The Slope Scaling Parameter for General Channels, Decoders, and Ensembles

Jérémie Ezri*, Andrea Montanari†, Sewoong Oh†, and Ruediger Urbanke*
* School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland,
Email: jeremie.ezri@epfl.ch and ruediger.urbanke@epfl.ch
†Electrical Engineering and Statistics Departments, Stanford University, Stanford, CA
Email: montanari@stanford.edu and swoh@stanford.edu

***Abstract*— Scaling laws are a powerful way to analyze the performance of moderately sized iteratively decoded sparse graph codes. Our aim is to provide an easily usable finite-length optimization tool that is applicable to the wide variety of channels, blocklengths, error probability requirements, and decoders that one encounters for practical systems. The tool is aimed at non-experts in the field, who need to quickly find code designs that are comparable with the best known codes available today but do not have the luxury of spending months in doing so. In previous work we have shown how to compute scaling parameters for transmission over the binary erasure channel, as well as general channels and general quantized message-passing decoders when applied to *regular* ensembles. In this paper we show how to compute the message variance for a fixed number of iterations for *irregular* low-density parity-check ensembles. From these calculations the basic scaling parameter α can be deduced by determining the leading term of the limiting expression when the number of iterations tends to infinity and the channel parameter approaches the density evolution threshold.**

I. Introduction

Let us first trace the past developments in our effort to use scaling laws for the efficient optimization of moderately sized and iteratively decoded sparse graph codes. In Section II we then present the recent progress.

The idea of scaling laws was introduced to coding theory by Montanari [9]. By means of a specific example (regular ensemble, binary symmetric channel with parameter ϵ) he showed that if one plots the block error probability as a function of the "scaled variable" $z = \sqrt{n}(\epsilon^{\text{BP}} - \epsilon)$ (where ϵ^{BP} is the density evolution threshold), then the curves corresponding to increasing blocklengths quickly converge to a single "mother curve". This common curve is called the *scaling function*. A particular example can be seen in the right-hand side graph in Figure 1. This suggested that if one were able to analytically determine the scaling function as well as the scaling parameters for a given system (degree distribution, channel, decoder) then it would be possible to efficiently and accurately predict the finite-length performance of iterative coding systems.

The first analytic result was derived by Amraoui, Montanari, Richardson, and Urbanke [2], [1]. They showed that, for transmission over the binary erasure channel (BEC) with parameter ϵ, the block error probability behaves like

$$P_B(z) = Q\left(\frac{z}{\alpha}\right)(1 + o_n(1)),$$

where $z = \sqrt{n}(\epsilon^{\text{BP}} - \epsilon)$ and where ϵ^{BP} is the threshold of the ensemble under BP decoding. They further conjectured that, more accurately,

$$P_B(z) = Q\left(\frac{z}{\alpha}\right)(1 + O(n^{-\frac{1}{3}})), \tag{1}$$

where $z = \sqrt{n}(\epsilon^{\text{BP}} - \beta n^{-\frac{2}{3}} - \epsilon)$, and where the term $\beta n^{-\frac{2}{3}}$ represents a *finite-length shift* of the threshold (the point where the block error probability takes on the value one-half). This refined scaling law was shown to be correct for left-regular right-Poisson ensembles by Dembo and Montanari [5].

Analytically, the scaling law promises the convergence of $P_B(z)$ for a fixed z and n tending to infinity. In practice, the scaling law provides accurate predictions already for moderate blocklengths and also away from the threshold; see e.g., the middle graph in Figure 1. The analysis put forward in [2], [1] was based on the so called *peeling decoder* by Luby, Mitzenmacher, Shokrollahi, Spielman, and Steman, [8]. This decoder is equivalent to the standard BP decoder but represents the decoding process as a sequence of discrete steps, where each step corresponds to determining a previously erased bit from its known neighbors. It was shown in [2], [1] how to determine the scaling parameter α from the solution of a system of differential equations, which were dubbed covariance evolution. Unfortunately no analytic solution of this system is known to date; therefore, α had to be computed numerically.

An alternative way to determine α for the BEC case was proposed in [3]. Recall that density evolution computes the *average* number of erased messages as a function of the iteration number. It was shown in [3] that α can be computed by determining the *variance* of the number of erased messages. This computation was accomplished in [4] and an explicit value for α as a function of the degree distribution pair (λ, ρ) was given. Further, explicit expressions for β were derived. Given these explicit expressions of the scaling parameters, it is then easy to accomplish a finite-length optimization.

All these developments were restricted to the BEC. The first step towards extending the scaling law to more general channels was accomplished in [6]. In this paper the authors considered transmission over the binary symmetric channel (BSC) and decoding via the Gallager algorithm A. This algorithm has the property that for most ensembles the threshold is given by a fixed point at the *beginning* of the decoding process.

This somewhat simplifies the determination of the scaling law and the scaling parameter.

As explained above, the derivation of α for the BEC in [4] is based on the peeling decoder. Unfortunately, no generalization of the peeling decoder to general channels is known. Ezri, Montanari, and Urbanke proposed in [7] an alternative computation model to determine the scaling parameter α. This is based on an EXIT-like curve, a concept which is easily extended to general BMS channel. In more detail, consider transmission over a BMS channel characterized by its channel entropy $\mathtt{h}$ and decoded with a generic quantized MP decoder which satisfies the standard message-passing symmetry conditions [10]. To be concrete, assume that the messages sent in the decoder are quantized and that they take their values in a real-valued finite alphabet $\{-w_m, \ldots, w_m\}$, $m \in \mathbb{N}$, where $0 = w_0 < w_1 < \cdots < w_m$. This representation is particularly convenient for the important class of quantized belief propagation (BP) decoders. In this case we assume that the messages represent actual log-likelihood values. The decoding rules are then the standard BP decoding rules followed by a quantization to the nearest element of the alphabet.

Given the symmetry of the channel and the decoder, we can assume that the all-zero codeword was transmitted. Let a denote the fixed point message density of density evolution. (This density is of course a function of the channel parameter, but to simplify notation we omit this dependence.) More precisely, this is the variable to check node message density. Pick a function $\mathfrak{E}$, $\mathfrak{E} : \mathbb{R}^{2m+1} \mapsto \mathbb{R}$. Although many choices are possible, for the sake of definiteness, let us assume that for a density $\mathsf{v} \in \mathbb{R}^{2m+1}$, $\mathfrak{E}(\mathsf{v}) = v_m$, where v_m is the value of the maximum component of v. Define $x = \mathfrak{E}(\mathsf{a})$, which "measures" the performance of the decoder. The special case x^{MP} is the value of x corresponding to the density evolution threshold. Define an EXIT-like curve as the curve which represents x as a function of $\mathtt{h}$ as depicted on the left-hand side of Figure 1.

In order to be suitable for the derivation of a scaling law we require that the slope of the EXIT-like curve at the threshold $\mathtt{h}^{\text{MP}}$ must be infinite and that its second derivative must be strictly non-zero at this point. It was shown in [7] that in this case α can be expressed as

$$\alpha = \frac{\partial^2 \mathtt{h}(x)}{\partial x^2}\Big|_{\text{MP}} \lim_{x \downarrow x^{\text{MP}}} (x - x^{\text{MP}}) \sqrt{\frac{\mathcal{V}}{\Lambda'(1)}}, \qquad (2)$$

where $\mathcal{V} = \lim_{\ell\to\infty} \lim_{n\to\infty} \mathcal{V}_n^{(\ell)}$, and where $\mathcal{V}_n^{(\ell)}$ is the message *variance*. More precisely, let us number all edges in the graph from 1 to $n\Lambda'(1)$. Define $\mu_i^{(\ell)} = 1$ if the variable to check node message along edge i equals w_m in iteration ℓ and 0 otherwise. To lighten the notation, let also define $\mathcal{E} = \{1, \ldots n\Lambda'(1)\}$. In this case,

$$\mathcal{V}_n^{(\ell)} = \frac{\text{Var}\left(\sum_{i\in\mathcal{E}} \mu_i^{(\ell)}\right)}{n\Lambda'(1)}.$$

As mentioned above, we are interested in the value of this variance in the limit as ℓ tends to infinity and when the channel parameter approaches the threshold from above. For each fixed channel parameter unequal to the threshold the limit $\lim_{\ell\to\infty} \lim_{n\to\infty} \mathcal{V}_n^{(\ell)}$ exists and is finite. But the value of this variance diverges when we let the channel parameter approach the threshold from above. One can show that the limit $\lim_{x\downarrow x^{\text{MP}}} \lim_{\ell\to\infty} \mathcal{V}^{(\ell)}(x - x^{\text{MP}})^2 = \xi$ exists and is finite. In other words, close to the critical value x^{MP}, $\lim_{\ell\to\infty} \mathcal{V}^{(\ell)}$ behaves like $\xi/(x - x^{\text{MP}})^2$. For the determination of the scaling parameter α one needs to determine the constant ξ. This has been accomplished for *regular* ensemble in [7]. Therefore, for regular ensembles the scaling parameter α can be computed and can be used to predict the performance. Figure 1 shows the prediction given by the scaling law for a $(3,4)$-regular ensemble used over a BAWGN channel and decoded with a quantized BP decoder.

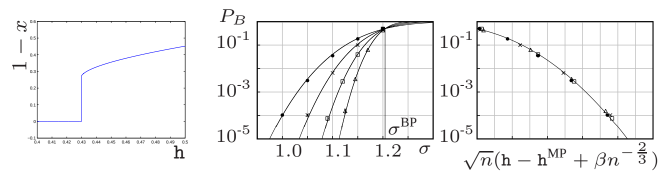

Fig. 1. *Left:* EXIT-like curve for the (3,6)-code over the BEC: $\mathtt{h}^{\text{MP}} \approx 0.42944$. Note that we plot $1 - x$ (and not x) versus $\mathtt{h}$ so that the curve looks like a standard EXIT curve. *Middle:* Block error probability of a (3,4)-code transmitted over an AWGNC and decoded with a 15-quantization levels BP where the messages are bounded by ± 8.12: $\sigma^{\text{BP}}_{15,8.12} = 1.2043$. Comparison of the block error probability $\mathbb{E}_{\mathtt{G}\in\text{LDPC}(nx^3, \frac{3}{4}nx^4)}[P_B^{\text{BP}}(\mathtt{G}, \sigma)]$ (crosses with 95% confidence intervals) determined via simulations with the curves given by the scaling law for $\alpha \simeq 0.97$. The lengths are $n = 1024$, 2048, 4096, and 8192, respectively. Since the scaling law predicts only large-sized errors, the ensembles were expurgated: only error events of size at least 25, 50, 100, and 200, respectively, were counted. *Right:* Rescaled block error probability with respect to $z = \sqrt{n}(\mathtt{h} - \mathtt{h}^{\text{MP}} + \beta n^{-\frac{2}{3}})$, where $\mathtt{h} = h_2(\sigma)$ and where $h_2(\cdot)$ is the binary entropy function. Note that the simulations points all cluster around the single *mother* curve. This curve is the standard Q-function.

II. Variance for a Fixed Number of Iterations

As explained above, the key to computing the scaling parameter α is to compute the variance $\mathcal{V}^{(\ell)} = \lim_{n\to\infty} \mathcal{V}_n^{(\ell)}$ and from this the limit $\mathcal{V} = \lim_{\ell\to\infty} \mathcal{V}^{(\ell)}$. Although strictly speaking we only need the dominant term of $\mathcal{V}$, we opted to compute the exact expression $\mathcal{V}^{(\ell)}$. This makes it possible to verify the expressions against simulations before taking the limit. The derivation is outlined in Section IV.

The final expressions for $\mathcal{V}^{(\ell)}$ are unwieldy and also not very insightful. We therefore do not reproduce them here. Rather, let us give a few examples.

Example 1: As a first example, we consider the (3,4)-regular ensemble used over a BAWGN channel. The decoding is performed by a quantized BP decoder with 15-quantization levels where the reliability in the decoder is bounded by 8.12. More precisely, the 15 quantization levels are chosen equally spaced in the domain $[-8.12, 8.12]$. The message-passing rules are the standard BP message-passing rules (applied always to pairs of incoming messages) followed by a quantization to the nearest level. The threshold of this decoder for a BAWGN channel is $\sigma^{\text{BP}}_{15,8.12} \approx 1.2043$. The left-hand side of Figure 2 represents the variance $\mathcal{V}^{(\ell)}$ as a function of the channel after $\ell = 0, \ldots, 6$ iterations. The crosses represent empirical measurements of $\mathcal{V}^{(\ell)}$ for $n = 10^4$. We see that there is an excellent agreement with the analytic expressions.

Example 2: In our second example, we consider the irregular ensemble characterized by $\lambda(x) = 0.06x + 0.82x^2 + 0.12x^3$ and $\rho(x) = 0.075x^2 + 0.8x^3 + 0.125x^4$. Again we consider transmission over a BAWGN channel. Assume a quantized BP decoder with 15-quantization levels, where the messages are bounded by ± 13.96. The corresponding threshold is $\sigma^{\text{BP}}_{15,13.96} \approx 1.0489$. We plot on the right-hand side in Figure 2 the variance $\mathcal{V}^{(\ell)}$ as a function of the channel after $\ell = 0, \ldots, 5$ iterations. We also plot some points which correspond to an empirical evaluation of the variance for $n = 10^4$. Again we see an excellent agreement.

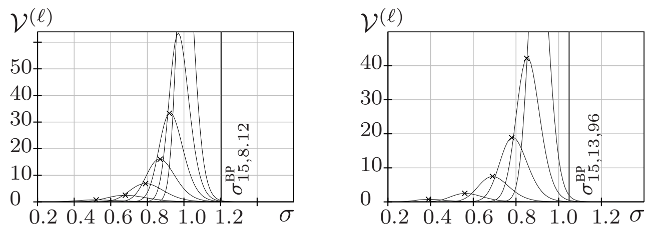

Fig. 2. *Left:* The variance $\mathcal{V}^{(\ell)}$ as a function of the channel parameter σ for $\ell = 0, \ldots, 6$. The ensemble is $(3,4)$-regular, transmission takes place over a BAWGN channel, and decoding is accomplished by a quantized BP decoder with 15-quantization levels and a maximum message of ± 8.12. The threshold of this combination is $\sigma^{\text{BP}}_{15,8.12} \approx 1.2043$. The crosses represent the empirically computed variances for $n = 10^4$. *Right:* The variance $\mathcal{V}^{(\ell)}$ as a function of the channel parameter σ for $\ell = 0, \ldots, 5$. The ensemble has degree distribution $\lambda(x) = 0.06x + 0.82x^2 + 0.12x^3$ and $\rho(x) = 0.075x^2 + 0.8x^3 + 0.125x^4$. Transmission takes place over a BAWGN channel and decoding by a quantized BP decoder with 15-quantization levels and a maximum message of ± 13.96. The threshold of this combination is $\sigma^{\text{BP}}_{15,13.96} \approx 1.0489$. The crosses represent the empirically computed variances for $n = 10^4$.

III. Future Work

As we discussed above, we now know how to compute the variance $\mathcal{V}^{(\ell)} = \lim_{n\to\infty} \mathcal{V}^{(\ell)}_n$ for a general quantized MP decoder and general irregular ensembles. In order to compute the scaling parameter α, it remains to determine

$$\xi = \lim_{x \downarrow x^*} \lim_{\ell\to\infty} \mathcal{V}^{(\ell)} (x - x^{\text{MP}})^2.$$

As might be gleaned from Figure 2 it is difficult to numerically approximate the dominant term from the variance curves for moderate values of ℓ; close to the threshold the variance grows very rapidly as a function of the iteration number and computational approximations quickly run into numerical difficulties.

It is therefore necessary to extract the dominant behavior of the finite-ℓ expressions analytically. For regular ensembles it was shown in [7] how this can be done. Unhappily the finite-ℓ expressions are considerably more complicated for the irregular case, as it will be explained in the next section. So it remains to be seen how the desired limit can be computed.

A further necessary prerequisite in order to be able to effect a meaningful finite-length optimization is the development of efficient expressions to determine the error-floor. Little is currently known on this subject despite its importance for practical applications.

IV. Computation of Variance for Fixed ℓ – Outline

Consider the standard $\text{LDPC}(n, \lambda, \rho) = \text{LDPC}(n, \Lambda, P)$ ensemble. Recall that $\mathbb{E}[\mu_i^{(\ell)}] = x^{(\ell)}$, where $x^{(\ell)} = \mathfrak{E}(\mathsf{a}^{(\ell)})$. As we have seen in Section II, we are interested in computing

$$\begin{aligned}
\mathcal{V}^{(\ell)} &= \lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\sum_{i\in\mathcal{E}} \mu_i^{(\ell)}\right)^2\right] - \mathbb{E}\left[\sum_{i\in\mathcal{E}} \mu_i^{(\ell)}\right]^2}{n\Lambda'(1)} \\
&= \lim_{n\to\infty} \frac{\sum_{j,i\in\mathcal{E}} \mathbb{E}\left[\mu_j^{(\ell)}\mu_i^{(\ell)}\right] - \sum_{j,i\in\mathcal{E}} \mathbb{E}\left[\mu_j^{(\ell)}\right]\mathbb{E}\left[\mu_i^{(\ell)}\right]}{n\Lambda'(1)} \\
&= \lim_{n\to\infty} \sum_{i\in\mathcal{E}} \mathbb{E}\left[\mu_1^{(\ell)}\mu_i^{(\ell)}\right] - \mathbb{E}\left[\mu_1^{(\ell)}\right] \sum_{i\mathcal{E}} \mathbb{E}\left[\mu_i^{(\ell)}\right] \\
&= \lim_{n\to\infty} \sum_{i\in\mathcal{E}} \mathbb{E}\left[\mu_1^{(\ell)}\mu_i^{(\ell)}\right] - n\Lambda'(1)(x^{(\ell)})^2. \qquad (3)
\end{aligned}$$

Let $\mathtt{T}^{(\ell)}$ be the set of indices of all messages whose computation tree ℓ intersect the computation tree of $\mu_1^{(\ell)}$, where both trees have depth ℓ. For convenience, we also add to $\mathtt{T}^{(\ell)}$, indices of edges which are connected to the same variable node as an edge already in $\mathtt{T}^{(\ell)}$. For instance, the edges (b) and (f) in the left-hand side graph of Figure 3 are added to $\mathtt{T}^{(\ell)}$, even though their computation trees do not intersect the computation tree of the root edge. Let $\mathrm{G}^{(\ell)}_{\mathtt{T}}$ be the tree formed by all edges which belong to $\mathtt{T}^{(\ell)}$. Thus, $\mathrm{G}^{(\ell)}_{\mathtt{T}}$ is formed by ℓ variable node layers "above" the root variable node and 2ℓ variable nodes layers "below" the root variable, as depicted in the right-hand side of Figure 3.

Let us partition $\mathtt{T}^{(\ell)}$ into four sets. Orient all edges in $\mathtt{G}$ from variable node to check node. Let $\mathtt{T}^{(\ell)}_1$ denote the set of edges which are in the "future" as seen by the root edge and which are directed in the same direction as the root edge itself. More precisely, these are the edges within $\mathrm{G}^{(\ell)}_{\mathtt{T}}$ which can be reached by paths starting at the root edge and which point in the same direction as the root edge. Next, let $\mathtt{T}^{(\ell)}_2$ denote the set of edges in the future of the root edge but which point in the opposite direction. Let $\mathtt{T}^{(\ell)}_3$ be the set of edges in the past of the root edge and which point in same direction and, finally, let $\mathtt{T}^{(\ell)}_4$ be the set of edges which are in the past of the root and in which point in the opposite direction. These four types of edges are depicted in the left of Figure 3. Let $(\mathtt{T}^{(\ell)})^c$ be the complement of $\mathtt{T}^{(\ell)}$ in $\{1, \ldots, n\Lambda'(1)\}$ and let $(\mathrm{G}^{(\ell)}_{\mathtt{T}})^c$ be the corresponding graph. We can then expand (3) as

$$\mathcal{V}^{(\ell)} = \lim_{n\to\infty} \mathbb{E}\left[(\mu_1^{(\ell)})^2\right] + \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i\in \mathtt{T}_1^{(\ell)}} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big]$$
$$+ \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i\in \mathtt{T}_2^{(\ell)}} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big] + \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i\in \mathtt{T}_3^{(\ell)}} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big]$$
$$+ \lim_{n\to\infty} \mathbb{E}\Big[\sum_{i\in \mathtt{T}_4^{(\ell)}} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big]$$
$$+ \lim_{n\to\infty} \bigg(\mathbb{E}\Big[\sum_{i\in (\mathtt{T}^{(\ell)})^c} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big] - n\Lambda'(1)(x^{(\ell)})^2\bigg). \quad (4)$$

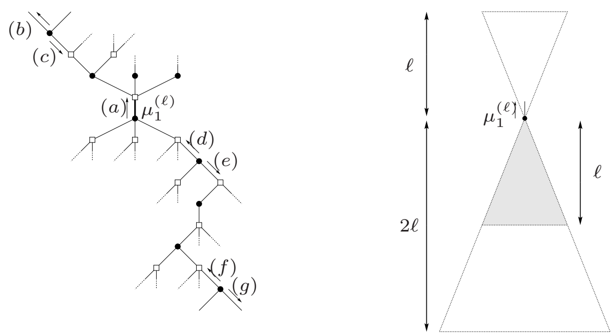

Fig. 3. *Left:* Graph representing the four types of edges contained in $\mathtt{T}^{(2)}$: (a) root edge, (b) element of $\mathtt{T}_1$, (c) element of $\mathtt{T}_2$, (d) and (f) are elements of $\mathtt{T}_3$ and (e) and (g) are elements of $\mathtt{T}_4$; *Right:* The graph $\mathrm{G}_{\mathtt{T}}^{(\ell)}$. It contains ℓ layers of variable nodes "above" the root node and 2ℓ "below." The gray area represents the computation tree of the root edge.

For $i \in \mathtt{T}^{(\ell)}$, it is clear that $\mu_1^{(\ell)}$ and $\mu_i^{(\ell)}$ are correlated since their computation trees intersect and thus they are based on some common received values. It was shown in [7] how to compute the first five terms in (4) for the case of regular ensembles (more precisely, it was shown how to compute the leading order term; in our current context we compute the exact such expressions). The computation of these terms for the irregular case is similar and, hence, we skip the details.

Let us rather concentrate on the sixth term, denote it by

$$S^c = \lim_{n\to\infty} \bigg(\mathbb{E}\Big[\sum_{i\in (\mathtt{T}^{(\ell)})^c} \mu_1^{(\ell)}\mu_i^{(\ell)}\Big] - n\Lambda'(1)(x^{(\ell)})^2\bigg).$$

At first one might think that the root message and a message on an edge in $(\mathtt{T}^{(\ell)})^c$ are uncorrelated since their computation graphs do not intersect. Indeed, this is the case for a regular ensemble, for which S^c is equal to $- \mid \mathtt{T}^{(\ell)} \mid (x^{(\ell)})^2$. The computation of S^c for irregular ensembles on the other hand is considerably more challenging; let us discuss this now.

A. Degree Distribution Correction

The message of the root node is a function of the specific realization of $\mathrm{G}_{\mathtt{T}}^{(\ell)}$. For the regular case, if we consider a fixed number of iterations and the limit of n tending to infinity, there is only one $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ which has a positive probability (namely a regular tree of the appropriate height). But for the irregular case many such computation graphs have a strictly positive probability in the asymptotic limit. Suppose, e.g., that $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ contains an unusual large number of variable nodes of high degree (as compared to $\lambda(x)$). In this case we expect the average (over the noise realization) reliability of the message emitted by the root node to be higher than what is predicted by density evolution. But $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ indirectly also influences the messages in $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$. This is true since the total number of nodes of a given degree is fixed. Therefore, in the above case we know that $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ contains fewer variable nodes of high degree than expected. This causes a small deviation of the degree distribution of $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ as compared to $\lambda(x)$ and, hence, a small deviation of average message in $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ as compared to density evolution. Even though this deviation is only of order $1/n$, there are of order n messages inside $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ and so this deviation gives a non-vanishing contribution even in the limit of infinite blocklengths.

Let us write down this effect explicitly. Given the degree distribution polynomials $\lambda(x)$ and $\rho(x)$, define the operators $\lambda(\mathsf{x}) = \sum_i \lambda_i \mathsf{x}^{\star(i-1)}$ and $\rho(\mathsf{x}) = \sum_i \rho_i \mathsf{x}^{\boxast(i-1)}$, where x is a density, $\star$ is the convolution at variable nodes and $\boxast$ is the convolution at check nodes. This extends to the respective derivatives in the natural way: $\lambda'(\mathsf{x}) = \sum_i (i-1)\lambda_i \mathsf{x}^{\star(i-2)}$ and $\rho'(\mathsf{x}) = \sum_i (i-1)\rho_i \mathsf{x}^{\boxast(i-2)}$. Let us also define the two functions $V^{\mathrm{G}_{\mathtt{T}}^{(\ell)}}(\mathsf{x}) = \sum_i V_i^{\mathrm{G}_{\mathtt{T}}^{(\ell)}} \mathsf{x}^{\star i}$ and $C^{\mathrm{G}_{\mathtt{T}}^{(\ell)}}(\mathsf{x}) = \sum_i C_i^{\mathrm{G}_{\mathtt{T}}^{(\ell)}} \mathsf{x}^{\boxast i}$, where $V_i^{\mathrm{G}_{\mathtt{T}}^{(\ell)}}$ and $C_i^{\mathrm{G}_{\mathtt{T}}^{(\ell)}}$ are the number of variable nodes and check nodes in $\mathrm{G}_{\mathtt{T}}^{(\ell)}$, respectively.

Consider the degree distribution of $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$. We know the overall degree distribution and we are given the degree distribution of $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ itself. Therefore, by removing $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ from the overall bipartite graph we see that the distribution of $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ changes by $\Delta\lambda(x) = \frac{V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\lambda(x) - V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(x)}{n\Lambda'(1)} + O(1/n^2)$ at the variable node side and $\Delta\rho(x) = \frac{C^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\rho(x) - C^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(x)}{n\Lambda'(1)} + O(1/n^2)$ at the check node side.

Consider the effect on density evolution which is caused by this small deviation of the degree distribution. Let $\mathsf{a}^{(j)}$ be the density of messages from variable to check nodes in iteration j and let $\tilde{\Delta}\mathsf{a}^{(j)}$ be the deviation of this quantity due to the deviation of the degree distribution. The density evolution recursion in $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ at time j becomes $\mathsf{a}^{(j)} - \tilde{\Delta}\mathsf{a}^{(j)} = \lambda(\rho(\mathsf{a}^{(j-1)} - \tilde{\Delta}\mathsf{a}^{(j-1)}) - \Delta\rho(\mathsf{a}^{(j-1)} - \tilde{\Delta}\mathsf{a}^{(j-1)})) - \Delta\lambda(\rho(\mathsf{a}^{(j-1)} - \tilde{\Delta}\mathsf{a}^{(j-1)}) - \Delta\rho(\mathsf{a}^{(j-1)} - \tilde{\Delta}\mathsf{a}^{(j-1)}))$, with $\tilde{\Delta}\mathsf{a}^{(0)} = 0$. By expanding this expression, we obtain

$$\tilde{\Delta}\mathsf{a}^{(j)} = \frac{1}{n\Lambda'(1)}\Big(V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\mathsf{a}^{(j)} - \mathsf{a}_{\mathsf{BMS}} \star V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(\mathsf{b}^{(j)})$$
$$+ C^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\mathsf{a}_{\mathsf{BMS}} \star \lambda'(\mathsf{b}^{(j)}) \star \mathsf{b}^{(j)}$$
$$- \mathsf{a}_{\mathsf{BMS}} \star \lambda'(\mathsf{b}^{(j)}) \star C^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(\mathsf{a}^{(j-1)})\Big)$$
$$+ \mathsf{a}_{\mathsf{BMS}} \star \lambda'(\mathsf{b}^{(j)}) \star \Big(\tilde{\Delta}\mathsf{a}^{(j-1)} \boxast \rho'(\mathsf{a}^{(j-1)})\Big) + O(1/n^2),$$

where $\mathsf{b}^{(j)}$ is the density of messages from check to variable nodes in iteration j and $\mathsf{a}_{\mathsf{BMS}}$ is the channel density.

B. Messages Correction

The second term which gives a non-vanishing contribution in the irregular case is due to messages that flow across the boundary from $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ to $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$. These messages influence the neighbors of $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ and thus the average densities in $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$. Indeed, assume that the average density of messages which flow from $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ across the boundary to $(\mathrm{G}_{\mathtt{T}}^{(\ell)})^c$ at time $j < \ell$ is not $\mathsf{a}^{(j)}$ but $\mathsf{a}_\star^{(j)}$. This influences messages up to a distance $\ell - j$ away from the boundary.

Let $\hat{\Delta}\mathsf{a}^{(\ell)}$ be the deviation of $\mathsf{a}^{(\ell)}$ due to the deviation of density which flows out of $\mathrm{G}_{\mathtt{T}}^{(\ell)}$. Due to space constraints we skip the derivation of $\hat{\Delta}\mathsf{a}^{(\ell)}$ but rather discuss how both these terms influence the computation of the variance.

C. Putting it Together

We have seen above that $\mu_1^{(\ell)}$ and $\mu_i^{(\ell)}$, with $i \in (\mathtt{T}^{(\ell)})^c$, are both influenced by the realization of $\mathrm{G}_{\mathtt{T}}^{(\ell)}$ through the degree distribution correction and the messages correction. We can thus write S^c as

$$
\begin{aligned}
&= \lim_{n\to\infty}\left(\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\Big[\sum_{i\in(\mathtt{T}^{(\ell)})^c}\mathbb{E}\Big[\mu_1^{(\ell)}\mu_i^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\Big] - n\Lambda'(1)(x^{(\ell)})^2\right)\\
&= \lim_{n\to\infty}\Big(\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\Big[(n\Lambda'(1)-|\mathtt{T}^{(\ell)}|)\mathbb{E}\Big[\mu_1^{(\ell)}\,|\,\mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\mathbb{E}\Big[\mu_i^{(\ell)}\,|\,\mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\Big]\\
&\quad - n\Lambda'(1)(x^{(\ell)})^2\Big), \quad i\in(\mathtt{T}^{(\ell)})^c\\
&= \lim_{n\to\infty} n\Lambda'(1)\Big(\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\Big[\mathbb{E}\Big[\mu_1^{(\ell)}|\mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\mathbb{E}\Big[\mu_i^{(\ell)}|\mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\Big] - (x^{(\ell)})^2\Big)\\
&\quad - \lim_{n\to\infty}\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\Big[|\ \mathtt{T}^{(\ell)}\ |\ \mathbb{E}\Big[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\mathbb{E}\Big[\mu_i^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\Big], \qquad (5)
\end{aligned}
$$

where $i \in (\mathtt{T}^{(\ell)})^c$ and where in the second equality we have used the fact that $\mu_1^{(\ell)}$ and $\mu_i^{(\ell)}$ are independent once we condition on $\mathrm{G}_{\mathtt{T}}^{(\ell)}$. Write

$$
\mathbb{E}\left[\mu_i^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\right] = x^{(\ell)} - \tilde{\Delta}x^{(\ell)} - \hat{\Delta}x^{(\ell)}, \quad i \in (\mathtt{T}^{(\ell)})^c,
$$

where $\tilde{\Delta}x^{(\ell)}$ is due to the degree distribution correction and $\hat{\Delta}x^{(\ell)}$ to the message correction. Now note that $\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\Big[\mathbb{E}\Big[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\Big]\Big] = x^{(\ell)}$. Further, $\tilde{\Delta}x^{(\ell)} = O(1/n) = \hat{\Delta}x^{(\ell)}$. We can therefore expand (5) as

$$
\begin{aligned}
&- \lim_{n\to\infty} n\Lambda'(1)\left(\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\left[\mathbb{E}\left[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\right]\tilde{\Delta}x^{(\ell)} + \hat{\Delta}x^{(\ell)}\right]\right)\\
&- \lim_{n\to\infty}\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\left[|\ \mathtt{T}^{(\ell)}\ |\ \mathbb{E}\left[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\right]x^{(\ell)}\right]\\
= &- \lim_{n\to\infty} n\Lambda'(1)\left(\mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\left[\mathbb{E}\left[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\right]\mathfrak{E}(\tilde{\Delta}\mathsf{a}^{(\ell)})\right]\right.\\
&\left. + \mathbb{E}_{\mathrm{G}_{\mathtt{T}}^{(\ell)}}\left[\mathbb{E}\left[\mu_1^{(\ell)} \mid \mathrm{G}_{\mathtt{T}}^{(\ell)}\right]\mathfrak{E}(\hat{\Delta}\mathsf{a}^{(\ell)})\right]\right)\\
&- \lim_{n\to\infty} x^{(\ell)}\mathbb{E}\left[V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\mu_1^{(\ell)}\right]\\
= &- \lim_{n\to\infty} n\Lambda'(1)\mathfrak{E}\left(\mathbb{E}\left[\mu_1^{(\ell)}\tilde{\Delta}\mathsf{a}^{(\ell)}\right]\right)\\
&- \lim_{n\to\infty} n\Lambda'(1)\mathfrak{E}\left(\mathbb{E}\left[\mu_1^{(\ell)}\hat{\Delta}\mathsf{a}^{(\ell)}\right]\right)\\
&- \lim_{n\to\infty} x^{(\ell)}\mathbb{E}\left[V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\mu_1^{(\ell)}\right],
\end{aligned}
$$

where we have used the fact that $V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)$ is equal to $|\ \mathtt{T}^{(\ell)}\ |$, the number of edges in $\mathrm{G}_{\mathtt{T}}^{(\ell)}$. To summarize, we have seen that in the case of irregular ensembles, due to the degree distribution correction and the message correction, we can write S^c as

$$
\begin{aligned}
S^c = &- \lim_{n\to\infty} n\Lambda'(1)\mathfrak{E}\left(\mathbb{E}\left[\mu_1^{(\ell)}\tilde{\Delta}\mathsf{a}^{(\ell)}\right]\right)\\
&- \lim_{n\to\infty} n\Lambda'(1)\mathfrak{E}\left(\mathbb{E}\left[\mu_1^{(\ell)}\hat{\Delta}\mathsf{a}^{(\ell)}\right]\right)\\
&- \lim_{n\to\infty} x^{(\ell)}\mathbb{E}\left[V^{(\mathrm{G}_{\mathtt{T}}^{(\ell)})'}(1)\mu_1^{(\ell)}\right].
\end{aligned}
$$

It remains to express each of the three contributions in terms of quantities that can be computed by DE. This indeed can be accomplished. As remarked above, we do not write down the final expression as it is rather long and not too illuminating. But Figure 2 shows some sample evaluations.

Acknowledgment

J.E. and R.U are partially supported by NCCR-MICS, a center supported by the Swiss National Science Foundation under grant number 5005-67322. A.M. is partially supported by a David Filo and Jerry Yang fellowship.

References

[1] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke, *Finite-length scaling for iteratively decoded LDPC ensembles*. Accepted IEEE Trans. on Information Theory.

[2] ———, *Finite-length scaling for iteratively decoded LDPC ensembles*, in Proc. 41th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, 2003.

[3] A. Amraoui, A. Montanari, and R. Urbanke, *Analytic determination of scaling parameters*, in Proc. of the IEEE Int. Symposium on Inform. Theory, Seattle, WA, July 2006.

[4] ———, *How to find good finite-length codes: From art towards science*, Euro. Trans Telecomms., 18 (2007), pp. 491–508.

[5] A. Dembo and A. Montanari, *Finite size scaling for the core of large random hypergraphs*. Xarch:math.PR/0702007, 2007.

[6] J. Ezri, A. Montanari, and R. Urbanke, *Finite-length scaling for Gallager a*, in 44th Allerton Conf. on Communication, Control, and Computing, Monticello, IL, Oct. 2006.

[7] ———, *A generalization of the finite-length scaling approach beyond the BEC*, in Proc. of the IEEE Int. Symposium on Inform. Theory, Nice, France, June 2007.

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

[9] A. Montanari, *Finite-size scaling of good codes*, in Proc. 39th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, 2001.

[10] T. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, 2008.