}}$, and right degree $j$, $r_j(\tau)$ for $j = 1, \ldots, r_{\text{max}}$, as the TEP decoder iterates. The normalized time $\tau$ is such that if $\ell$ is the TEP iteration index and $E \propto n$ is the total number of edges in the original graph then $\tau = \ell/E$. The normalized size of the residual graph along time is given by $e(\tau) = \sum_j r_j(\tau) = \sum_i l_i(\tau)$. The initial conditions for the TEP differential equations depend on both the DD in (1) and the erasure probability $\epsilon$.

Let us illustrate the accuracy of this model for the analysis of the TEP decoder properties. In Fig. 3, we compare the estimated TEP expected graph evolution (in solid lines) for $R_1(\tau) = r_1(\tau)E$ and $R_2(\tau) = r_2(\tau)E$ for a regular $(3,6)$ code with 20 particular decoding trajectories obtained through simulation (thin lines). We choose a very large code length, $n = 2^{17}$, so that the deviation around the expected evolution is small. Since the BP threshold is $\epsilon_{BP} = 0.4294$, we consider two cases: below the threshold, $\epsilon = 0.415$ in (a), and above, $\epsilon = 0.43$ in (b). We depict their evolution $e(\tau)$. As we can see, our model in [4] accurately estimates the average decoding trajectories. Note that, below the threshold, all the curves represent successful decoding.

To perform better than the BP decoder, the TEP decoder needs to create a significant amount of check nodes of degree one. In other words, the probability of the scenario in Fig. 2(a) should be large. In [4] we computed the probability that two variable nodes that share a check node of degree-two also share at least another check node (of any degree). Let $S$ denote this scenario. For the graph at time $\tau$, if we randomly choose a particular degree-two check node, the probability of this scenario is lower bounded by

$$P_S(\tau) \geq \frac{(l_{\text{avg}}(\tau) - 1)^2 (r_{\text{avg}}(\tau) - 1)}{e(\tau)E},$$

where $l_{\text{avg}}(\tau)$ and $r_{\text{avg}}(\tau)$ are, respectively, the average left and right edge degrees. The probability of the scenario in Fig. 2(a) is just a fraction of (2). For most codes of interest, in the limit $n \to \infty$, $P_S(\tau) \to 0$ for any $\tau$, proving that the TEP and BP present the same limiting performance and threshold [4]. For finite-length codes, $P_S(\tau)$ is higher than zero and the improvement with respect to the BP decoder is possible.

### IV. TEP Gain in the Finite-Length Regime

In [11], the authors prove that the BP performance for finite-length LDPC codes can be predicted by analyzing the average presence of degree-one check nodes at a finite set of time instants during the BP decoding. These points are referred to as critical points. For a given ensemble with $n \to \infty$, if $r_{\text{BP}}^1(\tau)$ is the expected evolution of the fraction of degree-one check
nodes under BP decoding, we say that $\tau^*$ is a BP-critical point if
\[
\lim_{\epsilon \to \epsilon_{\text{BP}}^*} r_1^\text{BP}(\tau^*) = 0,
\]
where $r_1^\text{BP}(\tau)$ is computed analytically in [10]:
\[
r_1^\text{BP}(u) = \epsilon \lambda(u) \left( u - 1 + \rho(1 - \epsilon \lambda(u)) \right),
\]
and
\[
\frac{\partial u}{u} = \frac{- \partial \tau}{\epsilon(\tau)} \Rightarrow u = \exp \left( - \int_0^\tau \frac{ds}{\epsilon(s)} \right).
\]
Given the parameter $u$ in (5), the decoding process starts at $u = 1$ and finishes, if it succeeds, at $u = 0$. In [11], it is shown that the BP decoder at $\epsilon = \epsilon_{\text{BP}} - \Delta \epsilon$ succeed with very high probability as long as there exists, on average, degree-one check nodes at the critical points. The finite-length performance is estimated by computing the cumulative probability distribution of individual trajectories at the critical points. We show that the same idea extends to the TEP case.

For the TEP decoder, we extend the notion of critical points: given an LDPC ensemble with $n \to \infty$, we say that $\tau'$ is a TEP-critical point of the LDPC ensemble if
\[
\lim_{\epsilon \to \epsilon_{\text{TEP}}} r_1(\tau') = 0,
\]
where $r_1(\tau')$ is the evolution of degree-one check nodes under TEP decoding, computed in [4]. The following lemma relates the number of TEP and BP critical points for the same LDPC ensemble. The complete proof of this lemma can be found in the extended version of this paper in [4]:

**Lemma 1:** For a given LDPC $[\lambda(x), \rho(x), n \to \infty]$ ensemble used for transmission over the BEC, the number of TEP critical points is equal to the number of BP critical points.

To estimate the TEP performance, we compute the cumulative probability distribution of decoding trajectories at the critical points. For a given finite-length ensemble, the TEP expected graph evolution provides the average presence of degree-one check nodes at any time $\tau$, i.e., $r_1(\tau, n, \epsilon)$. Unlike the BP case in (4), the average trajectory $r_1(\tau, n, \epsilon)$ is strongly dependent on the code length $n$. For instance, in Fig. 4, we include the solution for $r_1(\tau, n, \epsilon)$ for the $(3, 6)$ regular code and different code lengths $n = 2^i$ for $i = 12, \ldots, 17$ at $\epsilon = \epsilon_{\text{BP}} = \epsilon_{\text{TEP}}$. For $n \geq 2^{17}$, we locate the unique TEP-critical point at $\epsilon(\tau) \approx 0.1$, where $r_1(\tau', n, \epsilon_{\text{BP}})$ vanishes. For lower code lengths, $r_1(\tau', n, \epsilon_{\text{TEP}})$ does not vanish since the graph is small enough to make $P_S(\tau)$ in (2) significant. The fraction $r_1(\tau', n, \epsilon_{\text{TEP}})$ represents the number of degree-one check nodes that the TEP decoder is able to create with respect to the BP solution, which is zero at the critical point.

As we increase $n$, we observe that, above a certain length threshold $n_0$, $r_1(\tau', n, \epsilon_{\text{TEP}})$ becomes zero. On the one hand, for $n \geq n_0$, we do not expect significant mismatch between the BP and the TEP solutions. On the other hand, for $n < n_0$, a closer look to the curves in Fig. 4 shows that $r_1(\tau', n, \epsilon_{\text{TEP}})$ decays approximately as $O(1/n)$, which is consistent with with probability $P_S(\tau)$ in (2):
\[
r_1(\tau', n, \epsilon_{\text{TEP}}) \approx \gamma_{\text{TEP}}^{-1} n^{-1},
\]
where $\gamma_{\text{TEP}}$ depends on the ensemble. We obtain this value by linearly interpolating $r_1(\tau', n, \epsilon_{\text{TEP}})^{-1}$ for different code lengths. Indeed, we have observed that (7) is a general property, where the constant $\gamma_{\text{TEP}}$ depends on the ensemble. For instance, for the $(3, 6)$ regular code we obtain $\gamma_{\text{TEP}}^{-1} = 0.3194$ and for the following single-critical-point irregular LDPC ensemble with BP threshold $\epsilon_{\text{BP}} = 0.4828$ defined by
\[
\lambda(x) = \frac{1}{6} x + \frac{5}{6} x^3, \quad \rho(x) = x^5,
\]
we computed $\gamma_{\text{TEP}}^{-1} = 0.2925$.

\[\frac{\partial}{\partial \tau} \Rightarrow u = \exp \left( - \int_0^\tau \frac{ds}{\epsilon(s)} \right).\]
A. BP SL for single critical-point LDPC ensembles

Consider a transmission over the BEC using the ensemble LDPC(\(\lambda(x), \rho(x), n\)), which has a single non-zero critical point. Under these conditions, in [11] it was proposed the following SL to predict the BP error performance:

\[
\mathbb{E}_{\text{LDPC}}[\lambda(x), \rho(x), n][F_{\text{BP}}(C, \epsilon)] = \mathcal{Q}\left(\sqrt{\alpha_{\text{BP}}(\epsilon_{\text{BP}} - \beta_{\text{BP}} n^{-2/3} - \epsilon)}\right) \left(1 + \mathcal{O}(n^{-1/3})\right),
\]

where \(F_{\text{BP}}(C, \epsilon)\) is the average block error probability for \(C \in \text{LDPC}(\lambda(x), \rho(x), n)\), \(\alpha_{\text{BP}} = \alpha_{\text{BP}}(\lambda(x), \rho(x))\) and \(\beta_{\text{BP}} = \beta_{\text{BP}}(\lambda(x), \rho(x))\), i.e., \(\alpha_{\text{BP}}\) and \(\beta_{\text{BP}}\) are constants that depend only on the degree distribution. The analytical expression for the parameters \(\alpha_{\text{BP}}\) and \(\beta_{\text{BP}}\) can be found in [11], [6].

It is important for our work to provide further details about the derivation of the SL in (9). For \(\epsilon = \epsilon_{\text{BP}}\), we know by assumption that there only exists one critical point \(\tau^*\). If we now vary \(\epsilon\) slightly, the expected fraction of check nodes at \(\tau^*\) can be estimated by Taylor expansion:

\[
r_{1,\text{BP}}(\tau^*) = \left. \frac{\partial r_{1,\text{BP}}(\tau)}{\partial \epsilon} \right|_{\tau = \tau^*} \Delta \epsilon,
\]

where \(\Delta \epsilon = (\epsilon - \epsilon_{\text{BP}})\) [11]. Because of the channel dispersion, individual decoding trajectories are Gaussian distributed with a variance of order \(\mathcal{O}(1/n)\) [6], [11]. Hence, at point \(\tau^*\), the observed fraction of degree-one check nodes is a normal random variable with mean given in (10) and variance denoted by \(\sigma_{r_{1,\tau^*}}^2/n\). A coarse estimation of the error probability is obtained by computing the probability that this random variable is less than zero:

\[
\mathbb{E}_{\text{LDPC}}[\lambda(x), \rho(x), n][F_{\text{BP}}(C, \epsilon)] \approx 1 - \mathcal{Q}\left(\frac{\partial r_{1,\text{BP}}(\tau)}{\partial \epsilon}\right) \left|_{\tau = \tau^*} \frac{\Delta \epsilon}{\sqrt{\sigma_{r_{1,\tau^*}}^2/n}}\right.
\]

(11)

where

\[
\alpha_{\text{BP}} = \sqrt{\sigma_{r_{1,\tau^*}}^2/n} \left(\frac{\partial r_{1,\text{BP}}(\tau)}{\partial \epsilon}\right) \left|_{\tau = \tau^*}\right.^{-1}.
\]

(12)

For moderate block-lengths, we are underestimating the error probability since we ignore those trajectories for which the fraction of check nodes of degree one vanishes at \(\tau \leq \tau^*\). This effect is included in the scaling function by introducing the \(\beta_{\text{BP}}\) parameter in (9).

B. TEP SL for single critical-point LDPC ensembles

We now show that the TEP decoder can be modeled by a similar scaling law and we provide a first estimate of the scaling parameters as a function of the LDPC DD.

The function \(r_{1}(\tau, n, \epsilon_{\text{BP}})\) provides the average presence of degree-one check nodes at the code’s critical point for finite block-lengths. We proceed as in (10)-(12) to estimate the TEP error probability. The LDPC ensemble has threshold \(\epsilon_{\text{TEP}} = \epsilon_{\text{BP}}\) and a critical point at \(\tau^*\). Assume now a slight deviation of the erasure probability \(\epsilon = \epsilon_{\text{TEP}} + \Delta \epsilon\). According to (7), the average fraction of degree-one check nodes has the following Taylor expansion:

\[
r_{1}(\tau', n, \epsilon) = \gamma_{\text{TEP}} n^{-1} + \left. \frac{\partial r_{1}(\tau, n, \epsilon)}{\partial \epsilon} \right|_{\epsilon = \epsilon_{\text{TEP}}} \Delta \epsilon.
\]

(13)

To estimate the TEP error probability, we consider that the TEP decoder will succeed with high probability as long as the fraction of degree-one check nodes at \(\tau'\) is positive. Therefore, assuming a Gaussian distribution around the mean in (13) with a variance denoted by \(\sigma^2_{r_{1,\tau'}}\), we have

\[
\mathbb{E}_{\text{LDPC}}[\lambda(x), \rho(x), n][F_{\text{TEP}}(C, \epsilon)] = \mathcal{Q}\left(\frac{\sqrt{\gamma_{\text{TEP}}}}{\sigma_{r_{1,\tau'}}} \frac{\partial r_{1}(\tau, n, \epsilon)}{\partial \epsilon} \left|_{\epsilon = \epsilon_{\text{TEP}}} \right. \Delta \epsilon\right)\]

(14)

where \(F_{\text{TEP}}(C, \epsilon)\) is the TEP block error rate for the code \(C \in \text{LDPC}(\lambda(x), \rho(x), n)\) and

\[
\alpha_{\text{TEP}} = \left(\frac{\partial r_{1}(\tau, n, \epsilon)}{\partial \epsilon} \right) \left|_{\epsilon = \epsilon_{\text{TEP}}} \right.^{-1} \sigma^2_{r_{1,\tau'}}.
\]

(15)

The analytical computation of the normalized variance \(\sigma^2_{r_{1,\tau'}}\) at the critical point requires to solve, for the TEP and each particular ensemble, the covariance graph evolution equations [11]. Because the variance of the trajectories essentially comes from the channel dispersion [12] and from the LDPC ensemble [11], a first approach to estimate this
parameter is given by the BP variance at the corresponding critical point, i.e., \( \alpha_{BP}^{\tau_1, \tau_1}(\tau^*) \). Besides, for \( n \to \infty \), both the TEP and BP converge to the same performance, so we set \( \alpha_{TEP} = \alpha_{BP} \). Due to the assumption of these values for the TEP parameters, our model provides an upper bound to the TEP performance because, the better the decoder is, the less variance we expect among different trajectories.

\[
\delta_{BP}^{\tau_1, \tau_1}(\tau^*) \approx \frac{\delta_{TEP}^{\tau_1, \tau_1}(\tau^*)}{\alpha_{TEP}} \approx \frac{\delta_{TEP}^{\tau_1, \tau_1}(\tau^*)}{\alpha_{BP}} \approx \frac{\gamma_{TEP}}{\gamma_{BP}} \approx \frac{\tau_{BP}}{\tau_{TEP}} = \frac{\tau_{BP}}{\tau_{BP}}.
\]

Let us include a couple of examples. In Fig. 5 (a), we compare the SL solution in (14) in dashed lines with real performance data obtained through simulation in solid lines for the regular (3,6) LDPC ensemble for code lengths \( n = 2^i \) for \( i = 10, \ldots, 12 \). As shown in [2], for these code lengths the TEP significantly improves the BP solution. The match between dashed and solid lines is quite good, proving the accuracy of the model for the TEP performance and the parameter estimation proposed. Due to the assumption of the BP values for \( \delta_{TEP}^{\tau_1, \tau_1}(\tau^*)/n \) and \( \alpha_{TEP} \), we slightly overestimate the TEP error probability. In Fig. 5 (b), we include with similar conclusions the results for the irregular code defined in (8). In all cases, the real error curves have been averaged throughout one hundred code samples chosen to present large enough minimum distance so that the error floor is far below \( 10^{-4} \).

C. Future directions

In this work, we provide a method to analyze the TEP finite-length performance for an LDPC code. For ensembles that exhibit a single critical point, a model to estimate the error rate as a function of the channel parameter is provided. We show that the BP equivalent parameters provide a good first approach. However, irregular capacity-achieving LDPC ensembles typically have several critical points. In [13], the authors propose a BP scaling law that takes into account the effect of all the critical points and its extension to the TEP decoder should be done in the same way as in (14). The computation of the (multiple) scaling parameters, for both the TEP and the BP decoders, is the real challenge. On the other hand, the main open problem is, without any doubt, the analysis of the TEP decoder performance over general binary memoryless channels.

References


