Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform So Well over the BEC

Shrinivas Kudekar, *Member, IEEE*, Thomas J. Richardson, *Fellow, IEEE*, and Rüdiger L. Urbanke

Dedicated to the memory of Ralf Koetter (1963–2009)

***Abstract*—Convolutional low-density parity-check (LDPC) ensembles, introduced by Felström and Zigangirov, have excellent thresholds and these thresholds are rapidly increasing functions of the average degree. Several variations on the basic theme have been proposed to date, all of which share the good performance characteristics of convolutional LDPC ensembles. We describe the fundamental mechanism that explains why "convolutional-like" or "spatially coupled" codes perform so well. In essence, the spatial coupling of individual codes increases the belief-propagation (BP) threshold of the new ensemble to its maximum possible value, namely the maximum *a posteriori* (MAP) threshold of the underlying ensemble. For this reason, we call this phenomenon "threshold saturation." This gives an entirely new way of approaching capacity. One significant advantage of this construction is that one can create capacity-approaching ensembles with an error correcting radius that is increasing in the blocklength. Although we prove the "threshold saturation" only for a specific ensemble and for the binary erasure channel (BEC), empirically the phenomenon occurs for a wide class of ensembles and channels. More generally, we conjecture that for a large range of graphical systems a similar saturation of the "dynamical" threshold occurs once individual components are coupled sufficiently strongly. This might give rise to improved algorithms and new techniques for analysis.**

***Index Terms*—Belief-propagation (BP) decoder, capacity-achieving codes, convolutional low-density parity-check (LDPC) codes, density evolution (DE), EXIT curves, maximum *a posteriori* (MAP) decoder, protographs.**

I. Introduction

In this paper, we consider the design of capacity-approaching codes based on the connection between the belief-propagation (BP) and maximum *a posteriori* (MAP) threshold of sparse graph codes. Recall that the BP threshold is the threshold of the "locally optimum" BP message-passing algorithm. As such it has low complexity. The MAP threshold, on the other hand, is the threshold of the "globally optimum" decoder. No decoder can do better, but the complexity of the MAP decoder is in general high. The threshold itself is the unique channel parameter so that for channels with lower (better) parameter decoding succeeds with high probability (for large instances) whereas for channels with higher (worse) parameters decoding fails with high probability. Surprisingly, for sparse graph codes there is a connection between these two thresholds; see [1] and [2].[1]

We discuss a fundamental mechanism, which ensures that these two thresholds coincide (or at least are very close). We call this phenomenon "threshold saturation via spatial coupling." A prime example where this mechanism is at work are convolutional low-density parity-check (LDPC) ensembles.

It was Tanner who introduced the method of "unwrapping" a cyclic block code into a convolutional structure [3], [4]. The first *low-density* convolutional ensembles were introduced by Felström and Zigangirov [5]. Convolutional LDPC ensembles are constructed by *coupling* several standard (d_l, d_r)-regular LDPC ensembles together in a chain. Perhaps surprisingly, due to the coupling, and assuming that the chain is finite and properly terminated, the threshold of the resulting ensemble is considerably improved. Indeed, if we start with a $(3,6)$-regular ensemble, then on the binary erasure channel (BEC) the threshold is improved from $\epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6) \approx 0.4294$ to roughly 0.4881 (the capacity for this case is $\frac{1}{2}$). The latter number is the MAP threshold $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ of the underlying $(3,6)$-regular ensemble. This opens up an entirely new way of constructing capacity-approaching ensembles. It is a folk theorem that for standard constructions improvements in the BP threshold go hand in hand with increases in the error floor. More precisely, a large fraction of degree-two variable nodes is typically needed in order to get large thresholds under BP decoding. Unfortunately, the higher the fraction of degree-two variable nodes, the more low-weight codewords (small cycles, small stopping sets, etc.) appear. Under MAP decoding on the other hand these two

Manuscript received April 15, 2010; revised July 21, 2010; accepted August 20, 2010. Date of current version January 19, 2011. The work of S. Kudekar was supported by the Swiss National Science Foundation under Grant 200020-113412.

This paper is part of the special issue on "Facets of Coding Theory: From Algorithms to Networks," dedicated to the scientific legacy of Ralf Koetter.

S. Kudekar was with the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland. He is now with the New Mexico Consortium and CNLS, Los Alamos National Laboratory and the New Mexico Consortium, Los Alamos, NM 87544 USA (e-mail: skudekar@lanl.gov).

T. J. Richardson is with Qualcomm, Flarion Technologies, Bridgewater, NJ 08807 USA (e-mail: tjr@qualcomm.com).

R. Urbanke is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland (e-mail: ruediger.urbanke@epfl.ch).

Communicated by G. D. Forney, Jr., Associate Editor for the special issue on "Facets of Coding Theory: From Algorithms to Networks."

Digital Object Identifier 10.1109/TIT.2010.2095072

[1]There are some trivial instances in which the two thresholds coincide. This is, e.g., the case for so-called "cycle ensembles" or, more generally, for irregular LDPC ensembles that have a large fraction of degree-two variable nodes. In these cases, the reason for this agreement is that for both decoders the performance is dominated by small structures in the graph. But for general ensembles these two thresholds are distinct and, indeed, they can differ significantly.

quantities are positively correlated. To be concrete, if we consider the sequence of $(d_l, 2d_l)$-regular ensembles of rate one-half, by increasing d_l we increase both the MAP threshold as well as the typical minimum distance. It is therefore possible to construct ensembles that have large MAP thresholds *and* low error floors.

The potential of convolutional LDPC codes has long been recognized. Our contribution lies therefore not in the introduction of a new coding scheme, but in clarifying the basic mechanism that make convolutional-like ensembles perform so well.

There is a considerable literature on convolutional-like LDPC ensembles. Variations on the constructions as well as some analysis can be found in [6]–[9]. In [10] and [11], Sridharan *et al.* consider density evolution (DE) for convolutional LDPC ensembles and determine thresholds for the BEC. The equivalent observations for general channels were reported by Lentmaier *et al.* [11], [12]. The preceding two sets of works are perhaps the most pertinent to our setup. By considering the resulting thresholds and comparing them to the thresholds of the underlying ensembles under MAP decoding (see, e.g., [13]) it becomes quickly apparent that an interesting physical effect must be at work. Indeed, in a recent paper [14], Lentmaier and Fettweis followed this route and independently formulated the equality of the BP threshold of convolutional LDPC ensembles and the MAP threshold of the underlying ensemble as a conjecture. They attribute this numerical observation to G. Liva.

A representation of convolutional LDPC ensembles in terms of a protograph was introduced by Mitchell *et al.* [15]. The corresponding representation for terminated convolutional LDPC ensembles was introduced by Lentmaier *et al.* [16]. A pseudocodeword analysis of convolutional LDPC codes was performed by Smarandache *et al.* [17], [18]. In [19], Papaleo *et al.* consider windowed decoding of convolutional LDPC codes on the BEC to study the tradeoff between the decoding latency and the code performance.

In what follows, we will assume that the reader is familiar with basic notions of sparse graph codes and message-passing decoding, and in particular with the asymptotic analysis of LDPC ensembles for transmission over the BEC as it was accomplished in [20]. We summarized the most important facts that are needed for our proof in Section III-A, but this summary is not meant to be a gentle introduction to the topic. Our notation follows for the most part the one in [13].

II. Convolutional-Like LDPC Ensembles

The principle that underlies the good performance of convolutional-like LDPC ensembles is very broad and there are many degrees of freedom in constructing such ensembles. In what follows, we introduce two basic variants. The (d_l, d_r, L)-ensemble is very close to the ensemble discussed in [16]. Experimentally, it has a very good performance. We conjecture that it is capable of achieving capacity.

We also introduce the ensemble (d_l, d_r, L, w). Experimentally, it shows a worse tradeoff between rate, threshold, and blocklength. But it is easier to analyze and we will show that it is capacity achieving. One can think of w as a "smoothing parameter" and we investigate the behavior of this ensemble when w tends to infinity.

Fig. 1. Protograph of a standard $(3, 6)$-regular ensemble.

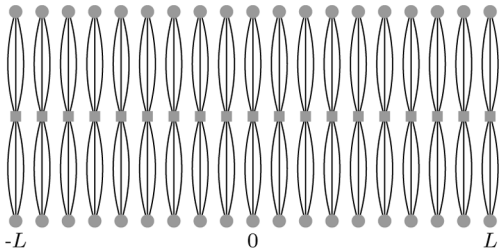

Fig. 2. A chain of $(2L + 1)$ protographs of the standard $(3, 6)$-regular ensembles for $L = 9$. These protographs do not interact.

A. The (d_l, d_r, L) Ensemble

To start, consider a protograph of a standard $(3, 6)$-regular ensemble (see [21] and [22] for the definition of protographs). It is shown in Fig. 1. There are two variable nodes and there is one check node. Let M denote the number of variable nodes at each position. For our example, $M = 100$ means that we have 50 copies of the protograph so that we have 100 variable nodes at each position. For all future discussions, we will consider the regime where M tends to infinity.

Next, consider a collection of $(2L + 1)$ such protographs as shown in Fig. 2. These protographs are noninteracting and so each component behaves just like a standard $(3, 6)$-regular component. In particular, the BP threshold of each protograph is just the standard threshold, call it $\epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6)$ (see Lemma 4 for an analytic characterization of this threshold). Slightly more generally: start with an $(d_l, d_r = kd_l)$-regular ensemble where d_l is odd so that $\hat{d}_l = (d_l - 1)/2 \in \mathbb{N}$.

An interesting phenomenon occurs if we couple these components. To achieve this coupling, connect each protograph to $\hat{d}_l$[2] protographs "to the left" and to $\hat{d}_l$ protographs "to the right." This is shown in Fig. 3 for the two cases $(d_l = 3, d_r = 6)$ and $(d_l = 7, d_r = 14)$. In this figure, $\hat{d}_l$ extra check nodes are added on each side to connect the "overhanging" edges at the boundary.

There are two main effects resulting from this coupling.

1) *Rate reduction*: Recall that the design rate of the underlying standard $(d_l, d_r = kd_l)$-regular ensemble is $1 - \frac{d_l}{d_r} = \frac{k-1}{k}$. Let us determine the design rate of the corresponding $(d_l, d_r = kd_l, L)$ ensemble. By design rate we mean here the rate that we get if we assume that every involved check node imposes a linearly independent constraint.
   The variable nodes are indexed from $-L$ to L so that in total there are $(2L + 1)M$ variable nodes. The check

[2]If we think of this as a convolutional code, then $2\hat{d}_l$ is the syndrome former memory of the code.

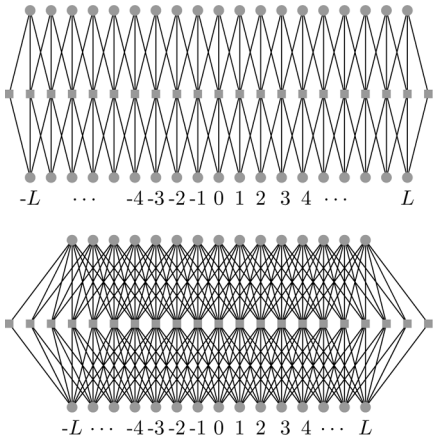

Fig. 3. Two coupled chains of protographs with $L = 9$ and $(d_l = 3, d_r = 6)$ (top) and $L = 7$ and $(d_l = 7, d_r = 14)$ (bottom), respectively.

nodes are indexed from $-(L+\hat{d}_l)$ to $(L+\hat{d}_l)$, so that in total there are $(2(L+\hat{d}_l)+1)M/k$ check nodes. We see that, due to boundary effects, the design rate is reduced to

$$R(d_l, d_r = kd_l, L) = \frac{(2L+1) - (2(L+\hat{d}_l)+1)/k}{2L+1} = \frac{k-1}{k} - \frac{2\hat{d}_l}{k(2L+1)}$$

where the first term on the right-hand side represents the design rate of the underlying standard $(d_l, d_r = kd_l)$-regular ensemble and the second term represents the rate loss. As we see, this rate reduction effect vanishes at a speed $1/L$.

2) *Threshold increase*: The threshold changes dramatically from $\epsilon^{\mathrm{BP}}(d_l, d_r)$ to something close to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ (the MAP threshold of the underlying standard (d_l, d_r)-regular ensemble; see Lemma 4). This phenomenon (which we call "threshold saturation") is much less intuitive and it is the aim of this paper to explain why this happens.

So far we have considered $(d_l, d_r = kd_l)$-regular ensembles. Let us now give a general definition of the (d_l, d_r, L)-ensemble that works for all parameters (d_l, d_r) so that d_l is odd. Rather than starting from a protograph, place variable nodes at positions $[-L, L]$. At each position there are M such variable nodes. Place $\frac{d_l}{d_r}M$ check nodes at each position $[-L-\hat{d}_l, L+\hat{d}_l]$. Connect exactly one of the d_l edges of each variable node at position i to a check node at position $i-\hat{d}_l, \ldots, i+\hat{d}_l$.

Note that at each position $i \in [-L+\hat{d}_l, L-\hat{d}_l]$, there are exactly $M\frac{d_l}{d_r}d_r = Md_l$ check node sockets.[3] Exactly M of those come from variable nodes at each position $i-\hat{d}_l, \ldots, i+\hat{d}_l$. For check nodes at the boundary the number of sockets is decreased linearly according to their position. The probability distribution of the ensemble is defined by choosing a random permutation on the set of all edges for each check node position.

The next lemma, whose proof can be found in Appendix I, asserts that the minimum stopping set distance of most codes in this ensemble is at least a fixed fraction of M. With respect to the technique used in the proof we follow the lead of [15], [18] and [17], [22] that consider distance and pseudodistance analysis of convolutional LDPC ensembles, respectively.

Lemma 1 (Stopping Set Distance of (d_l, d_r, L)-Ensemble): Consider the (d_l, d_r, L)-ensemble with $d_l = 2\hat{d}_l + 1$, $\hat{d}_l \geq 1$, and $d_r \geq d_l$. Define

$$\begin{aligned} p(x) &= \sum_{i\neq 1} \binom{d_r}{i} x^i \\ a(x) &= \frac{\left(\sum_{i\neq 1} \binom{d_r}{i} i x^i\right)}{\left(\sum_{i\neq 1} \binom{d_r}{i} x^i\right)} \\ b(x) &= -(d_l - 1)h_2\left(\frac{a(x)}{d_r}\right) + \frac{d_l}{d_r}\log_2(p(x)) - a(x)\frac{d_l}{d_r}\log_2(x) \\ \omega(x) &= \frac{a(x)}{d_r} \\ h_2(x) &= -x\log_2(x) - (1-x)\log_2(1-x). \end{aligned}$$

Let $\hat{x}$ denote the unique strictly positive solution of the equation $b(x) = 0$ and let $\hat{\omega}(d_l, d_r) = \omega(\hat{x})$. Then, for any $\delta > 0$

$$\lim_{M\to\infty} \mathbb{P}\{d_{\mathrm{ss}}(\mathcal{C})/M < (1-\delta)d_l\hat{\omega}(d_l, d_r)\} = 0$$

where $d_{\mathrm{ss}}(\mathcal{C})$ denotes the minimum stopping set distance of the code $\mathcal{C}$.

Discussion: The quantity $\hat{\omega}(d_l, d_r)$ is the relative weight (normalized to the blocklength) at which the exponent of the expected stopping set distribution of the underlying standard (d_l, d_r)-regular ensemble becomes positive. It is perhaps not too surprising that the same quantity also appears in our context. The lemma asserts that the minimum stopping set distance grows linearly in M. But the stated bound does *not* scale with L. We leave it as an interesting open problem to determine whether this is due to the looseness of our bound or whether our bound indeed reflects the correct behavior.

Example 2 $((d_l = 3, d_r = 6, l))$: An explicit calculation shows that $\hat{x} \approx 0.058$ and $3\hat{\omega}(3,6) \approx 0.056$. Let $n = M(2L+1)$ be the blocklength. If we assume that $2L+1 = M^{\alpha}$, $\alpha \in (0,1)$, then $M = n^{\frac{1}{1+\alpha}}$. Lemma 1 asserts that the minimum stopping set distance grows in the blocklength at least as $0.056n^{\frac{1}{1+\alpha}}$.

B. The (d_l, d_r, L, w) Ensemble

In order to simplify the analysis we modify the ensemble (d_l, d_r, L) by adding a randomization of the edge connections.

[3] *Sockets* are connection points where edges can be attached to a node. For example, if a node has degree 3 then we imagine that it has three sockets. This terminology arises from the so-called *configuration* model of LDPC ensembles. In this model, we imagine that we label all check-node sockets and all variable-node sockets with the set of integers from one to the cardinality of the sockets. To construct then a particular element of the ensemble we pick a permutation on this set uniformly at random from the set of all permutations and connect variable-node sockets to check-node sockets according to this permutation.

For the remainder of this paper we always assume that $d_r \geq d_l$, so that the ensemble has a nontrivial design rate.

We assume that the variable nodes are at positions $[-L, L]$, $L \in \mathbb{N}$. At each position there are M variable nodes, $M \in \mathbb{N}$. Conceptually, we think of the check nodes to be located at all integer positions from $[-\infty, \infty]$. Only some of these positions actually interact with the variable nodes. At each position there are $\frac{d_l}{d_r} M$ check nodes. It remains to describe how the connections are chosen.

Rather than assuming that a variable at position i has exactly one connection to a check node at position $[i - \hat{d}_l, \dots, i + \hat{d}_l]$, we assume that each of the d_l connections of a variable node at position i is uniformly and independently chosen from the range $[i, \dots, i + w - 1]$, where w is a "smoothing" parameter. In the same way, we assume that each of the d_r connections of a check node at position i is independently chosen from the range $[i - w + 1, \dots, i]$. We no longer require that d_l is odd.

More precisely, the ensemble is defined as follows. Consider a variable node at position i. The variable node has d_l outgoing edges. A *type* t is a w-tuple of nonnegative integers, $t = (t_0, t_1, \dots, t_{w-1})$, so that $\sum_{j=0}^{w-1} t_j = d_l$. The operational meaning of t is that the variable node has t_j edges that connect to a check node at position $i + j$. There are $\binom{d_l + w - 1}{w - 1}$ types. Assume that for each variable we order its edges in an arbitrary but fixed order. A *constellation* c is an d_l-tuple, $c = (c_1, \dots, c_{d_l})$ with elements in $[0, w - 1]$. Its operational significance is that if a variable node at position i has constellation c then its kth edge is connected to a check node at position $i + c_k$. Let $\tau(c)$ denote the type of a constellation. Since we want the position of each edge to be chosen independently we impose a uniform distribution on the set of all constellations. This imposes the following distribution on the set of all types. We assign the probability

$$p(t) = \frac{|\{c : \tau(c) = t\}|}{w^{d_l}}.$$

Pick M so that $Mp(t)$ is a natural number for all types t. For each position i pick $Mp(t)$ variables that have their edges assigned according to type t. Further, use a random permutation for each variable, uniformly chosen from the set of all permutations on d_l letters, to map a type to a constellation.

Under this assignment, and ignoring boundary effects, for each check position i, the number of edges that come from variables at position $i - j$, $j \in [0, w - 1]$, is $M \frac{d_l}{w}$. In other words, it is exactly a fraction $\frac{1}{w}$ of the total number $M d_l$ of sockets at position i. At the check nodes, distribute these edges according to a permutation chosen uniformly at random from the set of all permutations on $M d_l$ letters, to the $M \frac{d_l}{d_r}$ check nodes at this position. It is then not very difficult to see that, under this distribution, for each check node each edge is roughly independently chosen to be connected to one of its nearest w "left" neighbors. Here, "roughly independent" means that the corresponding probability deviates at most by a term of order $1/M$ from the desired distribution. As discussed beforehand, we will always consider the limit in which M first tends to infinity and then the number of iterations tends to infinity. Therefore, for any fixed number of rounds of DE the probability model is exactly the independent model described above.

Lemma 3 (Design Rate): The design rate of the ensemble (d_l, d_r, L, w), with $w \leq 2L$, is given by

$$R(d_l, d_r, L, w) = \left(1 - \frac{d_l}{d_r}\right) - \frac{d_l}{d_r} \frac{w + 1 - 2 \sum_{i=0}^{w} \left(\frac{i}{w}\right)^{d_r}}{2L + 1}.$$

Proof: Let V be the number of variable nodes and C be the number of check nodes that are connected to at least one of these variable nodes. Recall that we *define* the design rate as $1 - C/V$.

There are $V = M(2L + 1)$ variables in the graph. The check nodes that have potential connections to variable nodes in the range $[-L, L]$ are indexed from $-L$ to $L + w - 1$. Consider the $M \frac{d_l}{d_r}$ check nodes at position $-L$. Each of the d_r edges of each such check node is chosen independently from the range $[-L - w + 1, -L]$. The probability that such a check node has at least one connection in the range $[-L, L]$ is equal to $1 - \left(\frac{w-1}{w}\right)^{d_r}$. Therefore, the expected number of check nodes at position $-L$ that are connected to the code is equal to $M \frac{d_l}{d_r} \left(1 - \left(\frac{w-1}{w}\right)^{d_r}\right)$. In a similar manner, the expected number of check nodes at position $-L + i$, $i = 0, \dots, w - 1$, that are connected to the code is equal to $M \frac{d_l}{d_r} \left(1 - \left(\frac{w-i-1}{w}\right)^{d_r}\right)$. All check nodes at positions $-L + w, \dots, L - 1$ are connected. Further, by symmetry, check nodes in the range $L, \dots, L + w - 1$ have an identical contribution as check nodes in the range $-L, \dots, -L + w - 1$. Summing up all these contributions, we see that the number of check nodes, which are connected, is equal to

$$C = M \frac{d_l}{d_r} \left[2L - w + 2 \sum_{i=0}^{w} \left(1 - \left(\frac{i}{w}\right)^{d_r}\right)\right]. \qquad \square$$

Discussion: In the above lemma, we have *defined* the design rate as the normalized difference of the number of variable nodes and the number of check nodes that are involved in the ensemble. This leads to a relatively simple expression, which is suitable for our purposes. But in this ensemble there is a nonzero probability that there are two or more degree-one check nodes attached to the same variable node. In this case, some of these degree-one check nodes are redundant and do not impose constraints. This effect only happens for variable nodes close to the boundary. Since we consider the case where L tends to infinity, this slight difference between the "design rate" and the "true rate" does not play a role. We therefore opt for this simple definition. The design rate is a lower bound on the true rate.

C. Other Variants

There are many variations on the theme that show the same qualitative behavior. For real applications these and possibly other variations are vital to achieve the best tradeoffs. Let us give a few select examples.

1) Diminished rate loss: One can start with a *cycle* (as is the case for tailbiting codes) rather than a chain so that some of the extra check nodes, which we add at the boundary, can be used for the termination on both sides. This reduces the rate loss.
2) Irregular and structured ensembles: We can start with irregular or structured ensembles. Arrange a number of graphs next to each other in a horizontal order.

Couple them by connecting neighboring graphs up to some order. Empirically, once the coupling is "strong" enough and spreads out sufficiently, the threshold is "very close" to the MAP threshold of the underlying ensembles. See also [23] for a study of such ensembles.

The main aim of this paper is to explain why coupled LDPC codes perform so well rather than optimizing the ensemble. Therefore, despite the practical importance of these variations, we focus on the ensemble (d_l, d_r, L, w). It is the simplest to analyze.

III. General Principle

As mentioned before, the basic reason why coupled ensembles have such good thresholds is that their BP threshold is very close to the MAP threshold of the underlying ensemble. Therefore, as a starting point, let us review how the BP and the MAP threshold of the underlying ensemble can be characterized. A detailed explanation of the following summary can be found in [13].

A. The Standard (d_l, d_r)-Regular Ensemble: BP Versus MAP

Consider DE of the standard (d_l, d_r)-regular ensemble. More precisely, consider the fixed point (FP) equation

$$x = \epsilon \left(1 - (1-x)^{d_r-1}\right)^{d_l-1} \tag{1}$$

where ϵ is the channel erasure value and x is the average erasure probability flowing from the variable node side to the check node side. Both the BP as well as the MAP threshold of the (d_l, d_r)-regular ensemble can be characterized in terms of solutions (FPs) of this equation.

Lemma 4 (Analytic Characterization of Thresholds): Consider the (d_l, d_r)-regular ensemble. Let $\epsilon^{\mathrm{BP}}(d_l, d_r)$ denote its BP threshold and let $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ denote its MAP threshold. Define

$$p^{\mathrm{BP}}(x) = ((d_l-1)(d_r-1)-1)(1-x)^{d_r-2} - \sum_{i=0}^{d_r-3}(1-x)^i$$

$$p^{\mathrm{MAP}}(x) = x + \frac{1}{d_r}(1-x)^{d_r-1}(d_l + d_l(d_r-1)x - d_r x) - \frac{d_l}{d_r}$$

$$\epsilon(x) = \frac{x}{(1-(1-x)^{d_r-1})^{d_l-1}}.$$

Let x^{BP} be the unique positive solution of the equation $p^{\mathrm{BP}}(x) = 0$ and let x^{MAP} be the unique positive solution of the equation $p^{\mathrm{MAP}}(x) = 0$. Then, $\epsilon^{\mathrm{BP}}(d_l, d_r) = \epsilon(x^{\mathrm{BP}})$ and $\epsilon^{\mathrm{MAP}}(d_l, d_r) = \epsilon(x^{\mathrm{MAP}})$. We remark that above, for ease of notation, we drop the dependence of x^{BP} and x^{MAP} on d_l and d_r.

Example 5 (Thresholds of (3, 6)-Ensemble): Explicit computations show that $\epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6) \approx 0.42944$ and $\epsilon^{\mathrm{MAP}}(d_l = 3, d_r = 6) \approx 0.488151$.

Lemma 6 (Graphical Characterization of Thresholds): The left-hand side of Fig. 4 shows the so-called extended BP (EBP) EXIT curve associated to the $(3, 6)$-regular ensemble. This is the curve given by $\{\epsilon(x), (1-(1-x)^{d_r-1})^{d_l}\}$, $0 \le x \le 1$. For all regular ensembles with $d_l \ge 3$ this curve has a characteristic

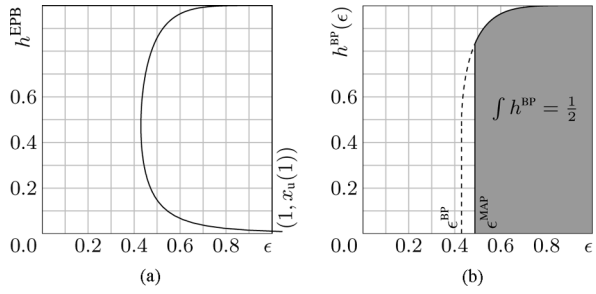

Fig. 4. Left: The EBP EXIT curve h^{EBP} of the $(d_l = 3, d_r = 6)$-regular ensemble. The curve goes "outside the box" at the point $(1, x_u(1))$ and tends to infinity. Right: The BP EXIT function $h^{\mathrm{BP}}(\epsilon)$. Both the BP as well as the MAP threshold are determined by $h^{\mathrm{BP}}(\epsilon)$.

"C" shape. It starts at the point $(1, 1)$ for $x = 1$ and then moves downwards until it "leaves" the unit box at the point $(1, x_u(1))$ and extends to infinity. The right-hand side of Fig. 4 shows the BP EXIT curve (dashed line). It is constructed from the EBP EXIT curve by "cutting off" the lower branch and by completing the upper branch via a vertical line.

The BP threshold $\epsilon^{\mathrm{BP}}(d_l, d_r)$ is the point at which this vertical line hits the x-axis. In other words, the BP threshold $\epsilon^{\mathrm{BP}}(d_l, d_r)$ is equal to the smallest ϵ-value, which is taken on along the EBP EXIT curve.

Lemma 7 (Lower Bound on x^{BP}): For the (d_l, d_r)-regular ensemble

$$x^{\mathrm{BP}}(d_l, d_r) \ge 1 - (d_l - 1)^{-\frac{1}{d_r-2}}.$$

Proof: Consider the polynomial $p^{\mathrm{BP}}(x)$. Note that $p^{\mathrm{BP}}(x) \ge \tilde{p}(x) = ((d_l-1)(d_r-1)-1)(1-x)^{d_r-2} - (d_r-2)$ for $x \in [0, 1]$. Since $p^{\mathrm{BP}}(0) \ge \tilde{p}(0) = (d_l-2)(d_r-1) > 0$, the positive root of $\tilde{p}(x)$ is a lower bound on the positive root of $p^{\mathrm{BP}}(x)$. But the positive root of $\tilde{p}(x)$ is at $1 - \left(\frac{d_r-2}{(d_l-1)(d_r-1)-1}\right)^{\frac{1}{d_r-2}}$. This in turn is lower bounded by $1 - (d_l-1)^{-\frac{1}{d_r-2}}$. □

To construct the MAP threshold $\epsilon^{\mathrm{MAP}}(d_l, d_r)$, integrate the BP EXIT curve starting at $\epsilon = 1$ until the area under this curve is equal to the design rate of the code. The point at which equality is achieved is the MAP threshold (see the right-hand side of Fig. 4).

Lemma 8 (MAP Threshold for Large Degrees): Consider the (d_l, d_r)-regular ensemble. Let $r(d_l, d_r) = 1 - \frac{d_l}{d_r}$ denote the design rate so that $d_r = \frac{d_l}{1-r}$. Then, for r fixed and d_l increasing, the MAP threshold $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ converges exponentially fast (in d_l) to $1 - r$.

Proof: Recall that the MAP threshold is determined by the unique positive solution of the polynomial equation $p^{\mathrm{MAP}}(x) = 0$, where $p^{\mathrm{MAP}}(x)$ is given in Lemma 4. A closer look at this equation shows that this solution has the form

$$x = (1-r)\left(1 - \frac{r^{\frac{d_l}{1-r}-1}(d_l + r - 1)}{1 - r^{\frac{d_l}{1-r}-2}(1 + d_l(d_l + r - 2))} + o\left(d_l r^{\frac{d_l}{1-r}}\right)\right).$$

We see that the root converges exponentially fast (in d_l) to $1-r$. Further, in terms of this root we can write the MAP threshold as

$$x\left(1+\frac{1-r-x}{(d_l+r-1)x}\right)^{d_l-1}. \qquad \square$$

Lemma 9 (Stable and Unstable FPs [13]): Consider the standard (d_l, d_r)-regular ensemble with $d_l \geq 3$. Define

$$h(x) = \epsilon\left(1-(1-x)^{d_r-1}\right)^{d_l-1} - x. \tag{2}$$

Then, for $\epsilon^{\mathrm{BP}}(d_l, d_r) < \epsilon \leq 1$, there are exactly two strictly positive solutions of the equation $h(x) = 0$ and they are both in the range $[0, 1]$.

Let $x_s(\epsilon)$ be the larger of the two and let $x_u(\epsilon)$ be the smaller of the two. Then, $x_s(\epsilon)$ is a strictly increasing function in ϵ and $x_u(\epsilon)$ is a strictly decreasing function in ϵ. Finally, $x_s(\epsilon^{\mathrm{BP}}) = x_u(\epsilon^{\mathrm{BP}})$.

Discussion: Recall that $h(x)$ represents the change of the erasure probability of DE in one iteration, assuming that the system has current erasure probability x. This change can be negative (erasure probability decreases), it can be positive, or it can be zero (i.e., there is an FP). We discuss some useful properties of $h(x)$ in Appendix II.

As the notation indicates, x_s corresponds to a *stable* FP whereas x_u corresponds to an unstable FP. Here stability means that if we initialize DE with the value $x_s(\epsilon) + \delta$ for a sufficiently small δ then DE converges back to $x_s(\epsilon)$.

B. The (d_l, d_r, L) Ensemble

Consider the EBP EXIT curve of the (d_l, d_r, L) ensemble. To compute this curve we proceed as follows. We fix a desired "entropy" value (see Definition 15), call it χ. We initialize DE with the constant χ. We then repeatedly perform one step of DE, where in each step we fix the channel parameter in such a way that the resulting entropy is equal to χ. This is equivalent to the procedure introduced in [24, Sec. VIII], to compute the EBP EXIT curve for general binary-input memoryless output-symmetric channels. Once the procedure has converged, we plot its EXIT value versus the resulting channel parameter. We then repeat the procedure for many different entropy values to produce a whole curve.

Note that DE here is not just DE for the underlying ensemble. Due to the spatial structure we in effect deal with a multiedge ensemble [25] with many edge types. For our current casual discussion the exact form of the DE equations is not important, but if you are curious please fast forward to Section V.

Why do we use this particular procedure? By using forward DE, one can only reach stable FPs. But the above procedure allows one to find points along the whole EBP EXIT curve, i.e., one can in particular also produce *unstable* FPs of DE.

The resulting curve is shown in Fig. 5 for various values of L. Note that these EBP EXIT curves show a dramatically different behavior compared to the EBP EXIT curve of the underlying ensemble. These curves appear to be "to the right" of the threshold $\epsilon^{\mathrm{MAP}}(3, 6) \approx 0.48815$. For small values of L one might be led to believe that this is true since the design rate of such an ensemble is considerably smaller than $1-d_l/d_r$. But even for large values of L, where the rate of the ensemble is close to $1-d_l/d_r$, this dramatic increase in the threshold is still true. Empirically, we see that, for L increasing, the EBP EXIT curve approaches the MAP EXIT curve of the underlying $(d_l = 3, d_r = 6)$-regular ensemble. In particular, for $\epsilon \approx \epsilon^{\mathrm{MAP}}(d_l, d_r)$ the EBP EXIT curve drops essentially vertically until it hits zero. We will see that this is a fundamental property of this construction.

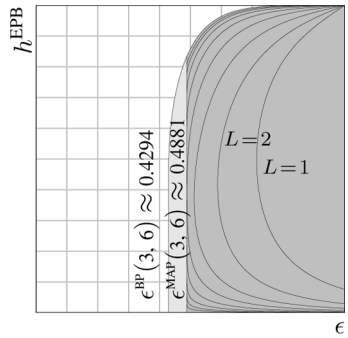

Fig. 5. EBP EXIT curves of the ensemble $(d_l = 3, d_r = 6, L)$ for $L = 1, 2, 4, 8, 16, 32, 64$, and 128. The BP/MAP thresholds are $\epsilon^{\mathrm{BP/MAP}}(3, 6, 1) = 0.714309/0.820987$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 2) = 0.587842/0.668951$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 4) = 0.512034/0.574158$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 8) = 0.488757/0.527014$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 16) = 0.488151/0.505833$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 32) = 0.488151/0.496366$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 64) = 0.488151/0.492001$, $\epsilon^{\mathrm{BP/MAP}}(3, 6, 128) = 0.488151/0.489924$. The light/dark gray areas mark the interior of the BP/MAP EXIT function of the underlying $(3, 6)$-regular ensemble, respectively.

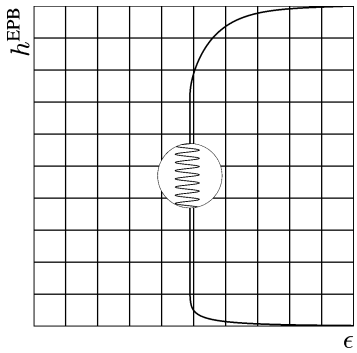

Fig. 6. EBP EXIT curve for the $(d_l = 3, d_r = 6, L = 32)$ ensemble. The circle shows a magnified portion of the curve. The horizontal magnification is 10^7, and the vertical one is 1.

C. Discussion

A look at Fig. 5 might convey the impression that the transition of the EBP EXIT function is completely flat and that the threshold of the ensemble (d_l, d_r, L) is exactly equal to the MAP threshold of the underlying (d_l, d_r)-regular ensemble when L tends to infinity.

Unfortunately, the actual behavior is more subtle. Fig. 6 shows the EBP EXIT curve for $L = 32$ with a small section of the transition greatly magnified. As one can see from this magnification, the curve is not flat but exhibits small "wiggles" in ϵ around $\epsilon^{\mathrm{MAP}}(d_l, d_r)$. These wiggles do not vanish as L tends to infinity but their width remains constant. As we will discuss in much more detail later, area considerations imply that, in the limit as L diverges to infinity, the BP threshold is slightly below $\epsilon^{\mathrm{MAP}}(d_l, d_r)$. Although this does not play a role

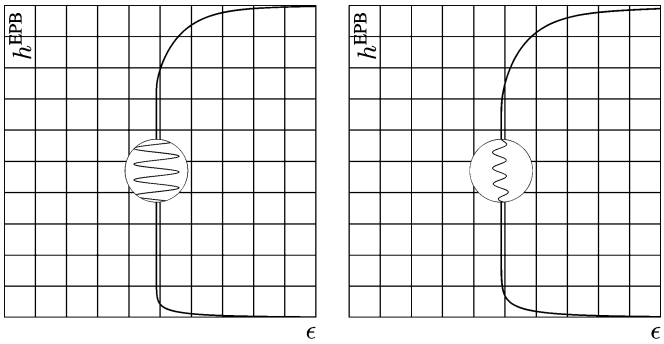

Fig. 7. EBP EXIT curve for the $(d_l = 3, d_r = 6, L = 16, w)$ ensemble. Left: $w = 2$. The circle shows a magnified portion of the curve. The horizontal magnification is 10^3, and the vertical one is 1. Right: $w = 3$. The circle shows a magnified portion of the curve. The horizontal magnification is 10^6, and the vertical one is 1.

in the remainder of the paper, let us remark that the number of wiggles is (up to a small additive constant) equal to L.

Where do these wiggles come from? They stem from the fact that the system is discrete. If, instead of considering a system with sections at integer points, we would deal with a continuous system where neighboring "sections" are infinitesimally close, then these wiggles would vanish. This "discretization" effect is well known in the physics literature. By letting w tend to infinity we can in effect create a continuous system. This is in fact our main motivation for introducing this parameter.

Empirically, these wiggles are very small (e.g., they are of width 10^{-7} for the $(d_l = 3, d_r = 6, L)$ ensemble), and further, these wiggles tend to 0 when d_l is increased. Unfortunately this is hard to prove.

We therefore study the ensemble (d_l, d_r, L, w). The wiggles for this ensemble are in fact larger; see, e.g., Fig. 7. But, as mentioned above, the wiggles can be made arbitrarily small by letting w (the smoothing parameter) tend to infinity. For example, in the left-hand side of Fig. 7, $w = 2$, whereas in the right-hand side, we have $w = 3$. We see that the wiggle size has decreased by more than a factor of 10^3.

IV. Main Statement and Interpretation

As pointed out in the introduction, numerical experiments indicate that there is a large class of convolutional-like LDPC ensembles that all have the property that their BP threshold is "close" to the MAP threshold of the underlying ensemble. Unfortunately, no general theorem is known to date that states when this is the case. The following theorem gives a particular instance of what we believe to be a general principle. The bounds stated in the theorem are loose and can likely be improved considerably. Throughout the paper we assume that $d_l \geq 3$.

A. Main Statement

Theorem 10 (BP Threshold of the (d_l, d_r, L, w) Ensemble): Consider transmission over the $\mathrm{BEC}(\epsilon)$ using random elements from the ensemble (d_l, d_r, L, w). Let $\epsilon^{\mathrm{BP}}(d_l, d_r, L, w)$ denote the BP threshold and let $R(d_l, d_r, L, w)$ denote the design rate of this ensemble.

Then, in the limit as M tends to infinity, and for

$$w > \max\left\{2^{16}, 2^4 d_l^2 d_r^2, \frac{\left(2d_l d_r\left(1+\frac{2d_l}{1-2^{-1/(d_r-2)}}\right)\right)^8}{\left(1-2^{-1/(d_r-2)}\right)^{16}\left(\frac{1}{2}\left(1-\frac{d_l}{d_r}\right)\right)^8}\right\}$$

we have

$$\begin{aligned}\epsilon^{\mathrm{BP}}(d_l, d_r, L, w) &\leq \epsilon^{\mathrm{MAP}}(d_l, d_r, L, w)\\ &\leq \epsilon^{\mathrm{MAP}}(d_l, d_r)\\ &\quad + \frac{w-1}{2L\left(1-(1-x^{\mathrm{MAP}}(d_l, d_r))^{d_r-1}\right)^{d_l}}\end{aligned} \tag{3}$$

$$\begin{aligned}\epsilon^{\mathrm{BP}}(d_l, d_r, L, w) \geq \Bigg(&\epsilon^{\mathrm{MAP}}(d_l, d_r)\\ &- w^{-\frac{1}{8}}\frac{8d_l d_r + \frac{4d_r d_l^2}{\left(1-4w^{-\frac{1}{8}}\right)^{d_r}}}{\left(1-2^{-\frac{1}{d_r}}\right)^2}\Bigg)\\ &\times (1-4w^{-1/8})^{d_r d_l}.\end{aligned} \tag{4}$$

In the limit as M, L, and w (in that order) tend to infinity

$$\begin{aligned}&\lim_{w\to\infty}\lim_{L\to\infty} R(d_l, d_r, L, w)\\ &\quad = 1 - \frac{d_l}{d_r}\end{aligned} \tag{5}$$

$$\begin{aligned}&\lim_{w\to\infty}\lim_{L\to\infty} \epsilon^{\mathrm{BP}}(d_l, d_r, L, w)\\ &\quad = \lim_{w\to\infty}\lim_{L\to\infty} \epsilon^{\mathrm{MAP}}(d_l, d_r, L, w)\\ &\quad = \epsilon^{\mathrm{MAP}}(d_l, d_r).\end{aligned} \tag{6}$$

Discussion:

1) The lower bound on $\epsilon^{\mathrm{BP}}(d_l, d_r, L, w)$ is the main result of this paper. It shows that, up to a term, which tends to zero when w tends to infinity, the threshold of the chain is equal to the MAP threshold of the underlying ensemble. The statement in the theorem is weak. As we discussed earlier, the convergence speed with respect to (w.r.t.) w is most likely exponential. We prove only a convergence speed of $w^{-\frac{1}{8}}$. We pose it as an open problem to improve this bound. We also remark that, as seen in (6), the MAP threshold of the (d_l, d_r, L, w) ensemble tends to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ for any finite w when L tends to infinity, whereas the BP threshold is bounded away from $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ for any finite w.
2) We right away prove the upper bound on $\epsilon^{\mathrm{BP}}(d_l, d_r, L, w)$. For the purpose of our proof, we first consider a "circular" ensemble. This ensemble is defined in an identical manner as the (d_l, d_r, L, w) ensemble except that the positions are now from 0 to $K-1$ and index arithmetic is performed modulo K. This circular ensemble has design rate equal to $1 - d_l/d_r$. Set $K = 2L + w$. The original ensemble is recovered by setting any consecutive $w-1$ positions to zero. We first provide a lower bound on the conditional

entropy for the circular ensemble when transmitting over a BEC with parameter ϵ. We then show that setting $w-1$ sections to 0, does not significantly decrease this entropy. Overall this gives an upper bound on the MAP threshold of the original ensemble.

It is not hard to see that the BP EXIT curve[4] is the same for both the (d_l, d_r)-regular ensemble and the circular ensemble. Indeed, the forward DE (see Definition 13) converges to the same FP for both ensembles. Consider the (d_l, d_r)-regular ensemble and let $\epsilon \in \left[\epsilon^{\mathrm{MAP}}(d_l, d_r), 1\right]$. The conditional entropy when transmitting over a BEC with parameter ϵ is at least equal to $1 - d_l/d_r$ minus the area under the BP EXIT curve between $[\epsilon, 1]$ (see [13, Th. 3.120]). Call this area $A(\epsilon)$. Here, the entropy is normalized by KM, where K is the length of the circular ensemble and M denotes the number of variable nodes per section. Assume now that we set $w-1$ consecutive sections of the circular ensemble to 0 in order to recover the original ensemble. As a consequence, we "remove" an entropy (degrees of freedom) of at most $(w-1)/K$ from the circular system. The remaining entropy is therefore positive (and hence we are above the MAP threshold of the circular ensemble) as long as $1 - d_l/d_r - (w-1)/K - A(\epsilon) > 0$. Thus, the MAP threshold of the circular ensemble is given by the supremum over all ϵ such that $1 - d_l/d_r - (w-1)/K - A(\epsilon) \leq 0$. Now note that $A(\epsilon^{\mathrm{MAP}}(d_l, d_r)) = 1 - d_l/d_r$, so that the above condition becomes $A(\epsilon^{\mathrm{MAP}}(d_l, d_r)) - A(\epsilon) \leq (w-1)/K$. But the BP EXIT curve is an increasing function in ϵ so that $A(\epsilon^{\mathrm{MAP}}(d_l, d_r)) - A(\epsilon) > (\epsilon - \epsilon^{\mathrm{MAP}}(d_l, d_r))(1 - (1 - x^{\mathrm{MAP}}(d_l, d_r))^{d_r - 1})^{d_l}$. We get the stated upper bound on $\epsilon^{\mathrm{MAP}}(d_l, d_r, L, w)$ by lower bounding K by $2L$.

3) According to Lemma 3, $\lim_{L\to\infty} \lim_{M\to\infty} R(d_l, d_r, L, w) = 1 - \frac{d_l}{d_r}$. This immediately implies the limit (5). The limit for the BP threshold $\epsilon^{\mathrm{BP}}(d_l, d_r, L, w)$ follows from (4).
4) According to Lemma 8, the MAP threshold $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ of the underlying ensemble quickly approaches the Shannon limit. We therefore see that convolutional-like ensembles provide a way of approaching capacity with low complexity. For example, for a rate equal to one-half, we get $\epsilon^{\mathrm{MAP}}(d_l = 3, d_r = 6) = 0.48815$, $\epsilon^{\mathrm{MAP}}(d_l = 4, d_r = 8) = 0.49774$, $\epsilon^{\mathrm{MAP}}(d_l = 5, d_r = 10) = 0.499486$, $\epsilon^{\mathrm{MAP}}(d_l = 6, d_r = 12) = 0.499876$, $\epsilon^{\mathrm{MAP}}(d_l = 7, d_r = 14) = 0.499969$.

B. Proof Outline

The proof of the lower bound in Theorem 10 is long. We therefore break it up into several steps. Let us start by discussing each of the steps separately. This hopefully clarifies the main ideas. But it will also be useful later when we discuss how the main statement can potentially be generalized. We will see that some steps are quite generic, whereas other steps require a rather detailed analysis of the particular chosen system.

[4]The BP EXIT curve is the plot of the extrinsic estimate of the BP decoder versus the channel erasure fraction (see [13] for details).

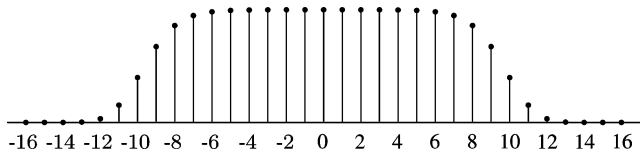

Fig. 8. Unimodal FP of the $(d_l = 3, d_r = 6, L = 16, w = 3)$ ensemble with small values towards the boundary, a fast transition, and essentially constant values in the middle.

1) Existence of FP: "The" key to the proof is to show the existence of a unimodal FP $(\epsilon^*, \underline{x}^*)$, which takes on an essentially constant value in the "middle," has a fast "transition," and has arbitrarily small values towards the boundary (see Definition 12). Fig. 8 shows a typical such example. We will see later that the associated channel parameter of such an FP, ϵ^*, is necessarily very close to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$.
2) Construction of EXIT curve: Once we have established the existence of such a special FP we construct from it a whole *FP family*. The elements in this family of FPs look essentially identical. They differ only in their "width." This width changes continuously, initially being equal to roughly $2L+1$ until it reaches zero. As we will see, this family "explains" how the overall constellation (see Definition 12) collapses once the channel parameter has reached a value close to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$: starting from the two boundaries, the whole constellation "moves in" like a wave until the two wave ends meet in the middle. The EBP EXIT curve is a projection of this wave (by computing the EXIT value of each member of the family). If we look at the EBP EXIT curve, this phenomenon corresponds to the very steep vertical transition close to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$.

   Where do the wiggles in the EBP EXIT curve come from? Although the various FPs look "almost" identical (other than the place of the transition) they are not exactly identical. The ϵ value changes very slightly (around ϵ^*). The larger we choose w the smaller we can make the changes (at the cost of a longer transition).

   When we construct the above family of FPs it is mathematically convenient to allow the channel parameter ϵ to depend on the position. Let us describe this in more detail.

   We start with a special FP as depicted in Fig. 8. From this we construct a smooth family $(\underline{\epsilon}(\alpha), \underline{x}(\alpha))$, parameterized by α, $\alpha \in [0,1]$, where $\underline{x}(1) = \underline{1}$ and where $\underline{x}(0) = \underline{0}$. The components of the vector $\underline{\epsilon}(\alpha)$ are essentially constants (for α fixed). The possible exceptions are components towards the boundary. We allow those components to take on larger (than in the middle) values.

   From the family $(\underline{\epsilon}(\alpha), \underline{x}(\alpha))$ we derive an EBP EXIT curve and we then measure the area enclosed by this curve. We will see that this area is close to the design rate. From this we will be able to conclude that $\epsilon^* \approx \epsilon^{\mathrm{MAP}}(d_l, d_r)$.

3) Operational meaning of EXIT curve: We next show that the EBP EXIT curve constructed in step 2) has an operational meaning. More precisely, we show that if we pick a channel parameter sufficiently below ϵ^* then forward DE converges to the trivial FP.
4) Putting it all together: The final step is to combine all the constructions and bounds discussed in the previous steps to show that $\epsilon^{\mathrm{BP}}(d_l, d_r, w, L)$ converges to $\epsilon^{\mathrm{MAP}}(d_l, d_r)$ when w and L tend to infinity.

V. Proof of Theorem 10

This section contains the technical details of Theorem 10. We accomplish the proof by following the steps outlined in the previous section. To enhance the readability of this section we have moved some of the long proofs to Appendixes I–V.

A. Step 1): Existence of FP

Definition 11 (DE of (d_l, d_r, L, w) Ensemble): Let x_i, $i \in \mathbb{Z}$, denote the average erasure probability, which is emitted by variable nodes at position i. For $i \notin [-L, L]$ we set $x_i = 0$. For $i \in [-L, L]$ the FP condition implied by DE is

$$x_i = \epsilon \left(1 - \frac{1}{w} \sum_{j=0}^{w-1} \left(1 - \frac{1}{w} \sum_{k=0}^{w-1} x_{i+j-k} \right)^{d_r - 1} \right)^{d_l - 1}. \quad (7)$$

If we define

$$f_i = \left(1 - \frac{1}{w} \sum_{k=0}^{w-1} x_{i-k} \right)^{d_r - 1} \quad (8)$$

then (7) can be rewritten as

$$x_i = \epsilon \left(1 - \frac{1}{w} \sum_{j=0}^{w-1} f_{i+j} \right)^{d_l - 1}.$$

In what follows, it will be handy to have an even shorter form for the right-hand side of (7). Therefore, let

$$g(x_{i-w+1}, \ldots, x_{i+w-1}) = \left(1 - \frac{1}{w} \sum_{j=0}^{w-1} f_{i+j} \right)^{d_l - 1}. \quad (9)$$

Note that

$$g(x, \ldots, x) = \left(1 - (1 - x)^{d_r - 1} \right)^{d_l - 1}$$

where the right-hand side represents DE for the underlying (d_l, d_r)-regular ensemble.

The function $f_i(x_{i-w+1}, \ldots, x_i)$ defined in (8) is decreasing in all its arguments $x_j \in [0, 1]$, $j = i - w + 1, \ldots, i$. In what follows, it is understood that $x_i \in [0, 1]$. The channel parameter ϵ is allowed to take values in $\mathbb{R}^+$.

Definition 12 (FPs of DE): Consider DE for the (d_l, d_r, L, w) ensemble. Let $\underline{x} = (x_{-L}, \ldots, x_L)$. We call $\underline{x}$ the *constellation*. We say that $\underline{x}$ forms an FP of DE with parameter ϵ if $\underline{x}$ fulfills (7) for $i \in [-L, L]$. As a short hand we then say that $(\epsilon, \underline{x})$ is an FP. We say that $(\epsilon, \underline{x})$ is a *non-trivial* FP if $\underline{x}$ is not identically zero. More generally, let

$$\underline{\epsilon} = (\epsilon_{-L}, \ldots, \epsilon_0, \ldots, \epsilon_L)$$

where $\epsilon \in \mathbb{R}^+$ for $i \in [-L, L]$. We say that $(\underline{\epsilon}, \underline{x})$ forms an FP if

$$x_i = \epsilon_i g(x_{i-w+1}, \ldots, x_{i+w-1}), \quad i \in [-L, L]. \quad (10)$$

□

Definition 13 (Forward DE and Admissible Schedules): Consider DE for the (d_l, d_r, L, w) ensemble. More precisely, pick a parameter $\epsilon \in [0, 1]$. Initialize $\underline{x}^{(0)} = (1, \ldots, 1)$. Let $\underline{x}^{(\ell)}$ be the result of ℓ rounds of DE. For example, $\underline{x}^{(\ell+1)}$ is generated from $\underline{x}^{(\ell)}$ by applying the DE (7) to each section $i \in [-L, L]$

$$x_i^{(\ell+1)} = \epsilon g \left(x_{i-w+1}^{(\ell)}, \ldots, x_{i+w-1}^{(\ell)} \right).$$

We call this the *parallel* schedule.

More generally, consider a schedule in which in each step ℓ an arbitrary subset of the sections is updated, constrained only by the fact that every section is updated in infinitely many steps. We call such a schedule *admissible*. Again, we call $\underline{x}^{(\ell)}$ the resulting sequence of constellations.

In what follows, we will refer to this procedure as *forward* DE by which we mean the appropriate *initialization* and the subsequent DE procedure. For example, in the next lemma we will discuss the FPs, which are reached under forward DE. These FPs have special properties and so it will be convenient to be able to refer to them in a succinct way and to be able to distinguish them from general FPs of DE.

Lemma 14 (FPs of Forward DE): Consider forward DE for the (d_l, d_r, L, w) ensemble. Let $\underline{x}^{(\ell)}$ denote the sequence of constellations under an admissible schedule. Then $\underline{x}^{(\ell)}$ converges to an FP of DE and this FP is independent of the schedule. In particular, it is equal to the FP of the parallel schedule.

Proof: Consider first the parallel schedule. We claim that the vectors $\underline{x}^{(\ell)}$ are ordered, i.e., $\underline{x}^{(0)} \geq \underline{x}^{(1)} \geq \cdots \geq \underline{0}$ (the ordering is pointwise). This is true since $\underline{x}^{(0)} = (1, \ldots, 1)$, whereas $\underline{x}^{(1)} \leq (\epsilon, \ldots, \epsilon) \leq (1, \ldots, 1) = \underline{x}^{(0)}$. It now follows by induction on the number of iterations that the sequence $\underline{x}^{(\ell)}$ is monotonically decreasing.

Since the sequence $\underline{x}^{(\ell)}$ is also bounded from below it converges. Call the limit $\underline{x}^{(\infty)}$. Since the DE equations are continuous it follows that $\underline{x}^{(\infty)}$ is an FP of DE (7) with parameter ϵ. We call $\underline{x}^{(\infty)}$ the forward FP of DE.

That the limit (exists in general and that it) does not depend on the schedule follows by standard arguments and we will be brief. The idea is that for any two admissible schedules the corresponding computation trees are nested. This means that if we look at the computation graph of schedule let us say one at time ℓ then there exists a time ℓ' so that the computation graph under schedule 2 is a superset of the first computation graph. To be able to come to this conclusion we have crucially used the fact that for an admissible schedule every section is updated infinitely often. This shows that the performance under schedule 2 is at least as good as the performance under schedule 1. The converse claim, and hence equality, follows by symmetry. □

Definition 15 (Entropy): Let $\underline{x}$ be a constellation. We define the (normalized) *entropy* of $\underline{x}$ to be

$$\chi(\underline{x}) = \frac{1}{2L+1} \sum_{i=-L}^{L} x_i.$$

Discussion: More precisely, we should call $\chi(\underline{x})$ the average *message* entropy. But we will stick with the shorthand entropy in the remainder of the paper.

Lemma 16 (Nontrivial FPs of Forward DE): Consider the ensemble (d_l, d_r, L, w). Let $\underline{x}$ be the FP of forward DE for the parameter ϵ. For $\epsilon \in (\ \frac{d_l}{d_r}, 1\ ]$ and $\chi \in \left[0, \epsilon^{\frac{1}{d_l-1}} \left(\epsilon - \frac{d_l}{d_r}\right)\right)$, if

$$L \geq \frac{w}{2\left(\frac{d_r}{d_l}\left(\epsilon - \chi\epsilon^{-\frac{1}{d_l-1}}\right) - 1\right)} \tag{11}$$

then $\chi(\underline{x}) \geq \chi$.

Proof: Let $R(d_l, d_r, L, w)$ be the design rate of the (d_l, d_r, L, w) ensemble as stated in Lemma 3. Note that the design rate is a lower bound on the actual rate. It follows that the system has at least $(2L+1)R(d_l, d_r, L, w)M$ degrees of freedom. If we transmit over a channel with parameter ϵ then in expectation at most $(2L+1)(1-\epsilon)M$ of these degrees of freedom are resolved. Recall that we are considering the limit in which M diverges to infinity. Therefore, we can work with averages and do not need to worry about the variation of the quantities under consideration. It follows that the number of degrees of freedom left unresolved, measured per position and normalized by M, is at least $(R(d_l, d_r, L, w) - 1 + \epsilon)$.

Let $\underline{x}$ be the forward DE FP corresponding to parameter ϵ. Recall that x_i is the average message, which flows from a variable at position i towards the check nodes. From this we can compute the corresponding probability that the node value at position i has not been recovered. It is equal to $\epsilon \left(\frac{x_i}{\epsilon}\right)^{\frac{d_l}{d_l-1}} = \epsilon^{-\frac{1}{d_l-1}} x_i^{\frac{d_l}{d_l-1}}$. Clearly, the BP decoder cannot be better than the MAP decoder. Further, the MAP decoder cannot resolve the unknown degrees of freedom. It follows that we must have

$$\epsilon^{-\frac{1}{d_l-1}} \frac{1}{2L+1} \sum_{i=-L}^{L} x_i^{\frac{d_l}{d_l-1}} \geq R(d_l, d_r, L, w) - 1 + \epsilon.$$

Note that $x_i \in [0,1]$ so that $x_i \geq x_i^{\frac{d_l}{d_l-1}}$. We conclude that

$$\chi(\underline{x}) = \frac{1}{2L+1} \sum_{i=-L}^{L} x_i \geq \epsilon^{\frac{1}{d_l-1}} (R(d_l, d_r, L, w) - 1 + \epsilon).$$

Assume that we want a constellation with entropy at least χ. Using the expression for $R(d_l, d_r, L, w)$ from Lemma 3, this leads to the inequality

$$\epsilon^{\frac{1}{d_l-1}} \left(-\frac{d_l}{d_r} - \frac{d_l}{d_r} \frac{w + 1 - 2\sum_{i=0}^{w} \left(\frac{i}{w}\right)^{d_r}}{2L+1} + \epsilon \right) \geq \chi. \tag{12}$$

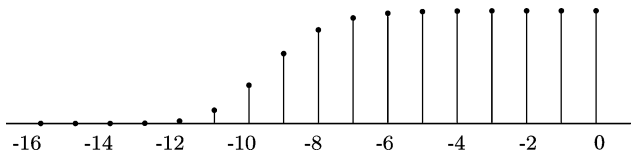

Fig. 9. A proper one-sided FP $(\epsilon, \underline{x})$ for the ensemble $(d_l = 3, d_r = 6, L = 16, w = 3)$, where $\epsilon = 0.488151$. As we will discuss in Lemma 23, for sufficiently large L, the maximum value of $\underline{x}$, namely x_0, approaches the stable value $x_s(\epsilon)$. Further, as discussed in Lemma 26, the width of the transition is of order $O\left(\frac{w}{\delta}\right)$, where $\delta > 0$ is a parameter that indicates, which elements of the constellation we want to include in the transition.

Solving for L and simplifying the inequality by upper bounding $1 - 2\sum_{i=0}^{w} \left(\frac{i}{w}\right)^{d_r}$ by 0 and lower bounding $2L+1$ by $2L$ leads to (11). □

Not all FPs can be constructed by forward DE. In particular, one can only reach (marginally) "stable" FPs by the above procedure. Recall from Section IV-B, step 1), that we want to construct an unimodal FP, which "explains" how the constellation collapses. Such an FP is by its very nature unstable.

It is difficult to prove the existence of such an FP by direct methods. We therefore proceed in stages. We first show the existence of a "one-sided" increasing FP. We then construct the desired unimodal FP by taking two copies of the one-sided FP, flipping one copy, and gluing these FPs together.

Definition 17 (One-Sided DE): Consider the tuple $\underline{x} = (x_{-L}, \ldots, x_0)$. The FP condition implied by *one-sided* DE is equal to (7) with $x_i = 0$ for $i < -L$ and $x_i = x_0$ for $i > 0$.

Definition 18 (FPs of One-Sided DE): We say that $\underline{x}$ is a one-sided FP (of DE) with parameter ϵ and length L if (7) is fulfilled for $i \in [-L, 0]$, with $x_i = 0$ for $i < -L$ and $x_i = x_0$ for $i > 0$.

In the same manner as we have done this for two-sided FPs, if $\underline{\epsilon} = (\epsilon_{-L}, \ldots, \epsilon_0)$, then we define one-sided FPs with respect to $\underline{\epsilon}$.

We say that $\underline{x}$ is *nondecreasing* if $x_i \leq x_{i+1}$ for $i = -L, \ldots, 0$.

Definition 19 (Entropy): Let $\underline{x}$ be a one-sided FP. We define the (normalized) *entropy* of $\underline{x}$ to be

$$\chi(\underline{x}) = \frac{1}{L+1} \sum_{i=-L}^{0} x_i.$$

Definition 20 (Proper One-Sided FPs): Let $(\epsilon, \underline{x})$ be a *nontrivial* and *nondecreasing* one-sided FP. As a short hand, we then say that $(\epsilon, \underline{x})$ is a proper one-sided FP.

A proper one-sided FP is shown in Fig. 9.

Definition 21 (One-Sided Forward DE and Schedules): Similar to Definition 13, one can define the one-sided forward DE by initializing all sections with 1 and by applying DE according to an admissible schedule.

Lemma 22 (FPs of One-Sided Forward DE): Consider an (d_l, d_r, L, w) ensemble and let $\epsilon \in [0,1]$. Let $\underline{x}^{(0)} = (1, \ldots, 1)$ and let $\underline{x}^{(\ell)}$ denote the result of applying ℓ steps of one-sided

forward DE according to an admissible schedule (cf., Definition 21). Then, we have the following.

1) $\underline{x}^{(\ell)}$ converges to a limit, which is an FP of one-sided DE. This limit is independent of the schedule and the limit is either proper or trivial. As a short hand we say that $(\epsilon, \underline{x})$ is a one-sided FP of forward DE.
2) For $\epsilon \in (\frac{d_l}{d_r}, 1]$ and $\chi \in \left[0, \epsilon^{\frac{1}{d_l-1}}\left(\epsilon - \frac{d_l}{d_r}\right)\right)$, if L fulfills (11) then $\chi(\underline{x}) \geq \chi$.

Proof: The existence of the FP and the independence of the schedule follows along the same line as the equivalent statement for two-sided FPs in Lemma 14. We hence skip the details. Assume that this limit $\underline{x}^{(\infty)}$ is nontrivial. We want to show that it is proper. This means we want to show that it is nondecreasing. We use induction. The initial constellation is nondecreasing. Let us now show that this property stays preserved in each step of DE if we apply a parallel schedule. More precisely, for any section $i \in [-L, 0]$

$$\begin{aligned} x_i^{(\ell+1)} &= \epsilon g\left(x_{i-w+1}^{(\ell)}, \ldots, x_{i+w-1}^{(\ell)}\right) \\ &\overset{(a)}{\leq} \epsilon g\left(x_{i+1-w+1}^{(\ell)}, \ldots, x_{i+1+w-1}^{(\ell)}\right) \\ &= x_{i+1}^{(\ell+1)} \end{aligned}$$

where (a) follows from the monotonicity of $g(\ldots)$ and the induction hypothesis that $\underline{x}^{(\ell)}$ is nondecreasing.

Let us now show that for $\epsilon \in \left(\frac{d_l}{d_r}, 1\right]$ and $\chi \in \left[0, \epsilon^{\frac{1}{d_l-1}}\left(\epsilon - \frac{d_l}{d_r}\right)\right)$, if L fulfills (11) then $\chi(\underline{x}) \geq \chi$. First, recall from Lemma 16 that the corresponding two-sided FP of forward DE has entropy at least χ under the stated conditions. Now compare one-sided and two-sided DE for the same initialization with the constant value 1 and the parallel schedule. We claim that for any step the values of the one-sided constellation at position i, $i \in [-L, 0]$, are larger than or equal to the values of the two-sided constellation at the same position i. To see this we use induction. The claim is trivially true for the initialization. Assume therefore that the claim is true at a particular iteration ℓ. For all points $i \in [-L, -w+1]$ it is then trivially also true in iteration $\ell+1$, using the monotonicity of the DE map. For points $i \in [-w+2, 0]$, recall that the one sided DE "sees" the value x_0 for all positions x_i, $i \geq 0$, and that x_0 is the largest of all x-values. For the two-sided DE on the other hand, by symmetry, $x_i = x_{-i} \leq x_0$ for all $i \geq 0$. Again by monotonicity, we see that the desired conclusion holds.

To conclude the proof, note that if for a unimodal two-sided constellation we compute the average over the positions $[-L, 0]$ then we get at least as large a number as if we compute it over the whole length $[-L, L]$. This follows since the value at position 0 is maximal. □

Let us establish some basic properties of proper one-sided FPs.

Lemma 23 (Maximum of FP): Let $(\epsilon, \underline{x})$, $0 \leq \epsilon \leq 1$, be a proper one-sided FP of length L. Then, $\epsilon > \epsilon^{\mathrm{BP}}(d_l, d_r)$ and

$$x_u(\epsilon) \leq x_0 \leq x_s(\epsilon)$$

where $x_s(\epsilon)$ and $x_u(\epsilon)$ denote the stable and unstable nonzero FP associated to ϵ, respectively.

Proof: We start by proving that $\epsilon \geq \epsilon^{\mathrm{BP}}(d_l, d_r)$. Assume to the contrary that $\epsilon < \epsilon^{\mathrm{BP}}(d_l, d_r)$. Then

$$x_0 = \epsilon g(x_{-w+1}, \ldots, x_{w-1}) \leq \epsilon g(x_0, \ldots, x_0) < x_0$$

a contradiction. Here, the last step follows since $\epsilon < \epsilon^{\mathrm{BP}}(d_l, d_r)$ and $0 < x_0 \leq 1$.

Let us now consider the claim that $x_u(\epsilon) \leq x_0 \leq x_s(\epsilon)$. The proof follows along a similar line of arguments. Since $\epsilon^{\mathrm{BP}}(d_l, d_r) \leq \epsilon \leq 1$, both $x_s(\epsilon)$ and $x_u(\epsilon)$ exist and are strictly positive. Suppose that $x_0 > x_s(\epsilon)$ or that $x_0 < x_u(\epsilon)$. Then

$$x_0 = \epsilon g(x_{-w+1}, \ldots, x_{w-1}) \leq \epsilon g(x_0, \ldots, x_0) < x_0$$

a contradiction.

A slightly more careful analysis shows that $\epsilon \neq \epsilon^{\mathrm{BP}}$, so that in fact we have strict inequality, namely $\epsilon > \epsilon^{\mathrm{BP}}(d_l, d_r)$. We skip the details. □

Lemma 24 (Basic Bounds on FP): Let $(\epsilon, \underline{x})$ be a proper one-sided FP of length L. Then, for all $i \in [-L, 0]$

$$\text{i) } x_i \leq \epsilon \left(1 - \left(1 - \frac{1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}\right)^{d_r-1}\right)^{d_l-1}$$

$$\text{ii) } x_i \leq \epsilon \left(\frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}\right)^{d_l-1}$$

$$\text{iii) } x_i \geq \epsilon \left(\frac{1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}\right)^{d_l-1}$$

$$\text{iv) } x_i \geq \epsilon \left(\left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+w-1-k}\right)^{d_r-2}\frac{d_r-1}{w^2} \times \sum_{j,k=0}^{w-1} x_{i+j-k}\right)^{d_l-1}$$

Proof: We have

$$x_i = \epsilon\left(1 - \frac{1}{w}\sum_{j=0}^{w-1}\left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}\right)^{d_r-1}\right)^{d_l-1}.$$

Let $\mathfrak{f}(x) = (1-x)^{d_r-1}$, $x \in [0,1]$. Since $\mathfrak{f}''(x) = (d_r-1)(d_r-2)(1-x)^{d_r-3} \geq 0$, $\mathfrak{f}(x)$ is convex. Let $y_j = \frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}$. We have

$$\frac{1}{w}\sum_{j=0}^{w-1}\left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}\right)^{d_r-1} = \frac{1}{w}\sum_{j=0}^{w-1}\mathfrak{f}(y_j).$$

Since $\mathfrak{f}(x)$ is convex, using Jensen's inequality, we obtain

$$\frac{1}{w}\sum_{j=0}^{w-1}\mathfrak{f}(y_j) \geq \mathfrak{f}\left(\frac{1}{w}\sum_{j=0}^{w-1} y_j\right)$$

which proves claim i).

The derivation of the remaining inequalities is based on the following identity:

$$1 - B^{d_r - 1} = (1 - B)\left(1 + B + \cdots + B^{d_r - 2}\right). \tag{13}$$

For $0 \leq B \leq 1$ this gives rise to the following inequalities:

$$1 - B^{d_r-1} \geq (d_r - 1)B^{d_r-2}(1 - B) \tag{14}$$
$$1 - B^{d_r-1} \geq (1 - B) \tag{15}$$
$$1 - B^{d_r-1} \leq (d_r - 1)(1 - B). \tag{16}$$

Let $B_j = 1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}$, so that $1 - f_{i+j} = 1 - B_j^{d_r-1}$ [recall the definition of f_{i+j} from (8)]. Using (15), this proves iii)

$$\begin{aligned} x_i &= \epsilon \left(\frac{1}{w}\sum_{j=0}^{w-1}(1 - f_{i+j})\right)^{d_l-1} \\ &\geq \epsilon \left(\frac{1}{w}\sum_{j=0}^{w-1}(1 - B_j)\right)^{d_l-1} \\ &= \epsilon \left(\frac{1}{w}\sum_{j=0}^{w-1}\frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}\right)^{d_l-1}. \end{aligned}$$

If we use (16) instead then we get ii). To prove iv), we use (14)

$$\begin{aligned} x_i &\geq \epsilon \left(\frac{d_r - 1}{w}\sum_{j=0}^{w-1}(1 - B_j)B_j^{d_r-2}\right)^{d_l-1} \\ &= \epsilon \left(\frac{d_r - 1}{w}\sum_{j=0}^{w-1}\left(\frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}\right)\right. \\ &\qquad \left.\times \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+j-k}\right)^{d_r-2}\right)^{d_l-1}. \end{aligned}$$

Since $\underline{x}$ is increasing, $\sum_{k=0}^{w-1} x_{i+j-k} \leq \sum_{k=0}^{w-1} x_{i+w-1-k}$. Hence

$$x_i \geq \epsilon \left(\left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+w-1-k}\right)^{d_r-2}\frac{d_r - 1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}\right)^{d_l-1}.$$

□

Lemma 25 (Spacing of FP): Let $(\epsilon, \underline{x})$, $\epsilon \geq 0$, be a proper one-sided FP of length L. Then for $i \in [-L+1, 0]$

$$\begin{aligned} x_i - x_{i-1} &\leq \epsilon \frac{(d_l - 1)(d_r - 1)\left(\frac{x_i}{\epsilon}\right)^{\frac{d_l-2}{d_l-1}}}{w^2}\left(\sum_{k=0}^{w-1} x_{i+k}\right) \\ &\leq \epsilon \frac{(d_l - 1)(d_r - 1)\left(\frac{x_i}{\epsilon}\right)^{\frac{d_l-2}{d_l-1}}}{w}. \end{aligned}$$

Let $\bar{x}_i$ denote the weighted average $\bar{x}_i = \frac{1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}$. Then, for any $i \in [-\infty, 0]$

$$\bar{x}_i - \bar{x}_{i-1} \leq \frac{1}{w^2}\sum_{k=0}^{w-1} x_{i+k} \leq \frac{1}{w}.$$

Proof: Represent both x_i as well as x_{i-1} in terms of the DE (10). Taking the difference

$$\begin{aligned} \frac{x_i - x_{i-1}}{\epsilon} &= \left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}\right)^{d_l-1} \\ &\quad - \left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j-1}\right)^{d_l-1}. \end{aligned} \tag{17}$$

Apply the identity

$$A^m - B^m = (A - B)(A^{m-1} + A^{m-2}B + \cdots + B^{m-1}) \tag{18}$$

where we set

$$\begin{aligned} A &= \left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}\right) \\ B &= \left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j-1}\right) \\ m &= d_l - 1. \end{aligned}$$

Note that $A \geq B$. Thus

$$\begin{aligned} &\left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}\right)^{d_l-1} - \left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j-1}\right)^{d_l-1} \\ &= A^{d_l-1} - B^{d_l-1} \\ &= (A - B)\left(A^{d_l-2} + A^{d_l-3}B + \cdots + B^{d_l-2}\right) \\ &\overset{(i)}{\leq} (d_l - 1)(A - B)A^{d_l-2} \\ &\overset{(ii)}{=} \frac{(d_l - 1)A^{d_l-2}}{w}(f_{i-1} - f_{i+w-1}). \end{aligned}$$

In step i), we used the fact that $A \geq B$ implies $A^{d_l-2} \geq A^p B^q$ for all $p, q \in \mathbb{N}$ so that $p + q = d_l - 2$. In step ii), we made the substitution $A - B = \frac{1}{w}(f_{i-1} - f_{i+w-1})$. Since $x_i = \epsilon A^{d_l-1}$, $A^{d_l-2} = \left(\frac{x_i}{\epsilon}\right)^{\frac{d_l-2}{d_l-1}}$. Thus

$$\frac{x_i - x_{i-1}}{\epsilon} \leq \frac{(d_l - 1)\left(\frac{x_i}{\epsilon}\right)^{\frac{d_l-2}{d_l-1}}}{w}(f_{i-1} - f_{i+w-1}).$$

Consider the term $(f_{i-1} - f_{i+w-1})$. Set $f_{i-1} = C^{d_r-1}$ and $f_{i+w-1} = D^{d_r-1}$, where $C = \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i-1-k}\right)$ and $D = \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i+w-1-k}\right)$. Note that $0 \leq C, D \leq 1$. Using again (18)

$$\begin{aligned} (f_{i-1} - f_{i+w-1}) &= (C - D)\left(C^{d_r-2} + C^{d_r-3}D + \cdots + D^{d_r-2}\right) \\ &\leq (d_r - 1)(C - D). \end{aligned}$$

Explicitly

$$(C-D)=\frac{1}{w}\left(\sum_{k=0}^{w-1}(x_{i+w-1-k}-x_{i-1-k})\right)\le\frac{1}{w}\sum_{k=0}^{w-1}x_{i+k}$$

which gives us the desired upper bound. By setting all $x_{i+k}=1$, we obtain the second, slightly weaker, form.

To bound the spacing for the weighted averages, we write $\bar{x}_i$ and $\bar{x}_{i-1}$ explicitly

$$\begin{aligned}&\bar{x}_i-\bar{x}_{i-1}\\&\quad=\frac{1}{w^2}\Big((x_{i+w-1}-x_{i+w-2})+2(x_{i+w-2}-x_{i+w-3})+\\&\qquad\cdots+w(x_i-x_{i-1})+(w-1)\\&\qquad\times(x_{i-1}-x_{i-2})+\cdots+(x_{i-w+1}-x_{i-w})\Big)\\&\quad\le\frac{1}{w^2}\sum_{k=0}^{w-1}x_{i+k}\le\frac{1}{w}.\end{aligned}$$ □

The proof of the following lemma is long. Hence we relegate it to Appendix III.

Lemma 26 (Transition Length): Let $w\ge 2d_l^2d_r^2$. Let $(\epsilon,\underline{x})$, $\epsilon\in(\epsilon^{\mathrm{BP}},1]$, be a proper one-sided FP of length L. Then, for all $0<\delta<\frac{3}{2^5d_l^4d_r^6(1+12d_ld_r)}$

$$|\{i:\delta<x_i<x_s(\epsilon)-\delta\}|\le w\frac{c(d_l,d_r)}{\delta}$$

where $c(d_l,d_r)$ is a strictly positive constant independent of L and ϵ.

Let us now show how we can construct a large class of one-sided FPs, which are not necessarily stable. In particular, we will construct increasing FPs. The proof of the following theorem is relegated to Appendix IV.

Theorem 27 (Existence of One-Sided FPs): Fix the parameters (d_l,d_r,w) and let $x_u(1)<\chi$. Let $L\ge L(d_l,d_r,w,\chi)$, where

$$\begin{aligned}&L(d_l,d_r,w,\chi)\\&\quad=\max\left\{\frac{4d_lw}{d_r\left(1-\frac{d_l}{d_r}\right)(\chi-x_u(1))},\frac{8w}{\kappa^*(1)(\chi-x_u(1))^2},\right.\\&\qquad\left.\frac{8w}{\lambda^*(1)(\chi-x_u(1))\left(1-\frac{d_l}{d_r}\right)},\frac{w}{\frac{d_r}{d_l}-1}\right\}.\end{aligned}$$

There exists a proper one-sided FP $\underline{x}$ of length L that *either* has entropy χ and channel parameter bounded by

$$\epsilon^{\mathrm{BP}}(d_l,d_r)<\epsilon<1$$

or has entropy bounded by

$$\frac{\left(1-\frac{d_l}{d_r}\right)(\chi-x_u(1))}{8}-\frac{d_lw}{2d_r(L+1)}\le\chi(\underline{x})\le\chi$$

and channel parameter $\epsilon=1$.

Discussion: We will soon see that, for the range of parameters of interest, the second alternative is not possible either. In the light of this, the previous theorem asserts for this range of parameters the existence of a proper FP of entropy χ. In what follows, this FP will be the key ingredient to construct the whole EXIT curve.

B. Step 2): Construction of EXIT Curve

Definition 28 [EXIT Curve for (d_l,d_r,L,w)-Ensemble]: Let $(\epsilon^*,\underline{x}^*)$, $0\le\epsilon^*\le 1$, denote a proper one-sided FP of length L' and entropy χ. Fix $1\le L<L'$.

The *interpolated* family of constellations based on $(\epsilon^*,\underline{x}^*)$ is denoted by $\{\underline{\epsilon}(\alpha),\underline{x}(\alpha)\}_{\alpha=0}^1$. It is indexed from $-L$ to L.

This family is constructed from the one-sided FP $(\epsilon^*,\underline{x}^*)$. By definition, each element $\underline{x}(\alpha)$ is symmetric. Hence, it suffices to define the constellations in the range $[-L,0]$ and then to set $x_i(\alpha)=x_{-i}(\alpha)$ for $i\in[0,L]$. As usual, we set $x_i(\alpha)=0$ for $i\notin[-L,L]$. For $i\in[-L,0]$ and $\alpha\in[0,1]$ define

$$x_i(\alpha)=\begin{cases}(4\alpha-3)+(4-4\alpha)x_0^*, & \alpha\in\left[\frac{3}{4},1\right]\\(4\alpha-2)x_0^*-(4\alpha-3)x_i^*, & \alpha\in\left[\frac{1}{2},\frac{3}{4}\right)\\a(i,\alpha), & \alpha\in\left(\frac{1}{4},\frac{1}{2}\right)\\4\alpha x_{i-L'+L}^*, & \alpha\in\left(0,\frac{1}{4}\right]\end{cases}$$

$$\epsilon_i(\alpha)=\frac{x_i(\alpha)}{g((x_{i-w+1}(\alpha)),\ldots,(x_{i+w-1}(\alpha))}$$

where for $\alpha\in\left(\frac{1}{4},\frac{1}{2}\right)$

$$a(i,\alpha)=x^{*\,4(L'-L)\left(\frac{1}{2}-\alpha\right)\bmod(1)}_{i-\lceil 4\left(\frac{1}{2}-\alpha\right)(L'-L)\rceil}\cdot x^{*\,1-4(L'-L)\left(\frac{1}{2}-\alpha\right)\bmod(1)}_{i-\lceil 4\left(\frac{1}{2}-\alpha\right)(L'-L)\rceil+1}.$$

The constellations $\underline{x}(\alpha)$ are increasing (componentwise) as a function of α, with $\underline{x}(\alpha=0)=(0,\ldots,0)$ and with $\underline{x}(\alpha=1)=(1,\ldots,1)$.

Remark: Let us clarify the notation occurring in the definition of the term $a(i,\alpha)$ above. The expression for $a(i,\alpha)$ consists of the product of two consecutive sections of $\underline{x}^*$, indexed by the subscripts $i-\lceil 4\left(\frac{1}{2}-\alpha\right)(L'-L)\rceil$ and $i-\lceil 4\left(\frac{1}{2}-\alpha\right)(L'-L)\rceil+1$. The erasure values at the two sections are first raised to the powers $4(L'-L)\left(\frac{1}{2}-\alpha\right)\bmod(1)$ and $1-4(L'-L)\left(\frac{1}{2}-\alpha\right)\bmod(1)$, before taking their product. Here, $\bmod(1)$ represents real numbers in the interval $[0,1]$.

Discussion: The interpolation is split into four phases. For $\alpha\in\left[\frac{3}{4},1\right]$, the constellations decrease from the constant value 1 to the constant value x_0^*. For the range $\alpha\in\left[\frac{1}{2},\frac{3}{4}\right]$, the constellation decreases further, mainly towards the boundaries, so that at the end of the interval it has reached the value x_i^* at position i (hence, it stays constant at position 0). The third phase is the most interesting one. For $\alpha\in\left[\frac{1}{4},\frac{1}{2}\right]$ we "move in" the

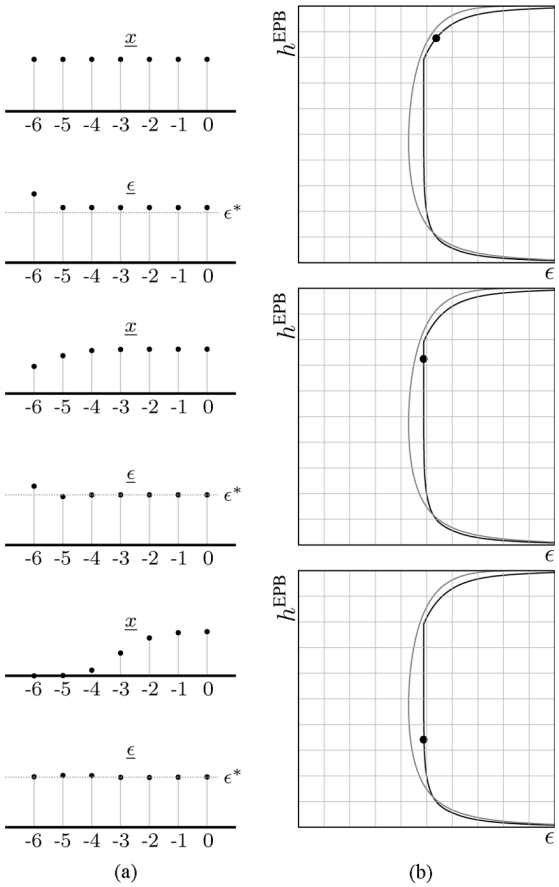

Fig. 10. Construction of EXIT curve for $(3,6,6,2)$-ensemble. The figure shows three particular points in the interpolation, namely the points $\alpha = 0.781$ [phase i)], $\alpha = 0.61$ [phase ii)], and $\alpha = 0.4$ [phase iii)]. For each parameter both the constellation $\underline{x}$ as well as the local channel parameters $\underline{\epsilon}$ are shown in the figure on left. The right column of the figure illustrates a projection of the EXIT curve. For example, we plot the average EXIT value of the constellation versus the channel value of the zeroth section. For reference, also the EBP EXIT curve of the underlying $(3,6)$-regular ensemble is shown (gray line).

constellation $\underline{x}^*$ by "taking out" sections in the middle and interpolating between two consecutive points. In particular, the value $a(i,\alpha)$ is the result of "interpolating" between two consecutive x^* values, call them x^*_j and x^*_{j+1}, where the interpolation is done in the exponents, i.e., the value is of the form $x^{*\beta}_j \cdot x^{*1-\beta}_{j+1}$. Finally, in the last phase all values are interpolated in a linear fashion until they have reached 0.

Example 29 [EXIT Curve for $(3,6,6,2)$-Ensemble]: Fig. 10 shows a small example, which illustrates this interpolation for the $(d_l = 3, d_r = 6, L = 6, w = 2)$-ensemble. We start with an FP of entropy $\chi = 0.2$ for $L' = 12$. This constellation has $\epsilon^* = 0.488223$ and

$$\underline{x}^* = (0,0,0,0,0,0.015,\\ 0.131,0.319,0.408,0.428,0.431,0.432,0.432).$$

Note that, even though the constellation is quite short, ϵ^* is close to $\epsilon^{\mathrm{MAP}}(d_l = 3, d_r = 6) \approx 0.48815$, and x^*_0 is close to $x_s(\epsilon^{\mathrm{MAP}}) \approx 0.4323$. From $(\epsilon^*, \underline{x}^*)$, we create an EXIT curve for $L = 6$. The figure shows three particular points of the interpolation, one in each of the first three phases.

Consider, e.g., the top figure corresponding to phase i). The constellation $\underline{x}$ in this case is completely flat. Correspondingly, the local channel values are also constant, except at the left boundary, where they are slightly higher to compensate for the "missing" x-values on the left.

The second figure from the top shows a point corresponding to phase ii). As we can see, the x-values close to 0 have not changed, but the x-values close to the left boundary decrease towards the solution $\underline{x}^*$. Finally, the last figure shows a point in phase iii). The constellation now "moves in." In this phase, the ϵ values are close to ϵ^*, with the possible exception of ϵ values close to the right boundary (of the one-sided constellation). These values can become large.

The proof of the following theorem can be found in Appendix V.

Theorem 30 (Fundamental Properties of EXIT Curve): Consider the parameters (d_l, d_r, w). Let $(\epsilon^*, \underline{x}^*)$, $\epsilon^* \in (\epsilon^{\mathrm{BP}}, 1]$, denote a proper one-sided FP of length L' and entropy $\chi > 0$. Then, for $1 \le L < L'$, the EXIT curve of Definition 28 has the following properties.

i) *Continuity:* The curve $\{\underline{\epsilon}(\alpha), \underline{x}(\alpha)\}_{\alpha=0}^{1}$ is continuous for $\alpha \in [0,1]$ and differentiable for $\alpha = [0,1]$ except for a finite set of points.

ii) Bounds in phase i): For $\alpha \in \left[\frac{3}{4}, 1\right]$

$$\epsilon_i(\alpha) \begin{cases} = \epsilon_0(\alpha), & i \in [-L+w-1, 0] \\ \ge \epsilon_0(\alpha), & i \in [-L, 0]. \end{cases}$$

iii) Bounds in phase ii): For $\alpha \in \left[\frac{1}{2}, \frac{3}{4}\right]$ and $i \in [-L, 0]$

$$\epsilon_i(\alpha) \ge \epsilon\left(x_0^*\right) \frac{x^*_{-L}}{x_0^*}$$

where $\epsilon(x) = \frac{x}{(1-(1-x)^{d_r-1})^{d_l-1}}$.

iv) Bounds in phase iii): Let

$$\gamma = \left(\frac{(d_r-1)(d_l-1)(\epsilon^*)^{\frac{1}{d_l-1}}(1+w^{1/8})}{w} \right)^{d_l-1}. \tag{19}$$

Let $\alpha \in \left[\frac{1}{4}, \frac{1}{2}\right]$. For $x_i(\alpha) > \gamma$

$$\epsilon_i(\alpha) \begin{cases} \le \epsilon^* \left(1 + \dfrac{1}{w^{1/8}}\right), \\ \quad i \in [-L+w-1, -w+1] \\ \ge \epsilon^* \left(1 - \dfrac{1}{1+w^{1/8}}\right), \\ \quad i \in [-L, 0]. \end{cases}$$

For $x_i(\alpha) \le \gamma$ and $w > \max\left\{2^4 d_l^2 d_r^2, 2^{16}\right\}$

$$\epsilon_i(\alpha) \ge \epsilon^* \left(1 - \frac{4}{w^{1/8}}\right)^{(d_r-2)(d_l-1)}, \qquad i \in [-L, 0].$$

v) Area under EXIT curve: The EXIT value at position $i \in [-L, L]$ is defined by

$$h_i(\alpha) = \left(g(x_{i-w+1}(\alpha), \ldots, x_{i+w-1}(\alpha))\right)^{\frac{d_l}{d_l-1}}.$$

Let

$$A(d_l,d_r,w,L)=\int_0^1 \frac{1}{2L+1}\sum_{i=-L}^{L} h_i(\alpha)d\epsilon_i(\alpha)$$

denote the area of the EXIT integral. Then

$$\left|A(d_l,d_r,w,L)-\left(1-\frac{d_l}{d_r}\right)\right|\le \frac{w}{L}d_l d_r.$$

vi) Bound on ϵ^*: For $w>\max\left\{2^4d_l^2d_r^2,2^{16}\right\}$

$$\left|\epsilon^{\mathrm{MAP}}(d_l,d_r)-\epsilon^*\right| \le \frac{2d_l d_r|x_0^*-x_s(\epsilon^*)|+c(d_l,d_r,w,L)}{\left(1-(d_l-1)^{-\frac{1}{d_r-2}}\right)^2}$$

where

$$c(d_l,d_r,w,L) = 4d_l d_r w^{-\frac{1}{8}}+\frac{wd_l(2+d_r)}{L} + d_l d_r\left(x^*_{-L'+L}+x_0^*-x^*_{-L}\right)+\frac{2d_r d_l^2}{\left(1-4w^{-\frac{1}{8}}\right)^{d_r}}w^{-\frac{7}{8}}.$$

C. Step 3): Operational Meaning of EXIT Curve

Lemma 31 [Stability of $\{(\underline{\epsilon}(\alpha),\underline{x}(\alpha))\}_{\alpha=0}^1$]: Let $\{(\underline{\epsilon}(\alpha),\underline{x}(\alpha))\}_{\alpha=0}^1$ denote the EXIT curve constructed in Definition 28. For $\beta\in(0,1)$, let

$$\epsilon^{(\beta)}=\inf_{\beta\le\alpha\le 1}\{\epsilon_i(\alpha): i\in[-L,L]\}.$$

Consider forward DE (cf., Definition 13) with parameter ϵ, $\epsilon<\epsilon^{(\beta)}$. Then, the sequence $\underline{x}^{(\ell)}$ (indexed from $-L$ to L) converges to an FP, which is pointwise upper bounded by $\underline{x}(\beta)$.

Proof: Recall from Lemma 14 that the sequence $\underline{x}^{(\ell)}$ converges to an FP of DE, call it $\underline{x}^{(\infty)}$. We claim that $\underline{x}^{(\infty)}\le\underline{x}(\beta)$.

We proceed by contradiction. Assume that $\underline{x}^{(\infty)}$ is not pointwise dominated by $\underline{x}(\beta)$. Recall that by construction of $\underline{x}(\alpha)$ the components are decreasing in α and that they are continuous. Further, $\underline{x}^{(\infty)}\le\epsilon<\underline{x}(1)$. Therefore

$$\gamma=\inf_{\beta\le\alpha\le 1}\{\alpha\,|\,\underline{x}^{(\infty)}\le\underline{x}(\alpha)\}$$

is well defined. By assumption $\gamma>\beta$. Note that there must exist at least one position $i\in[-L,0]$ so that $x_i(\gamma)=x_i^{(\infty)}$.[5] But since $\epsilon<\epsilon_i(\gamma)$ and since $g(\ldots)$ is monotone in its components

$$\begin{aligned}x_i(\gamma)&=\epsilon_i(\gamma)g(x_{i-w+1}(\gamma),\ldots,x_{i+w-1}(\gamma))\\&>\epsilon g\left(x_{i-w+1}^{(\infty)},\ldots,x_{i+w-1}^{(\infty)}\right)=x_i^{(\infty)}\end{aligned}$$

a contradiction. □

D. Step 4): Putting it All Together

We have now all the necessary ingredients to prove Theorem 10. In fact, the only statement that needs proof is (4). First note that $\epsilon^{\mathrm{BP}}(d_l,d_r,L,w)$ is a nonincreasing function in L. This follows by comparing DE for two constellations, one, say, of length L_1 and one of length L_2, $L_2>L_1$. It therefore suffices to prove (4) for the limit of L tending to infinity.

Let (d_l,d_r,w) be fixed with $w>w(d_l,d_r)$, where

$$w(d_l,d_r)=\max\left\{2^{16},2^4d_l^2d_r^2,\frac{\left(2d_ld_r\left(1+\frac{2d_l}{1-2^{-1/(d_r-2)}}\right)\right)^8}{\left(1-2^{-1/(d_r-2)}\right)^{16}\left(\frac{1}{2}(1-\frac{d_l}{d_r})\right)^8}\right\}.$$

Our strategy is as follows. We pick L' (length of constellation) sufficiently large (we will soon see what "sufficiently" means) and choose an entropy, call it $\hat{\chi}$. Then, we apply Theorem 27. Throughout this section, we will use $\underline{x}^*$ and ϵ^* to denote the FP and the corresponding channel parameter guaranteed by Theorem 27. We are faced with two possible scenarios. Either there exists an FP with the desired properties or there exists an FP with parameter $\epsilon^*=1$ and entropy at most $\hat{\chi}$. We will then show (using Theorem 30) that for sufficiently large L' the second alternative is not possible. As a consequence, we will have shown the existence of an FP with the desired properties. Using again Theorem 30 we then show that ϵ^* is close to ϵ^{MAP} and that ϵ^* is a lower bound for the BP threshold of the coupled code ensemble.

Let us make this program precise. Pick $\hat{\chi}=\frac{x_u(1)+x^{\mathrm{BP}}(d_l,d_r)}{2}$ and L' "large." In many of the subsequent steps we require specific lower bounds on L'. Our final choice is one, which obeys all these lower bounds. Apply Theorem 27 with parameters L' and $\hat{\chi}$. We are faced with two alternatives.

Consider first the possibility that the constructed one-sided FP $\underline{x}^*$ has parameter $\epsilon^*=1$ and entropy bounded by

$$\frac{\left(1-\frac{d_l}{d_r}\right)\left(x^{\mathrm{BP}}-x_u(1)\right)}{16}-\frac{d_l w}{2d_r(L'+1)}\le\chi(\underline{x}^*)\le\frac{x^{\mathrm{BP}}+x_u(1)}{2}.$$

For sufficiently large L' this can be simplified to

$$\frac{\left(1-\frac{d_l}{d_r}\right)\left(x^{\mathrm{BP}}-x_u(1)\right)}{32}\le\chi(\underline{x}^*)\le\frac{x^{\mathrm{BP}}+x_u(1)}{2}. \quad (20)$$

Let us now construct an EXIT curve based on $(\epsilon^*,\underline{x}^*)$ for a system of length L, $1\le L<L'$. According to Theorem 30, it must be true that

$$\epsilon^*\le\epsilon^{\mathrm{MAP}}(d_l,d_r)+\frac{2d_ld_r|x_0^*-x_s(\epsilon^*)|+c(d_l,d_r,w,L)}{\left(1-(d_l-1)^{-\frac{1}{d_r-2}}\right)^2}. \quad (21)$$

We claim that by choosing L' sufficiently large and by choosing L appropriately we can guarantee that

$$|x_0^*-x_s(\epsilon^*)|\le\delta,\left|x_0^*-x^*_{-L}\right|\le\delta,\ x^*_{-L'+L}\le\delta \quad (22)$$

where δ is any strictly positive number. If we assume this claim for a moment, then we see that the right-hand side of (21) can

[5]It is not hard to show that under forward DE, the constellation $\underline{x}^{(\ell)}$ is unimodal and symmetric around 0. This immediately follows from an inductive argument using Definition 13.

be made strictly less than 1. Indeed, this follows from $w > w(d_l, d_r)$ (hypothesis of the theorem) by choosing δ sufficiently small (by making L' large enough) and by choosing L to be proportional to L' (we will see how this is done in the remainder of the paper). This is a contradiction, since by assumption $\epsilon^* = 1$. This will show that the second alternative must apply.

Let us now prove the bounds in (22). In what follows, we say that sections with values in the interval $[0, \delta]$ are part of the *tail*, that sections with values in $[\delta, x_s(\epsilon^*) - \delta]$ form the *transition*, and that sections with values in $[x_s(\epsilon^*) - \delta, x_s(\epsilon^*)]$ represent the *flat* part. Recall from Definition 15 that the entropy of a constellation is the average (over all the $2L+1$ sections) erasure fraction. The bounds in (22) are equivalent to saying that *both* the tail as well as the flat part must have length at least L. From Lemma 26, for sufficiently small δ, the transition has length at most $\frac{wc(d_l,d_r)}{\delta}$ (i.e., the number of sections i with erasure value, x_i, in the interval $[\delta, x_s(\epsilon^*) - \delta]$), a constant independent of L'. Informally, therefore, most of the length L' consists of the tail or the flat part.

Let us now show all this more precisely. First, we show that the flat part is large, i.e., it is at least a fixed fraction of L'. We argue as follows. Since the transition contains only a constant number of sections, its contribution to the entropy is small. More precisely, this contribution is upper bounded by $\frac{wc(d_l,d_r)}{(L'+1)\delta}$. Further, the contribution to the entropy from the tail is small as well, namely at most δ. Hence, the total contribution to the entropy stemming from the tail plus the transition is at most $\frac{wc(d_l,d_r)}{(L'+1)\delta} + \delta$. However, the entropy of the FP is equal to $\frac{x^{\mathrm{BP}} + x_u(1)}{2}$. As a consequence, the flat part must have length, which is at least a fraction $\frac{x^{\mathrm{BP}} + x_u(1)}{2} - \frac{wc(d_l,d_r)}{(L'+1)\delta} - \delta$ of L'. This fraction is strictly positive if we choose δ small enough and L' large enough.

By a similar argument we can show that the tail length is also a strictly positive fraction of L'. From Lemma 23, $x_s(\epsilon^*) > x^{\mathrm{BP}}$. Hence, the flat part cannot be too large since the entropy is equal to $\frac{x^{\mathrm{BP}} + x_u(1)}{2}$, which is strictly smaller than x^{BP}. As a consequence, the tail has length at least a fraction $1 - \frac{x^{\mathrm{BP}} + x_u(1)}{2(x^{\mathrm{BP}} - \delta)} - \frac{1 + \frac{wc(d_l,d_r)}{\delta}}{L'+1}$ of L'. As before, this fraction is also strictly positive if we choose δ small enough and L' large enough. Hence, by choosing L to be the lesser of the length of the flat part and the tail, we conclude that the bounds in (22) are valid and that L can be chosen arbitrarily large (by increasing L').

Consider now the second case. In this case, $\underline{x}^*$ is a proper one-sided FP with entropy equal to $\frac{x^{\mathrm{BP}} + x_u(1)}{2}$ and with parameter $\epsilon^{\mathrm{BP}}(d_l, d_r) < \epsilon^* < 1$. Now, using again Theorem 30, we can show

$$\epsilon^* > \epsilon^{\mathrm{MAP}}(d_l, d_r) - 2w^{-\frac{1}{8}} \frac{4d_l d_r + \frac{2d_r d_l^2}{\left(1 - 4w^{-\frac{1}{8}}\right)^{d_r}}}{\left(1 - (d_l - 1)^{-\frac{1}{d_r - 1}}\right)^2}$$
$$\stackrel{d_l \geq 3}{\geq} \epsilon^{\mathrm{MAP}}(d_l, d_r) - 2w^{-\frac{1}{8}} \frac{4d_l d_r + \frac{2d_r d_l^2}{\left(1 - 4w^{-\frac{1}{8}}\right)^{d_r}}}{\left(1 - 2^{-\frac{1}{d_r}}\right)^2}.$$

To obtain the above expression, we take L' to be sufficiently large in order to bound the term in $c(d_l, d_r, w, L)$, which contains L. We also use (22) and choose δ to be sufficiently small to bound the corresponding terms. We also replace $w^{-7/8}$ by $w^{-1/8}$ in $c(d_l, d_r, w, L)$.

To summarize, we conclude that for an entropy equal to $\frac{x^{\mathrm{BP}}(d_l,d_r) + x_u(1)}{2}$, for sufficiently large L', $\underline{x}^*$ must be a proper one-sided FP with parameter ϵ^* bounded as above.

Finally, let us show that $\epsilon^* \left(1 - \frac{4}{w^{1/8}}\right)^{d_r d_l}$ is a lower bound on the BP threshold. We start by claiming that

$$\epsilon^* \left(1 - \frac{4}{w^{1/8}}\right)^{d_r d_l} < \epsilon^* \left(1 - \frac{4}{w^{1/8}}\right)^{(d_r - 2)(d_l - 1)} = \inf_{\frac{1}{4} \leq \alpha \leq 1} \{\epsilon_i(\alpha) : i \in [-L, L]\}.$$

To prove the above claim we just need to check that $\epsilon(x_0^*) x_{-L}^* / x_0^*$ [see bounds in phase ii) of Theorem 30] is greater than the above infimum. Since in the limit of $L' \to \infty$, $\epsilon(x_0^*) x_{-L}^* / x_0^* \to \epsilon^*$, for sufficiently large L' the claim is true.

From the hypothesis of the theorem, we have $w > 2^{16}$. Hence, $\epsilon^*(1 - 4w^{-1/8})^{d_r d_l} > 0$. Apply forward DE (cf., Definition 13) with parameter $\epsilon < \epsilon^*(1 - 4w^{-1/8})^{d_r d_l}$ and length L. Denote the FP by $\underline{x}^\infty$ (with indices belonging to $[-L, L]$). From Lemma 31, we then conclude that $\underline{x}^\infty$ is pointwise upper bounded by $\underline{x}\left(\frac{1}{4}\right)$. But for $\alpha = 1/4$, we have

$$x_i(1/4) \leq x_0(1/4) = x_{-L'+L}^* \leq \delta < x_u(1) \qquad \forall\, i$$

where we make use of the fact that δ can be chosen arbitrarily small. Thus, $x_i^{(\infty)} < x_u(1)$ for all $i \in [-L, L]$. Consider a one-sided constellation $\underline{y}$ with $y_i = x_0(1/4) < x_u(1)$ for all $i \in [-L, 0]$. Recall that for a one-sided constellation $y_i = y_0$ for all $i > 0$ and as usual $y_i = 0$ for $i < -L$. Clearly, $\underline{x}^{(\infty)} \leq \underline{y}$. Now apply one-sided forward DE to $\underline{y}$ with parameter ϵ (same as the one we applied to get $\underline{x}^\infty$) and call its limit $\underline{y}^{(\infty)}$. From part i) of Lemma 22, we conclude that the limit $\underline{y}^{(\infty)}$ is either proper or trivial. Suppose that $\underline{y}^\infty$ is proper (implies nontrivial). Clearly, $y_i^\infty < x_u(1)$ for all $i \in [-L, 0]$. But from Lemma 23, we have that for any proper one-sided FP $y_0 \geq x_u(\epsilon) \geq x_u(1)$, a contradiction. Hence, we conclude that $\underline{y}^\infty$ must be trivial and so must be $\underline{x}^\infty$.

VI. Discussion and Possible Extensions

A. New Paradigm for Code Design

The explanation of why convolutional-like LDPC ensembles perform so well given in this paper gives rise to a new paradigm in code design.

In most designs of codes based on graphs one encounters a tradeoff between the threshold and the error floor behavior. For example, for standard irregular graphs an optimization of the threshold tends to push up the number of degree-two variable nodes. The same quantity, on the other hand, favors the existence of low weight (pseudo)codewords.

For convolutional-like LDPC ensembles the important operational quantity is the MAP threshold of the underlying ensemble. As, e.g., regular LDPC ensembles show, it is simple to improve the MAP threshold *and* to improve the error-floor

performance—just increase the minimum variable-node degree. From this perspective one should simply pick as large a variable-node degree as possible.

There are some drawbacks to picking large degrees. First, picking large degrees also increases the complexity of the scheme. Second, although currently little is known about the scaling behavior of the convolutional-like LDPC ensembles, it is likely that large degrees imply a slowing down of the convergence of the performance of finite-length ensembles to the asymptotic limit. This implies that one has to use large block lengths. Third, the larger we pick the variable-node degrees the higher the implied rate loss. Again, this implies that we need very long codes in order to bring down the rate loss to acceptable levels. It is tempting to conjecture that the minimum rate loss that is required in order to achieve the change of thresholds is related to the area under the EXIT curve between the MAP and the BP threshold. For example, in Fig. 5, this is the light gray area. For the underlying ensemble this is exactly the amount of guessing (help) that is needed so that a local algorithm can decode correctly, assuming that the underlying channel parameter is the MAP threshold.

Due to the above reasons, an actual code design will therefore try to maintain relatively small average degrees so as to keep this gray area small. But the additional degree of freedom can be used to design codes with good thresholds *and* good error floors.

B. Scaling Behavior

In our design there are three parameters that tend to infinity. The number of variables nodes at each position, called M, the length of the constellation L, and the length of the smoothing window w. Assume we fix w and we are content with achieving a threshold slightly below the MAP threshold. How should we scale M with respect to L so that we achieve the best performance? This question is of considerable practical importance. Recall that the total length of the code is of order $L \cdot M$. We would therefore like to keep this product small. Further, the rate loss is of order $1/L$ (so L should be large) and M should be chosen large so as to approach the performance predicted by DE. Finally, how does the number of required iterations scale as a function of L?

Also, in the proof we assumed that we fix L and let M tend to infinity so that we can use DE techniques. We have seen that in this limit the boundary conditions of the system dictate the performance of the system regardless of the size of L (as long as L is fixed and M tends to infinity). Is the same behavior still true if we let L tend to infinity as a function of M? At what scaling does the behavior change?

C. Tightening of Proof

As mentioned already in the introduction, our proof is weak—it promises that the BP threshold approaches the MAP threshold of the underling ensemble at a speed of $w^{-1/8}$. Numerical experiments indicate that the actual convergence speed is likely to be exponential and that the prefactors are very small. Why is the analytic statement so loose and how can it be improved?

Within our framework it is clear that at many places the constants could be improved at the cost of a more involved proof. It is therefore likely that a more careful analysis following the same steps will give improved convergence speeds.

More importantly, for mathematical convenience we constructed an "artificial" EXIT curve by interpolating a particular FP and we allowed the channel parameter to vary as a function of the position. In the proof we then coarsely bounded the "operational" channel parameter by the minimum of all the individual channel parameters. This is a significant source for the looseness of the bound. A much tighter bound could be given if it were possible to construct the EXIT curve by direct methods. As we have seen, it is possible to show the existence of FPs of DE for a wide range of EXIT values. The difficulty consists in showing that all these individual FPs form a smooth 1-D manifold so that one can use the Area Theorem and integrate with respect to this curve.

D. Extensions to BMS Channels and General Ensembles

Preliminary numerical evidence suggests that the behavior of the convolutional-like LDPC ensembles discussed in this paper is not restricted to the BEC channel or to regular ensembles but is a general phenomenon. We will be brief. A more detailed discussion can be found in the two recent papers [26], [27]. Let us quickly discuss how one might want to attack the more general setup.

We have seen that the proof consists essentially of three steps.

1) Existence of FP: As long as we stay with the BEC, a similar procedure as the one used in the proof of Theorem 27 can be used to show the existence of the desired FP for more general *ensembles*.
   General BMS *channels* are more difficult to handle, but FP theorems do exist also in the setting of infinite-dimensional spaces. The most challenging aspect of this step is to prove that the constructed FP has the essential basic characteristics that we relied upon for our later steps. In particular, we need it to be unimodal, to have a short transition period, and to approach the FP density of the underlying standard ensemble.
2) Construction of EXIT curve and bounds: Recall that in order to create a whole EXIT curve, we started with an FP and interpolated the value of neighboring points. In order to ensure that each such interpolated constellation is indeed an FP, we allowed the local channel parameters to vary. By choosing the interpolation properly, we were then able to show that this variation is small. As long as one remains in the realm of BEC channels, the same technique can in principle be applied to other ensembles. For general channels the construction seems more challenging. It is not true in general that, given a constellation, one can always find "local" channels that make this constellation an FP. It is therefore not clear how an interpolation for general channels can be accomplished. This is perhaps the most challenging hurdle for any potential generalization.
3) *Operational interpretation*: For the operational interpretation we relied upon the notion of physical degradation. We showed that, starting with a channel parameter of

a channel, which is upgraded w.r.t. to any of the local channels used in the construction of the EXIT curve, we do not get stuck in a nontrivial FP. For the BEC, the notion of degradation is very simple, it is the natural order on the set of erasure probabilities, and this is a total order. For general channels, an order on channels still exists in terms of degradation, but this order is partial. We therefore require that the local channels used in the construction of the EXIT curve are all degraded w.r.t. a channel of the original channel family (e.g., the family of Gaussian channels) with a parameter, which is only slightly better than the parameter, which corresponds to the MAP threshold.

E. Extension to General Coupled Graphical Systems

Codes based on graphs are just one instance of graphical systems that have distinct thresholds for "local" algorithms (what we called the BP threshold) and for "optimal" algorithms (what we called the MAP threshold). To be sure, coding is somewhat special—it is conjectured that the so-called replica-symmetric solution always determines the threshold under MAP processing for codes based on graphs. Nevertheless, it is interesting to investigate to what extent the coupling of general graphical systems shows a similar behavior. Is there a general class of graphical models in which the same phenomenon occurs? If so, can this phenomenon either be used to analyze systems or to devise better algorithms?

Appendix I
Proof of Lemma 1

We proceed as follows. We first consider a "circular" ensemble. This ensemble is defined in an identical manner as the (d_l, d_r, L) ensemble except that the positions are now from 0 to $K-1$ and index arithmetic is performed modulo K. This circular definition symmetrizes all positions, which in turn simplifies calculations.

As we will see shortly, most codes in this circular ensemble have a minimum stopping set distance, which is a linear fraction of M. To make contact with our original problem we now argue as follows. Set $K = 2L + d_l$. If, for the circular ensemble, we take $d_l - 1$ consecutive positions and set them to 0 then this "shortened" ensemble has length $2L+1$ and it is in one-to-one correspondence with the (d_l, d_r, L) ensemble. Clearly, no new stopping sets are introduced by shortening the ensemble. This proves the claim.

Let $A(d_l, d_r, M, K, w)$ denote the expected number of stopping sets of weight w of the "circular" ensemble. Let $\mathcal{C}$ denote a code chosen uniformly at random from this ensemble.

Recall that every variable node at position i connects to a check node at positions $i - \hat{d}_l, \ldots, i + \hat{d}_l$, modulo K. There are M variable nodes at each position and $M\frac{d_l}{d_r}$ check nodes at each position. Conversely, the Md_l edges entering the check nodes at position i come equally from variable nodes at position $i - \hat{d}_l, \ldots, i + \hat{d}_l$. These Md_l edges are connected to the check nodes via a random permutation.

Let $w_k, k \in \{0, \ldots, K-1\}, 0 \le w_k \le M$, denote the weight at position i, i.e., the number of variable nodes at position i that have been set to 1. Call $\underline{w} = (w_0, \ldots, w_{K-1})$ the *type*. We are interested in the expected number of stopping sets for a particular type; call this quantity $A(d_l, d_r, M, K, \underline{w})$. Since the parameters (d_l, d_r, M, K) are understood from the context, we shorten the notation to $A(\underline{w})$. We claim that (23), shown at the bottom of the page, holds, where $p(x) = \sum_{i \ne 1} \binom{d_r}{i} x^i$. This expression is easily explained. The w_k variable nodes at position k that are set to 1 can be distributed over the M variable nodes in $\binom{M}{w_k}$ ways. Next, we have to distribute the $\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}$ ones among the $M\frac{d_l}{d_r}$ check nodes in such a way that every check node is fulfilled (since we are looking for stopping sets, "fulfilled" means that a check node is either connected to no variable node with associated value "1" or to *at least* two such nodes). This is encoded by $\operatorname{coef}\left\{p(x)^{M\frac{d_l}{d_r}}, x^{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}\right\}$. Finally, we have to divide by the total number of possible connections; there are Md_l check node sockets at position k and we distribute $\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}$ ones. This can be done in $\binom{Md_l}{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}$ ways. To justify step (a) note that

$$\prod_{i=-\hat{d}_l}^{\hat{d}_l} \binom{M}{w_{k+i}}^{\frac{1}{d_l}} \le 2^{M\frac{1}{d_l}\sum_{i=-\hat{d}_l}^{\hat{d}_l} h\left(\frac{w_{k+i}}{M}\right)} \overset{\text{Jensen}}{\le} 2^{Mh\left(\frac{1}{d_l}\sum_{i=-\hat{d}_l}^{\hat{d}_l} \frac{w_{k+i}}{M}\right)} \le (M+1)\binom{M}{\frac{1}{d_l}\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}.$$

$$A(\underline{w}) = \frac{\prod_{k=0}^{K-1} \binom{M}{w_k} \operatorname{coef}\left\{p(x)^{M\frac{d_l}{d_r}}, x^{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}\right\}}{\prod_{k=0}^{K-1} \binom{Md_l}{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}} \overset{(a)}{\le} \prod_{k=0}^{K-1} \frac{(M+1)\binom{M}{\frac{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}{d_l}} \operatorname{coef}\left\{p(x)^{M\frac{d_l}{d_r}}, x^{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}\right\}}{\binom{Md_l}{\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i}}}. \tag{23}$$

Note that, besides the factor $(M+1)$, which is negligible, each term in the product (23) has the exact form of the average stopping set weight distribution of the standard (d_l, d_r)-ensemble of length M and weight $\frac{1}{d_l}\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_k$. (Potentially this weight is nonintegral but the expression is nevertheless well defined.)

We can therefore leverage known results concerning the stopping set weight distribution for the underlying (d_l, d_r)-regular ensembles. For the (d_l, d_r)-regular ensembles we know that the relative minimum distance is at least $\hat{\omega}(d_l, d_r)$ with high probability [13, Lemma D.17]. Therefore, as long as $\frac{1}{d_l M}\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i} < \hat{\omega}(d_l, d_r)$, for all $0 \le k < K$, $\frac{1}{MK}\log A(\underline{w})$ is strictly negative and so most codes in the ensemble do not have stopping sets of this type. The claim now follows since in order for the condition $\frac{1}{d_l M}\sum_{i=-\hat{d}_l}^{\hat{d}_l} w_{k+i} < \hat{\omega}(d_l, d_r)$ to be violated for at least one position k we need $\frac{1}{M}\sum_{k=0}^{K-1} w_k$ to exceed $d_l \hat{\omega}(d_l, d_r)$.

Appendix II
Basic Properties of $h(x)$

Recall the definition of $h(x)$ from (2). We have the following.

Lemma 32 [Basic Properties of $h(x)$]: Consider the (d_l, d_r)-regular ensemble with $d_l \ge 3$ and let $\epsilon \in (\epsilon^{\mathrm{BP}}, 1]$.

i) $h'(x_u(\epsilon)) > 0$ and $h'(x_s(\epsilon)) < 0$; $|h'(x)| \le d_l d_r$ for $x \in [0,1]$.

ii) There exists a unique value $0 \le x_*(\epsilon) \le x_u(\epsilon)$ so that $h'(x_*(\epsilon)) = 0$, and there exists a unique value $x_u(\epsilon) \le x^*(\epsilon) \le x_s(\epsilon)$ so that $h'(x^*(\epsilon)) = 0$.

iii) Let

$$\kappa_*(\epsilon) = \min\left\{-h'(0), \frac{-h(x_*(\epsilon))}{x_*(\epsilon)}\right\}$$
$$\lambda_*(\epsilon) = \min\left\{h'(x_u(\epsilon)), \frac{-h(x_*(\epsilon))}{x_u(\epsilon) - x_*(\epsilon)}\right\}$$
$$\kappa^*(\epsilon) = \min\left\{h'(x_u(\epsilon)), \frac{h(x^*(\epsilon))}{x^*(\epsilon) - x_u(\epsilon)}\right\}$$
$$\lambda^*(\epsilon) = \min\left\{-h'(x_s(\epsilon)), \frac{h(x^*(\epsilon))}{x_s(\epsilon) - x^*(\epsilon)}\right\}.$$

The quantities $\kappa_*(\epsilon)$, $\lambda_*(\epsilon)$, $\kappa^*(\epsilon)$, and $\lambda^*(\epsilon)$ are nonnegative and depend only on the channel parameter ϵ and the degrees (d_l, d_r). In addition, $\kappa_*(\epsilon)$ is strictly positive for all $\epsilon \in [0,1]$.

iv) For $0 \le \epsilon \le 1$

$$x_*(\epsilon) > \frac{1}{d_l^2 d_r^2}.$$

v) For $0 \le \epsilon \le 1$

$$\kappa_*(\epsilon) \ge \frac{1}{8 d_r^2}.$$

vi) If we draw a line from 0 with slope $-\kappa_*$, then $h(x)$ lies below this line for $x \in [0, x_*]$.
If we draw a line from $x_u(\epsilon)$ with slope λ_*, then $h(x)$ lies below this line for all $x \in [x_*, x_u(\epsilon)]$.
If we draw a line from $x_u(\epsilon)$ with slope κ^*, then $h(x)$ lies above this line for $x \in [x_u(\epsilon), x^*]$.
Finally, if we draw a line from $x_s(\epsilon)$ with slope $-\lambda^*$, then $h(x)$ lies above this line for all $x \in [x^*, x_s(\epsilon)]$.

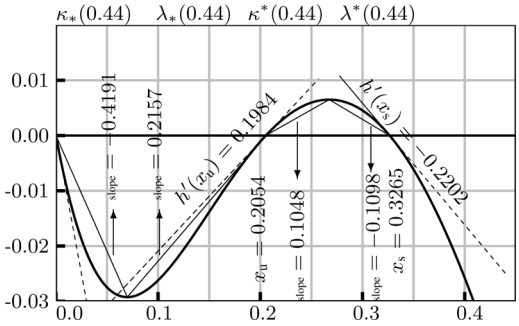

Fig. 11. Pictorial representation of the various quantities, which appear in Lemma 32. We use the $(3,6)$ ensemble to transmit over a BEC with erasure probability $\epsilon = 0.44$. The function $h(x) = 0.44(1-(1-x)^5)^2 - x$ is represented in the figure by the smooth bold curve. The roots of $h(x) = 0$ or, equivalently, the FPs of DE are given by 0, $x_u(0.44) \approx 0.2054$, and $x_s(0.44) \approx 0.3265$. There are only two stationary points of $h(x)$, i.e., only two points at which $h'(x) = 0$. They are given by $x_*(0.44) \approx 0.0697$ and $x^*(0.44) \approx 0.2673$. Along with the curve $h(x)$, the figure contains three dashed lines representing the tangents at the points 0, $x_u(0.44)$, and $x_s(0.44)$. The slopes of the tangents at 0, $x_u(0.44)$ and $x_s(0.44)$ are $h'(0) = -1$, $h'(x_u) = 0.1984$, and $h'(x_s) = -0.2202$, respectively. Also shown are the four lines, which bound $h(x)$ in the various regions. These lines are given (their endpoints) by: $\{(0,0),(x_*,h(x_*))\}$, $\{(x_*,h(x_*)),(x_u(0.44),0)\}$, $\{(x_u(0.44),0),(x^*,h(x^*))\}$, and $\{(x^*,h(x^*)),(x_s(0.44),0)\}$ and have slopes $-0.4191, 0.2157, 0.1048$, and -0.1098, respectively. Thus, we have $\kappa^*(0.44) = 0.1048$, $\lambda^*(0.44) = 0.1098$, $\kappa_*(0.44) = 0.4191$, and $\lambda_*(0.44) = 0.1984$.

Example 33 [$(3,6)$-Ensemble]: Consider transmission using a code from the $(3,6)$ ensemble over a BEC with $\epsilon = 0.44$. The FP equation for the BP decoder is given by

$$x = 0.44(1-(1-x)^5)^2.$$

The function $h(x) = 0.44(1-(1-x)^5)^2 - x$ is shown in Fig. 11. The equation $h(x) = 0$ has exactly three real roots, namely, $0, x_u(0.44) \approx 0.2054$, and $x_s(0.44) \approx 0.3265$. Further properties of $h(x)$ are shown in Fig. 11.

Let us prove each part separately. In order to lighten our notation, we drop the ϵ dependence for quantities like x_u, x_s, x_*, or x^*.

i) Note that $h(x) > 0$ for all $x \in (x_u, x_s)$, with equality at the two ends. This implies that $h'(x_u) > 0$ and that $h'(x_s) < 0$. With respect to the derivative, we have

$$\begin{aligned}|h'(x)| &= \big|\epsilon(d_l-1)(d_r-1)(1-x)^{d_r-2} \\ &\qquad \times (1-(1-x)^{d_r-1})^{d_l-2} - 1\big| \\ &\le (d_l-1)(d_r-1)+1 \le d_l d_r.\end{aligned}$$

ii) We claim that $h''(x) = 0$ has exactly one real solution in $(0,1)$. We have

$$\begin{aligned}h''(x) &= \epsilon(d_l-1)(d_r-1)(1-x)^{d_r-3} \\ &\quad \times \left(1-(1-x)^{d_r-1}\right)^{d_l-3} \\ &\quad \times \left[(1-x)^{d_r-1}(d_l d_r - d_l - d_r) - d_r + 2\right].\end{aligned} \tag{24}$$

Thus, $h''(x) = 0$ for $x \in (0,1)$ only at

$$x = 1 - \left(\frac{d_r - 2}{d_l d_r - d_l - d_r}\right)^{\frac{1}{d_r-1}}. \tag{25}$$

Since $d_l \ge 3$, the above solution is in $(0,1)$.

Since $h(0) = h(x_u) = h(x_s) = 0$, we know from Rolle's theorem that there must exist an $0 \leq x_* \leq x_u$ and an $x_u \leq x^* \leq x_s$, such that $h'(x_*) = h'(x^*) = 0$.
Now suppose that there exists a $y \in (0,1)$, $x_* \neq y \neq x^*$, such that $h'(y) = 0$, so that $h'(\cdot)$ vanishes at three distinct places in $(0,1)$. Then, by Rolle's theorem, we conclude that $h''(x) = 0$ has at least two roots in the interval $(0,1)$, a contradiction.

iii) To check that the various quantities in part iii) are strictly positive, it suffices to verify that $h(x_*) \neq 0$ and $h(x^*) \neq 0$. But we know from Lemma 9 that $h(x) = 0$ has exactly two solutions, namely x_u and x_s, and neither of them is equal to x^* or x_* since $h'(x_u) > 0$.

iv) From (24), for all $x \in [0,1]$, we can upper bound $|h''(x)|$ by

$$(d_l - 1)(d_r - 1)[d_l d_r - d_l - d_r - d_r + 2] < d_l^2 d_r^2. \quad (26)$$

Note that $h'(0) = -1$ and, by definition, $h'(x_*) = 0$, so that $\frac{1}{x_*} = \frac{h'(x_*) - h'(0)}{x_* - 0}$. Consider the function $h'(x)$, $x \in [0, x_*]$. From the continuity of the function $h'(x)$ and, using the mean-value theorem, we conclude that there exists an $\eta \in (0, x_*)$ such that $h''(\eta) = \frac{h'(x_*) - h'(0)}{x_*}$. But from (26), we know that $h''(\eta) < d_l^2 d_r^2$. It follows that $\frac{1}{x_*} = \frac{h'(x_*) - h'(0)}{x_*} < d_l^2 d_r^2$.

v) To get the universal lower bound on $\kappa_*(\epsilon)$ note that the dominant (i.e., smaller) term in the definition of $\kappa_*(\epsilon)$ is $\frac{-h(x_*(\epsilon))}{x_*(\epsilon)}$. (The second term, $-h'(0)$, is 1.) Recall that x_* is the point where $h(x)$ takes on the minimum value in the range $[0, x_u(\epsilon)]$. We can therefore rewrite $\kappa_*(\epsilon)$ in the form $\frac{1}{x_*} \max_{0 \leq x \leq x_u(\epsilon)} \{-h(x)\}$. To get a lower bound on $\kappa_*(\epsilon)$ we use the trivial upper bound $x_*(\epsilon) \leq 1$. It therefore remains to lower bound $\max_{0 \leq x \leq x_u(\epsilon)} \{-h(x)\}$. Notice that $-h(x)$ is a decreasing function of ϵ for every $x \in [0, x_u(1)]$. Thus, inserting $\epsilon = 1$, we get

$$\begin{aligned}
&\max_{0 \leq x \leq x_u(1)} \left[x - \left(1 - (1-x)^{d_r - 1}\right)^{d_l - 1} \right] \\
&\quad = \max_{0 \leq x \leq x_u(1)} \left[\left(x^{\frac{1}{d_l - 1}}\right)^{d_l - 1} - \left(1 - (1-x)^{d_r - 1}\right)^{d_l - 1} \right] \\
&\quad \geq \max_{0 \leq x \leq (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}} \left(x^{\frac{1}{d_l - 1}} - (d_r - 1)x \right) x^{\frac{d_l - 2}{d_l - 1}}.
\end{aligned}$$

Let us see how we derived the last inequality. First, we claim that for $x \in \left[0, (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}\right]$ we have $x^{\frac{1}{d_l - 1}} \geq (d_r - 1)x \geq 1 - (1 - x)^{d_r - 1}$. Indeed, this can be easily seen by using the identity $1 - (1-x)^{d_r - 1} = x(1 + (1-x) + \cdots + (1-x)^{d_r - 2})$ and $x \leq 1$. Then, we use $A^{d_l - 1} - B^{d_l - 1} = (A - B)(A^{d_l - 2} + A^{d_l - 3}B + \cdots + B^{d_l - 2}) \geq (A - B)A^{d_l - 2}$ for all $0 \leq B \leq A$. Finally, we use

$$(x_u(1))^{\frac{1}{d_l - 1}} = \left(1 - (1 - x_u(1))^{d_r - 1}\right) \leq (d_r - 1)x_u(1)$$

so that

$$x_u(1) \geq (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}. \quad (27)$$

As a consequence $\left[0, (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}\right] \subseteq [0, x_u(1)]$ and hence we get the last inequality. Now we can further lower bound the right-hand side above by evaluating it at any element of $\left[0, (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}\right]$.
We pick $\hat{x} = 2^{-\frac{d_l - 1}{d_l - 2}} (d_r - 1)^{-\frac{d_l - 1}{d_l - 2}}$. Continuing the chain of inequalities, we get

$$\begin{aligned}
&\overset{x = \hat{x}}{\geq} \left(2^{d_l - 1}(d_r - 1)\right)^{-\frac{1}{d_l - 2}} (\hat{x})^{\frac{d_l - 2}{d_l - 1}} \\
&\quad = \left(2^{d_l - 1}(d_r - 1)\right)^{-\frac{1}{d_l - 2}} \left(2^{-1}(d_r - 1)^{-1}\right) \\
&\quad = \frac{1}{2^{\frac{2d_l - 3}{d_l - 2}} (d_r - 1)^{\frac{d_l - 1}{d_l - 2}}} \overset{(a)}{\geq} \frac{1}{8(d_r - 1)^2} \geq \frac{1}{8 d_r^2}.
\end{aligned}$$

Since $d_l \geq 3$ we have $\frac{2d_l - 3}{d_l - 2} \leq 3$ and $\frac{d_l - 1}{d_l - 2} \leq 2$. Hence, we obtain (a).

vi) Let us prove that for all $x \in (x_u, x^*)$, $h(x)$ is strictly above the line, which contains the point $(x_u, 0)$ and has slope κ^*. Denote this line by $l(x)$. More precisely, we have $l(x) = \kappa^*(x - x_u)$. Suppose to the contrary that there exists a point $y \in (x_u, x^*)$ such that $h(y) < l(y)$. In this case, we claim that the equation $h(x) - l(x) = 0$ must have at least four roots.
This follows from a) $h(x_u) = l(x_u)$, b) $h'(x_u) \geq l'(x_u)$, c) $h(y) < l(y)$, d) $h(x^*) \geq l(x^*)$, and, finally, e) $h(1) < l(1)$, where $x_u < y < x^* < 1$. If all these inequalities are strict then the four roots are distinct. Otherwise, some roots will have higher multiplicities.
But if $h(x) - l(x) = 0$ has at least four roots then $h''(x) - l''(x) = 0$ has at least two roots. Note that $l''(x) = 0$, since $l(x)$ is a linear function. This leads to a contradiction, since, as discussed in part ii), $h''(x)$ has only one (single) root in $(0,1)$.
The other cases can be proved along similar lines.

Appendix III
Proof of Lemma 26

We split the transition into several stages. Generically, in each of the ensuing arguments we consider a section with associated value just above the lower bound of the corresponding interval. We then show that, after a fixed number of further sections, the value must exceed the upper bound of the corresponding interval. Depending on the length L and the entropy of the constellation there might not be sufficiently many sections left in the constellation to pass all the way to $x_s(\epsilon) - \delta$. In this case, the conclusion of the lemma is trivially fulfilled. Therefore, in what follows, we can always assume that there are sufficiently many points in the constellation.

In the remainder of the paper, $\kappa_*(\epsilon)$ and $x_*(\epsilon)$ are the specific quantities for a particular ϵ, whereas κ_* and x_* are the strictly positive universal bounds valid for all ϵ, discussed in Lemma 32. We write κ_* and x_* instead of $\frac{1}{8d_r^2}$ and $\frac{1}{d_l^2 d_r^2}$ to emphasize their operational meaning.

i) *Let $\delta > 0$. Then, there are at most $w\left(\frac{1}{\kappa_* \delta} + 1\right)$ sections i with value x_i in the interval $[\delta, x_*(\epsilon)]$.*

Let i be the smallest index so that $x_i \geq \delta$. If $x_{i+(w-1)} \geq x_*(\epsilon)$, then the claim is trivially fulfilled. Assume therefore that $x_{i+(w-1)} \leq x_*(\epsilon)$. Using the monotonicity of $g(\cdot)$

$$\begin{aligned} x_i &= \epsilon g(x_{i-(w-1)}, \dots, x_i, \dots, x_{i+(w-1)}) \\ &\leq \epsilon g(x_{i+(w-1)}, \dots, x_{i+(w-1)}). \end{aligned}$$

This implies

$$\begin{aligned} &x_{i+(w-1)} - x_i \\ &\quad \geq x_{i+(w-1)} - \epsilon g(x_{i+(w-1)}, \dots, x_{i+(w-1)}) \\ &\quad \overset{(2)}{=} -h(x_{i+(w-1)}) \overset{\text{Lemma 32(vi)}}{\geq} -l(x_{i+(w-1)}) \\ &\quad \geq -l(x_i) \geq -l(\delta) = \kappa_*(\epsilon)\delta. \end{aligned}$$

This is equivalent to

$$x_{i+(w-1)} \geq x_i + \kappa_*(\epsilon)\delta.$$

More generally, using the same line of reasoning

$$x_{i+l(w-1)} \geq x_i + l\kappa_*(\epsilon)\delta$$

as long as $x_{i+l(w-1)} \leq x_*(\epsilon)$.

We summarize. The total distance we have to cover is $x_* - \delta$ and every $(w-1)$ steps we cover a distance of at least $\kappa_*(\epsilon)\delta$ as long as we have not surpassed $x_*(\epsilon)$. Therefore, after $(w-1)\left\lfloor \frac{x_*(\epsilon)-\delta}{\kappa_*(\epsilon)\delta} \right\rfloor$ steps, we have either passed x_* or we must be strictly closer to x_* than $\kappa_*(\epsilon)\delta$. Hence, to cover the remaining distance, we need at most $(w-2)$ extra steps. The total number of steps needed is therefore upper bounded by $w - 2 + (w-1)\left\lfloor \frac{x_*(\epsilon)-\delta}{\kappa_*(\epsilon)\delta} \right\rfloor$, which, in turn, is upper bounded by $w\left(\frac{x_*(\epsilon)}{\kappa_*(\epsilon)\delta} + 1\right)$. The final claim follows by bounding $x_*(\epsilon)$ with 1 and $\kappa_*(\epsilon)$ by κ_*.

ii) *From $x_*(\epsilon)$ up to $x_u(\epsilon)$ it takes at most $w\left(\frac{8}{3\kappa_*(x_*)^2} + 2\right)$ sections.*

Recall that $\bar{x}_i$ is defined by $\bar{x}_i = \frac{1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}$. From Lemma 24 part i), $x_i \leq \epsilon g(\bar{x}_i, \bar{x}_i, \dots, \bar{x}_i) = \bar{x}_i + h(\bar{x}_i)$. Sum this inequality over all sections from $-\infty$ to $k \leq 0$

$$\sum_{i=-\infty}^{k} x_i \leq \sum_{i=-\infty}^{k} \bar{x}_i + \sum_{i=-\infty}^{k} h(\bar{x}_i).$$

Writing $\sum_{i=-\infty}^{k} \bar{x}_i$ in terms of the x_i, for all i, and rearranging terms

$$\begin{aligned} -\sum_{i=-\infty}^{k} h(\bar{x}_i) &\leq \frac{1}{w^2}\sum_{i=1}^{w-1} \binom{w-i+1}{2}(x_{k+i} - x_{k-i+1}) \\ &\leq \frac{w}{6}(x_{k+(w-1)} - x_{k-(w-1)}). \end{aligned}$$

Let us summarize

$$x_{k+(w-1)} - x_{k-(w-1)} \geq -\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i). \tag{28}$$

From i) and our discussion at the beginning, we can assume that there exists a section k so that $x_*(\epsilon) \leq x_{k-(w-1)}$. Consider sections $x_{k-(w-1)}, \dots, x_{k+(w+1)}$, so that in addition $x_{k+(w-1)} \leq x_u(\epsilon)$. If no such k exists then there are at most $2w-1$ points in the interval $[x_*(\epsilon), x_u(\epsilon)]$, and the statement is correct *a fortiori*.

From (28), we know that we have to lower bound $-\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i)$. Since by assumption $x_{k+(w-1)} \leq x_u(\epsilon)$, it follows that $\bar{x}_k \leq x_u(\epsilon)$, so that every contribution in the sum $-\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i)$ is positive. Further, by (the spacing) Lemma 25, $w(\bar{x}_i - \bar{x}_{i-1}) \leq 1$. Hence

$$-\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i) \geq -6\sum_{i=-\infty}^{k} h(\bar{x}_i)(\bar{x}_i - \bar{x}_{i-1}). \tag{29}$$

Since by assumption $x_*(\epsilon) \leq x_{k-(w-1)}$, it follows that $\bar{x}_k \geq x_*(\epsilon)$ and by definition $x_{-\infty} = 0$. Finally, according to Lemma 32 part iii), $-h(x) \geq \kappa_*(\epsilon)x$ for $x \in [0, x_*(\epsilon)]$. Hence

$$\begin{aligned} -6\sum_{i=-\infty}^{k} h(\bar{x}_i)(\bar{x}_i - \bar{x}_{i-1}) &\geq 6\kappa_*(\epsilon)\int_0^{\frac{x_*(\epsilon)}{2}} x dx \\ &= \frac{3}{4}\kappa_*(\epsilon)(x_*(\epsilon))^2. \end{aligned} \tag{30}$$

The inequality in (30) follows since there must exist a section with value greater than $\frac{x_*(\epsilon)}{2}$ and smaller than $x_*(\epsilon)$. Indeed, suppose, on the contrary, that there is no section with value between $\left(\frac{x_*(\epsilon)}{2}, x_*(\epsilon)\right)$. Since $\bar{x}_k \geq x_*(\epsilon)$, we must then have that $\bar{x}_k - \bar{x}_{k-1} > \frac{x_*(\epsilon)}{2}$. But by the spacing Lemma 25 we have that $\bar{x}_k - \bar{x}_{k-1} \leq \frac{1}{w}$. This would imply that $\frac{1}{w} > \frac{x_*(\epsilon)}{2}$. In other words, $w < \frac{2}{x_*(\epsilon)}$. Using the universal lower bound on $x_*(\epsilon)$ from Lemma 32 part iv), we conclude that $w < 2d_l^2 d_r^2$, a contradiction to the hypothesis of the lemma.

Combined with (28) this implies that

$$x_{k+(w-1)} - x_{k-(w-1)} \geq \frac{3}{4}\kappa_*(\epsilon)(x_*(\epsilon))^2.$$

We summarize. The total distance we have to cover is $x_u(\epsilon) - x_*(\epsilon)$ and every $2(w-1)$ steps we cover a distance of at least $\frac{3}{4}\kappa_*(\epsilon)(x_*(\epsilon))^2$ as long as we have not surpassed $x_u(\epsilon)$. Allowing for $2(w-1)-1$ extra steps to cover the last part, bounding again $w-1$ by w, bounding $x_u(\epsilon) - x_*(\epsilon)$ by 1 and replacing $\kappa_*(\epsilon)$ and $x_*(\epsilon)$ by their universal lower bounds, proves the claim.

iii) *From $x_u(\epsilon)$ to $x_u(\epsilon) + \frac{3\kappa_*(x_*)^2}{4(1+12d_l d_r)}$ it takes at most $2w$ sections.*

Let k be the smallest index so that $x_u(\epsilon) \leq x_{k-(w-1)}$. It follows that $\bar{x}_{k-2w+1} \leq x_u(\epsilon) \leq \bar{x}_k$. Let $\hat{k}$ be the largest index so that $\bar{x}_{\hat{k}} \leq x_u(\epsilon)$. From the previous line we deduce that $k - 2w + 1 \leq \hat{k} < k$, so that $k - \hat{k} \leq 2w - 1$. We use again (28). Therefore, let us bound $-\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i)$. We have

$$\begin{aligned} -\frac{6}{w}\sum_{i=-\infty}^{k} h(\bar{x}_i) &= -\frac{6}{w}\sum_{i=-\infty}^{\hat{k}} h(\bar{x}_i) - \frac{6}{w}\sum_{i=\hat{k}+1}^{k} h(\bar{x}_i) \\ &\overset{(a)}{\geq} \frac{3}{4}\kappa_*(\epsilon)(x_*(\epsilon))^2 \\ &\quad - 12 d_l d_r (x_{k+(w-1)} - x_u(\epsilon)). \end{aligned}$$

We obtain (a) as follows. There are two sums, one from $-\infty$ to $\hat{k}$ and another from $\hat{k}+1$ to k. Let us begin with the sum from $-\infty$ to $\hat{k}$. First, we claim that $\bar{x}_{\hat{k}} \geq \frac{x_*(\epsilon)}{2}$.

Indeed, suppose $\bar{x}_{\hat{k}} < \frac{x_*(\epsilon)}{2}$. Then, using the definition of $\hat{k}$

$$\bar{x}_{\hat{k}+1} - \bar{x}_{\hat{k}} > x_u(\epsilon) - \frac{x_*(\epsilon)}{2} \geq \frac{x_u(\epsilon)}{2} \geq \frac{x_u(1)}{2} \overset{(27)}{\geq} \frac{(d_r-1)^{-\frac{d_l-1}{d_l-2}}}{2} \geq \frac{1}{2d_r^2}.$$

But from (the spacing) Lemma 25, $\bar{x}_{\hat{k}+1} - \bar{x}_{\hat{k}} \leq \frac{1}{w}$, a contradiction, since from the hypothesis of the lemma $w \geq 2d_r^2$. Using (29) and (30) with the integral from 0 to $x_*(\epsilon)/2$ we get the first expression in (a). Note that the integral until $x_*(\epsilon)/2$ suffices because either $\bar{x}_{\hat{k}} \leq x_*(\epsilon)$ or, following an argument similar to the one after (30), there must exist a section with value between $\left(\frac{x_*(\epsilon)}{2}, x_*(\epsilon)\right)$.

We now focus on the sum from $\hat{k}+1$ to k. From the definition of $\hat{k}$, for all $i \in [\hat{k}+1, k]$, $|h(\bar{x}_i)| \leq d_l d_r(\bar{x}_i - x_u(\epsilon))$. Indeed, recall from Lemma 32 that $|h'(x)| \leq d_l d_r$ for $x \in [0,1]$. In particular, this implies that the line with slope $d_l d_r$ going through the point $(x_u(\epsilon), 0)$ lies above $h(x)$ for $x \geq x_u(\epsilon)$. Further, $\bar{x}_i - x_u(\epsilon) \leq \bar{x}_k - x_u(\epsilon) \leq x_{k+w-1} - x_u(\epsilon)$. Finally, using $k - \hat{k} \leq 2w-1$, we get the second expression in (a).

From (28), we now conclude that

$$x_{k+w-1} - x_u(\epsilon) \geq \frac{3}{4}\kappa_*(\epsilon)(x_*(\epsilon))^2 - 12 d_l d_r (x_{k+w-1} - x_u(\epsilon))$$

which is equivalent to

$$x_{k+(w-1)} - x_u(\epsilon) \geq \frac{3\kappa_*(\epsilon)(x_*(\epsilon))^2}{4(1+12 d_l d_r)}.$$

The final claim follows by replacing again $\kappa_*(\epsilon)$ and $x_*(\epsilon)$ by their universal lower bounds κ_* and x_*.

iv) *From $x_u(\epsilon) + \frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$ to $x_s(\epsilon) - \delta$ it takes at most $w\frac{1}{\delta \min\{\kappa^{\min}, \lambda^{\min}\}}$ steps, where*

$$\kappa^{\min} = \min_{\epsilon^{\min} \leq \epsilon \leq 1} \kappa^*(\epsilon), \qquad \lambda^{\min} = \min_{\epsilon^{\min} \leq \epsilon \leq 1} \lambda^*(\epsilon).$$

From step iii), we know that within a fixed number of steps we reach at least $\frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$ above $x_u(\epsilon)$. On the other hand, we know from Lemma 23 that $x_0 \leq x_s(\epsilon)$. We conclude that $x_s(\epsilon) - x_u(\epsilon) \geq \frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$. From Lemma 9, we know that $x_s(\epsilon^{\mathrm{BP}}) - x_u(\epsilon^{\mathrm{BP}}) = 0$ and that this distance is strictly increasing for $\epsilon \geq \epsilon^{\mathrm{BP}}$. Therefore, there exists a unique number, call it $\epsilon^{\min}$, $\epsilon^{\min} > \epsilon^{\mathrm{BP}}(d_l, d_r)$, so that

$$x_s(\epsilon) - x_u(\epsilon) \geq \frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$$

if and only if $\epsilon \geq \epsilon^{\min}$. As defined above, let

$$\kappa^{\min} = \min_{\epsilon^{\min} \leq \epsilon \leq 1} \kappa^*(\epsilon), \qquad \lambda^{\min} = \min_{\epsilon^{\min} \leq \epsilon \leq 1} \lambda^*(\epsilon).$$

Since $\epsilon^{\min} > \epsilon^{\mathrm{BP}}(d_l, d_r)$, both $\kappa^{\min}$ and $\lambda^{\min}$ are strictly positive. Using similar reasoning as in step i), we conclude that in order to reach from $x_u(\epsilon) + \frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$ to $x_s(\epsilon) - \delta$ it takes at most $w\frac{x_s(\epsilon) - x_s(\epsilon)}{\delta \min\{\kappa^{\min}, \lambda^{\min}\}}$ steps, where we have used the fact that, by assumption, $\delta \leq \frac{3\kappa_*(x_*)^2}{4(1+12 d_l d_r)}$.

From these four steps, we see that we need at most

$$w\left(\frac{1}{\delta}\left[\frac{1}{\kappa_*} + \frac{1}{\min\{\kappa^{\min}, \lambda^{\min}\}}\right] + \left[\frac{2}{3\kappa_*(x_*)^2} + 5\right]\right) \leq w\frac{1}{\delta}\left[\frac{1}{\kappa_*} + \frac{1}{\min\{\kappa^{\min}, \lambda^{\min}\}} + \frac{2}{3\kappa_*(x_*)^2} + 5\right] \triangleq w\frac{c(d_l, d_r)}{\delta}$$

sections in order to reach $x_s(\epsilon) - \delta$ once we reach δ. This constant depends on (d_l, d_r) but it is independent of L and ϵ.

Appendix IV
Proof of Theorem 27

To establish the existence of $\underline{x}$ with the desired properties, we use the Brouwer FP theorem: it states that every continuous function f from a convex compact subset S of a Euclidean space to S itself has an FP.

Let $\underline{z}$ denote the one-sided forward DE FP for parameter $\epsilon = 1$. Let the length L be chosen in accordance with the statement of the theorem. By assumption $L > \frac{w}{\frac{d_r}{d_l} - 1}$. Using Lemma 22 part ii), we conclude that $\chi(\underline{z}) \geq \frac{1}{2}\left(1 - \frac{d_l}{d_r}\right)$, i.e., $\underline{z}$ is nontrivial. By Lemma 22 part i), it is therefore proper, i.e., it is nondecreasing. Suppose that $\chi(\underline{z}) \leq \chi$. In this case, it is easy to verify that the second statement of the theorem is true. So in the remainder of the proof we assume that $\chi(\underline{z}) > \chi$.

Consider the Euclidean space $[0,1]^{L+1}$. Let $S(\chi)$ be the subspace

$$S(\chi) = \Big\{\underline{x} \in [0,1]^{L+1} : \chi(\underline{x}) = \chi; x_i \leq z_i, \; i \in [-L, 0]; x_{-L} \leq x_{-L+1} \leq \cdots \leq x_0\Big\}.$$

First note that $S(\chi)$ is nonempty since $\underline{z}$ is nontrivial and has entropy at least χ. We claim that $S(\chi)$ is convex and compact. Indeed, convexity follows since $S(\chi)$ is a convex polytope (defined as the intersection of half spaces). Since $S(\chi) \subset [0,1]^{L+1}$ and $S(\chi)$ is closed, $S(\chi)$ is compact.

Note that any constellation belonging to $S(\chi)$ has entropy χ and is increasing, i.e., any such constellation is proper. Our first step is to define a map $V(\underline{x})$, which "approximates" the DE equation and is well-suited for applying the Brouwer FP theorem. The final step in our proof is then to show that the FP of the map $V(\underline{x})$ is in fact an FP of one-sided DE.

The map $V(\underline{x})$ is constructed as follows. For $\underline{x} \in S(\chi)$, let $U(\underline{x})$ be the map

$$(U(\underline{x}))_i = g(x_{i-w+1}, \ldots, x_{i+w-1}), \qquad i \in [-L, 0].$$

Define $V : S(\chi) \to S(\chi)$ to be the map

$$V(\underline{x}) = \begin{cases} U(\underline{x})\dfrac{\chi}{\chi(U(\underline{x}))}, & \chi \leq \chi(U(\underline{x})) \\ \alpha(\underline{x})U(\underline{x}) + (1 - \alpha(\underline{x}))\underline{z}, & \text{otherwise} \end{cases}$$

where

$$\alpha(\underline{x}) = \frac{\chi(\underline{z}) - \chi}{\chi(\underline{z}) - \chi(U(\underline{x}))}.$$

Let us show that this map is well defined. First, consider the case $\chi \le \chi(U(\underline{x}))$. Since $\underline{x} \in S(\chi)$, $\underline{x} \le \underline{z}$ (componentwise). By construction, it follows that $U(\underline{x}) \le U(\underline{z}) = \underline{z}$, where the last step is true since $\underline{z}$ is the forward FP of DE for $\epsilon = 1$. We conclude that $U(\underline{x})\frac{\chi}{\chi(U(\underline{x}))} \le \underline{z}$. Further, by construction $\chi\left(U(\underline{x})\frac{\chi}{\chi(U(\underline{x}))}\right) = \chi$. It is also easy to check that $U(\underline{x})$ is nonnegative and that it is nondecreasing. It follows that in this case $V(\underline{x}) \in S(\chi)$.

Consider next the case $\chi > \chi(U(\underline{x}))$. As we have seen, $\underline{x} \le \underline{z}$ so that $\chi(U(\underline{x})) \le \chi(U(\underline{z})) = \chi(\underline{z})$. Together with $\chi > \chi(U(\underline{x}))$ this shows that $\alpha(\underline{x}) \in [0,1]$. Further, the choice of $\alpha(\underline{x})$ guarantees that $\chi(V(\underline{x})) = \chi$. It is easy to check that $V(\underline{x})$ is increasing and bounded above by $\underline{z}$. This shows that also in this case $V(\underline{x}) \in S(\chi)$.

We summarize that V maps $S(\chi)$ into itself.

In order to be able to invoke Brouwer's theorem we need to show that $V(\underline{x})$ is continuous. This means we need to show that for every $\underline{x} \in S(\chi)$ and for any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $\underline{y} \in B(\underline{x}, \delta) \cap S(\chi)$ then $\|V(\underline{y}) - V(\underline{x})\|_2 \le \varepsilon$.

First, note that $U(\underline{x})$ and $\chi(\underline{x})$ are continuous maps on $S(\chi)$. As a result, $\chi(U(\underline{x}))$, which is the composition of two continuous maps, is also continuous.

Fix $\underline{x} \in S(\chi)$. We have three cases: i) $\chi(U(\underline{x})) > \chi$, ii) $\chi(U(\underline{x})) < \chi$, and iii) $\chi(U(\underline{x})) = \chi$.

We start with i). Let $\rho = \chi(U(\underline{x})) - \chi$ and fix $\varepsilon > 0$. From the continuity of $\chi(U(\underline{x}))$ we know that there exists a ball $B(\underline{x}, \nu_1)$ of radius $\nu_1 > 0$ so that if $\underline{y} \in B(\underline{x}, \nu_1) \cap S(\chi)$ then $|\chi(U(\underline{x})) - \chi(U(\underline{y}))| \le \rho$, so that $\chi(U(\underline{y})) \ge \chi$. It follows that for those y, $V(\underline{y}) = U(\underline{y})\frac{\chi}{\chi(U(\underline{y}))}$.

For a subsequent argument we will need also a tight bound on $|\chi(U(\underline{x})) - \chi(U(\underline{y}))|$ itself. Let us therefore choose $\gamma = \min\{\varepsilon, \rho\}, \gamma > 0$. And let us choose ν_1 so that if $\underline{y} \in B(\underline{x}, \nu_1) \cap S(\chi)$ then $|\chi(U(\underline{x})) - \chi(U(\underline{y}))| \le \frac{\gamma\chi}{2(L+1)}$, so that $\chi(U(\underline{y})) \ge \chi$.

Further, since $U(\cdot)$ is continuous, there exists $\nu_2 > 0$ such that for all $\underline{y} \in B(\underline{x}, \nu_2) \cap S(\chi)$, $\|U(\underline{x}) - U(\underline{y})\|_2 \le \frac{\varepsilon}{2}$. Choose $\nu = \min\{\nu_1, \nu_2\}$. Then, for all $\underline{y} \in B(\underline{x}, \nu) \cap S(\chi)$

$$\begin{aligned}
\|V(\underline{x}) - V(\underline{y})\|_2 &= \chi \left\| \frac{U(\underline{x})}{\chi(U(\underline{x}))} - \frac{U(\underline{y})}{\chi(U(\underline{y}))} \right\|_2 \\
&\le \chi \left\| \frac{U(\underline{x})}{\chi(U(\underline{x}))} - \frac{U(\underline{y})}{\chi(U(\underline{x}))} \right\|_2 \\
&\quad + \chi \left\| \frac{U(\underline{y})}{\chi(U(\underline{x}))} - \frac{U(\underline{y})}{\chi(U(\underline{y}))} \right\|_2 \\
&\overset{\chi(U(\underline{x})) > \chi}{\le} \|U(\underline{x}) - U(\underline{y})\|_2 \\
&\quad + \frac{\|U(\underline{y})\|_2}{\chi} |\chi(U(\underline{x})) - \chi(U(\underline{y}))| \\
&\le \frac{\varepsilon}{2} + \frac{(L+1)}{\chi} \left|\chi(U(\underline{x})) - \chi(U(\underline{y}))\right| \\
&\le \frac{\varepsilon}{2} + \frac{(L+1)}{\chi} \frac{\gamma\chi}{2(L+1)} \le \varepsilon
\end{aligned}$$

where above we used the bound $\|U(\underline{y})\|_2 \le (L+1)$.

Using similar logic, one can prove ii).

Consider claim iii). In this case $\chi(U(\underline{x})) = \chi$, which implies that $V(\underline{x}) = U(\underline{x})$. As before, there exists $0 < \nu_1$ such that for all $\underline{y} \in B(\underline{x}, \nu_1) \cap S(\chi)$, $\|U(\underline{x}) - U(\underline{y})\|_2 < \frac{\varepsilon}{2}$. Let $\gamma = \min\{\chi(\underline{z}) - \chi, \chi\}$. Since we assumed that $\chi(\underline{z}) > \chi$, we have $\gamma > 0$. Furthermore, there exists $0 < \nu_2$ such that for all $\underline{y} \in B(\underline{x}, \nu_2) \cap S(\chi)$, $|\chi(U(\underline{x})) - \chi(U(\underline{y}))| < \frac{\gamma\varepsilon}{2(L+1)}$. Choose $\nu = \min\{\nu_1, \nu_2\}$. Consider $\underline{y} \in B(\underline{x}, \nu) \cap S(\chi)$. Assume first that $\chi(U(\underline{y})) \ge \chi$. Thus, as before

$$\|V(\underline{x}) - V(\underline{y})\|_2 \le \varepsilon.$$

Now let us assume that $\chi(U(\underline{y})) < \chi$. Then, we have

$$\begin{aligned}
\|V(\underline{x}) - V(\underline{y})\|_2 &= \|U(\underline{x}) - \alpha(\underline{y})U(\underline{y}) - (1 - \alpha(\underline{y}))\underline{z}\|_2 \\
&\le \alpha(\underline{y})\|U(\underline{x}) - U(\underline{y})\|_2 \\
&\quad + |1 - \alpha(\underline{y})|\|U(\underline{x}) - U(\underline{z})\|_2 \\
&\le \frac{\varepsilon}{2} + (L+1)\left|\frac{\chi(U(\underline{y})) - \chi(U(\underline{x}))}{\chi(\underline{z}) - \chi(U(\underline{y}))}\right| \\
&\le \frac{\varepsilon}{2} + \frac{1}{2}\left|\frac{\gamma\varepsilon}{\chi(\underline{z}) - \chi}\right| < \varepsilon
\end{aligned}$$

where above we used: i) $\|U(\underline{x}) - U(\underline{z})\|_2 \le L + 1$, ii) $\chi(U(\underline{y})) < \chi$, and iii) $\chi(U(\underline{x})) = \chi$ (when we explicitly write $|1 - \alpha(\underline{y})|$).

We can now invoke Brouwer's FP theorem to conclude that $V(\cdot)$ has an FP in $S(\chi)$, call it $\underline{x}$.

Let us now show that, as a consequence, either there exists a one-sided FP of DE with parameter $\epsilon = 1$ and entropy bounded between $\frac{(1-\frac{d_l}{d_r})(\chi - x_u(1))}{8} - \frac{d_l w}{2d_r(L+1)}$ and χ, or $\underline{x}$ itself is a proper one-sided FP of DE with entropy χ. Clearly, either $\chi \le \chi(U(\underline{x}))$ or $\chi(U(\underline{x})) < \chi$. In the first case, i.e., if $\chi \le \chi(U(\underline{x}))$, then $\underline{x} = V(\underline{x}) = U(\underline{x})\frac{\chi}{\chi(U(\underline{x}))}$. Combined with the nontriviality of $\underline{x}$, we conclude that $\underline{x}$ is a proper one-sided FP with entropy χ and the channel parameter (given by $\frac{\chi}{\chi(U(\underline{x}))}$) less than or equal to 1. Also, from Lemma 23, we then conclude that the channel parameter is strictly greater than $\epsilon^{\mathrm{BP}}(d_l, d_r)$.

Assume now the second case, i.e., assume that $\chi(U(\underline{x})) < \chi$. This implies that

$$\underline{x} = \alpha(\underline{x})(U(\underline{x})) + (1 - \alpha(\underline{x}))\underline{z}.$$

But since $\underline{x} \le \underline{z}$

$$\alpha(\underline{x})\underline{x} + (1 - \alpha(\underline{x}))\underline{z} \ge \underline{x} = \alpha(\underline{x})(U(\underline{x})) + (1 - \alpha(\underline{x}))\underline{z}.$$

As a result, $\underline{x} \ge (U(\underline{x}))$. We will now show that this implies the existence of a one-sided FP of DE with parameter $\epsilon = 1$ and with entropy bounded between $\frac{(1-\frac{d_l}{d_r})(\chi - x_u(1))}{8} - \frac{d_l w}{2d_r(L+1)}$ and χ.

Let $\underline{x}^{(0)} = \underline{x}$ and define $\underline{x}^{(\ell)} = U(\underline{x}^{(\ell-1)})$, $\ell \ge 1$. By assumption, $\underline{x} \ge U(\underline{x})$, i.e., $\underline{x}^{(0)} \ge \underline{x}^{(1)}$. By induction this implies that $\underline{x}^{(\ell-1)} \ge \underline{x}^{(\ell)}$, i.e, the sequence $\underline{x}^{(\ell)}$ is monotonically decreasing. Since it is also bounded from below, it converges to an FP of DE with parameter $\epsilon = 1$, call it $\underline{x}^{(\infty)}$.

We want to show that $\underline{x}^{(\infty)}$ is nontrivial and we want to give a lower bound on its entropy. We do this by comparing $\underline{x}^{(\ell)}$ with a

constellation that lower bounds $\underline{x}^{(\ell)}$ and which converges under DE to a nontrivial FP.

We claim that at least the last $N = (L+1)\frac{\chi - x_u(1)}{2}$ components of $\underline{x}$ are above $\frac{\chi + x_u(1)}{2}$

$$\chi(L+1) = \chi(\underline{x})(L+1) \leq N + (L+1-N)\frac{\chi + x_u(1)}{2}$$

where on the right-hand side we assume (worst case) that the last N components have height 1 and the previous $(L+1-N)$ components have height $\frac{\chi + x_u(1)}{2}$. If we solve the inequality for N we get $N \geq (L+1)\frac{\chi - x_u(1)}{2 - \chi - x_u(1)} \geq (L+1)\frac{\chi - x_u(1)}{2}$.

Consider standard DE for the underlying regular (d_l, d_r) ensemble and $\epsilon = 1$. We claim that it takes at most m

$$m = \max\left\{\frac{2}{\kappa^*(1)(\chi - x_u(1))}, \frac{2}{\lambda^*(1)\left(1 - \frac{d_l}{d_r}\right)}\right\}.$$

DE steps to go from the value $\frac{\chi + x_u(1)}{2}$ to a value above $\frac{1 + \frac{d_l}{d_r}}{2}$. The proof idea is along the lines used in the proof of Lemma 26. Consider the function $h(x)$ as defined in (2) for $\epsilon = 1$. Note that $x_u(1) < \frac{\chi + x_u(1)}{2}$ and that $\frac{1 + \frac{d_l}{d_r}}{2} < x_s(1) = 1$. Further, the function $h(x)$ is unimodal and strictly positive in the range $(x_u(1), x_s(1))$ and $h(x)$ is equal to the change in x, which happens during one iteration, assuming that the current value is x. If $\frac{\chi + x_u(1)}{2} \geq \frac{1 + \frac{d_l}{d_r}}{2}$ then the statement is trivially true. Otherwise, the progress in each required step is at least equal to

$$\min\left\{h\left(\frac{\chi + x_u(1)}{2}\right), h\left(\frac{1 + \frac{d_l}{d_r}}{2}\right)\right\}$$

$$\geq \min\left\{\kappa^*(1)\left(\frac{\chi + x_u(1)}{2} - x_u(1)\right), \lambda^*(1)\left(1 - \frac{1 + \frac{d_l}{d_r}}{2}\right)\right\}.$$

We use Lemma 32 part vi) to get the last inequality. The claim now follows by observing that the total distance that has to be covered is no more than 1.

Consider the constellation $\underline{y}^{(0)}$, which takes the value 0 for $[-L, -N]$ and the value $\frac{\chi + x_u(1)}{2}$ for $[-N+1, 0]$. By construction, $\underline{y} = \underline{y}^{(0)} \leq \underline{x}^{(0)} = \underline{x}$. Define $\underline{y}^{(\ell)} = U(\underline{y}^{(\ell-1)})$, $\ell \geq 1$. By monotonicity, we know that $U(\underline{y}^{(\ell)}) \leq U(\underline{x}^{(\ell)})$ (and hence $\underline{y}^{(\infty)} \leq \underline{x}^{(\infty)}$). In particular, this is true for $\ell = m$. But note that at least the last $N - wm$ positions of $y^{(m)}$ are above $\frac{1 + \frac{d_l}{d_r}}{2}$. Also, by the choice of L, $N - wm \geq N/2$.

Define the constellation $\underline{v}^{(0)}$, which takes the value 0 for $[-L, -N/2]$ and the value $\frac{1 + \frac{d_l}{d_r}}{2}$ for $[-N/2 + 1, 0]$. Define $\underline{v}^{(\ell)} = \frac{1 + \frac{d_l}{d_r}}{2} U(\underline{v}^{(\ell-1)})$, $\ell \geq 0$. Again, observe that by definition $\underline{v}^{(0)} \leq \underline{y}^{(m)}$ and $\frac{1 + \frac{d_l}{d_r}}{2} \leq 1$, hence we have $\underline{v}^{(\infty)} \leq \underline{y}^{(\infty)}$. From Lemma 22 we know that for a length $N/2 = (L+1)\frac{\chi - x_u(1)}{4}$ and a channel parameter $\frac{1 + \frac{d_l}{d_r}}{2}$ the resulting FP of forward DE has entropy at least

$$\chi' = \frac{1 - \frac{d_l}{d_r}}{4} - \frac{d_l w}{d_r(\chi - x_u(1))(L+1)} > 0.$$

Above, $\chi' > 0$ follows from the first assumption on L in the hypothesis of the theorem. It follows that $\underline{v}^{(\infty)}$ has (unnormalized) entropy at least equal to $\chi'(N/2)$ and therefore normalized entropy at least $\frac{\chi'(\chi - x_u(1))}{4}$.

Since $\underline{x}^{(\infty)} \geq \underline{y}^{(\infty)} \geq \underline{v}^{(\infty)}$, we conclude that $\underline{x}^{(\infty)}$ is a one-sided FP of DE for parameter $\epsilon = 1$ with entropy bounded between $\frac{(1 - \frac{d_l}{d_r})(\chi - x_u(1))}{8} - \frac{d_l w}{2d_r(L+1)}$ and χ.

Appendix V
Proof of Theorem 30

i) *Continuity*: In phases i), ii), and iv), the map is differentiable by construction. In phase iii), the map is differentiable in each "period." Further, by definition of the map, the (sub)phases are defined in such a way that the map is continuous at the boundaries.

ii) Bounds in phase i): Consider $\alpha \in \left[\frac{3}{4}, 1\right]$. By construction of the EXIT curve, all elements $x_i(\alpha)$, $i \in [-L, 0]$, are the same. In particular, they are all equal to $x_0(\alpha)$. Therefore, all values $\epsilon_i(\alpha)$, $i \in [-L + w - 1, 0]$, are identical, and equal to $\epsilon_0(\alpha)$.
For points close to the boundary, i.e., for $i \in [-L, -L + w - 2]$, some of the inputs involved in the computation of $\epsilon_i(\alpha)$ are 0 instead of $x_0(\alpha)$. Therefore, the local channel parameter $\epsilon_i(\alpha)$ has to be strictly bigger than $\epsilon_0(\alpha)$ in order to compensate for this. This explains the lower bound on $\epsilon_i(\alpha)$.

iii) Bounds in phase ii): Let $i \in [-L, 0]$ and $\alpha \in \left[\frac{1}{2}, \frac{3}{4}\right]$. Then

$$x^*_{-L} \leq x_i(\alpha) = \epsilon_i(\alpha) g(x_{i-w+1}(\alpha), \ldots, x_{i+w-1}(\alpha))$$
$$\leq \epsilon_i(\alpha) g(x^*_0, \ldots, x^*_0) = \epsilon_i(\alpha)\frac{x^*_0}{\epsilon(x^*_0)}.$$

This gives the lower bound $\epsilon_i(\alpha) \geq \epsilon(x^*_0)\frac{x^*_{-L}}{x^*_0}$.

iv) Bounds in phase iii): Let $\alpha \in \left[\frac{1}{4}, \frac{1}{2}\right]$ and $i \in [-L, 0]$. Note that $x_0(\frac{1}{2}) = x^*_0$ but that $x_0(\frac{1}{4}) = x^*_{-L'+L}$. The range $\left[\frac{1}{4}, \frac{1}{2}\right]$ is therefore split into $L' - L$ "periods." In each period, the original solution x^* is "moved in" by one segment. Let $p \in \{1, \ldots, L' - L\}$ denote the current period we are operating in. In what follows, we think of p as fixed and consider in detail the interpolation in this period. To simplify our notation, we reparameterize the interpolation so that if α goes from 0 to 1, we moved in the original constellation exactly by one more segment. This alternative parametrization is only used in this section. In part vi), when deriving bounds on ϵ^*, we use again the original parametrization.
Taking this reparameterization into account, for $\alpha \in [0, 1]$, according to Definition 28

$$x_i(\alpha) = \begin{cases} (x^*_{i-p})^{\alpha}\left(x^*_{i-p+1}\right)^{1-\alpha}, & i \in [-L, 0] \\ 0, & i < -L. \end{cases}$$

We remark that $x_i(\alpha)$ decreases with α. Thus, we have for any α, $x_i(1) \le x_i(\alpha) \le x_i(0)$. By symmetry, $x_i(\alpha) = x_{-i}(\alpha)$ for $i \ge 1$.

We start by showing that if $x_i(\alpha) > \gamma$ and $i \in [-L + w - 1, -w + 1]$ then $\epsilon_i(\alpha)/\epsilon^* \le 1 + \frac{1}{w^{1/8}}$. For $\alpha \in [0,1]$, define

$$f_i(\alpha) = \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x_{i-k}(\alpha)\right)^{d_r-1}.$$

Further, define

$$f_i^* = \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+1-k}\right)^{d_r-1}.$$

Note that the values x_i^* in the last definition are the values of the one-sided FP. In particular, this means that for $i \ge 0$ we have $x_i^* = x_0^*$.

From the definition of the EXIT curve, we have

$$\epsilon_i(\alpha) = \frac{x_i(\alpha)}{\left(1 - \frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}(\alpha)\right)^{d_l-1}}. \tag{31}$$

By monotonicity

$$\left(1 - \frac{\sum_{j=0}^{w-1} f_{i+j}(\alpha)}{w}\right)^{d_l-1} \ge \left(1 - \frac{\sum_{j=0}^{w-1} f^*_{i+j-1}}{w}\right)^{d_l-1} = \frac{x^*_{i-p}}{\epsilon^*}.$$

In the first step, we used the fact that $-L + w - 1 \le i \le -w + 1$ and the second step is true by definition.

Substituting this into the denominator of (31) results in

$$\frac{\epsilon_i(\alpha)}{\epsilon^*} \le \left(\frac{x^*_{i-p+1}}{x^*_{i-p}}\right)^{1-\alpha} \le \frac{x^*_{i-p+1}}{x^*_{i-p}} = 1 + \frac{1}{\frac{x^*_{i-p+1}}{(\Delta x^*)_{i-p+1}} - 1}$$

where we defined $(\Delta x^*)_i = x_i^* - x^*_{i-1}$. If we plug the upper bound on $(\Delta x^*)_{i-p+1}$ due to (the spacing) Lemma 25 into this expression, we get

$$\frac{1}{x^*_{i-p+1}/(\Delta x^*)_{i-p+1} - 1} \le \frac{1}{\left(\frac{x^*_{i-p+1}}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \frac{w}{(d_l-1)(d_r-1)} - 1}.$$

By assumption $x_i(\alpha) > \gamma$. But from the monotonicity, we have $x^*_{i-p+1} = x_i(0) \ge x_i(\alpha)$. Thus, $x^*_{i-p+1} > \gamma$. This is equivalent to

$$\left(\frac{x^*_{i-p+1}}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \frac{w}{(d_r-1)(d_l-1)} - 1 \ge w^{1/8}. \tag{32}$$

As a consequence

$$\frac{\epsilon_i(\alpha)}{\epsilon^*} \le 1 + \frac{1}{x^*_{i-p+1}/(\Delta x^*)_i - 1} \le 1 + \frac{1}{w^{1/8}}$$

the promised upper bound.

Let us now derive the lower bounds. First, suppose that $x_i(\alpha) > \gamma$. For $i \in [-L, 0]$, we can use again monotonicity to conclude that

$$x^*_{i-p} \le x_i(\alpha) = \epsilon_i(\alpha)\left(1 - \frac{\sum_{j=0}^{w-1} f_{i+j}(\alpha)}{w}\right)^{d_l-1} \le \epsilon_i(\alpha)\frac{x^*_{i-p+1}}{\epsilon^*}.$$

This proves that

$$\frac{\epsilon_i(\alpha)}{\epsilon^*} \ge \frac{x^*_{i-p}}{x^*_{i-p+1}} = 1 - \frac{(\Delta x^*)_{i-p+1}}{x^*_{i-p+1}}.$$

Note that this sequence of inequalities is true for the whole range $i \in [-L, 0]$. Since $x^*_{i-p+1} = x_i(0) \ge x_i(\alpha)$, we have $x^*_{i-p+1} > \gamma$ and using (32), we have

$$\frac{(\Delta x^*)_{i-p+1}}{x^*_{i-p+1}} \le \frac{1}{1 + w^{1/8}}.$$

As a consequence

$$\frac{\epsilon_i(\alpha)}{\epsilon^*} \ge 1 - \frac{(\Delta x^*)_{i-p+1}}{x^*_{i-p+1}} \ge 1 - \frac{1}{1 + w^{1/8}}.$$

It remains to consider the last case, i.e., we assume that $x_i(\alpha) \le \gamma$. From Lemma 24 part iv), we have

$$\left(x^*_{i-p}/\epsilon^*\right)^{\frac{1}{d_l-1}} \ge \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-1-k}\right)^{d_r-2} \times \frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1} x^*_{i-p+j-k} \ge \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k}\right)^{d_r-2} \times \frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1} x^*_{i-p+j-k}$$

and

$$\left(x^*_{i-p+1}/\epsilon^*\right)^{\frac{1}{d_l-1}} \ge \left(1 - \frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k}\right)^{d_r-2} \times \frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1} x^*_{i-p+1+j-k}.$$

We start with (31). Write $x_i(\alpha)$ in the numerator explicitly as $(x^*_{i-p})^{\alpha}(x^*_{i-p+1})^{1-\alpha}$ and bound each of the two terms by the above expressions. This yields

$$\left(\frac{\epsilon_i(\alpha)}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \geq \left(1-\frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k}\right)^{(d_r-2)} \times \frac{d_r-1}{w^2}\frac{\left(\sum_{j,k=0}^{w-1} x^*_{i-p+j-k}\right)^{\alpha}\left(\sum_{j,k=0}^{w-1} x^*_{i-p+1+j-k}\right)^{1-\alpha}}{1-\frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}(\alpha)}.$$

Applying steps, similar to those used to prove Lemma 24 part ii), to the above denominator, we get

$$1-\frac{1}{w}\sum_{j=0}^{w-1} f_{i+j}(\alpha) \leq \frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1} x_{i+j-k}(\alpha) \leq \frac{d_r-1}{w^2}\sum_{j,k=0}^{w-1}\left(x^*_{i-p+j-k}\right)^{\alpha}\left(x^*_{i-p+1+j-k}\right)^{1-\alpha}.$$

Combining all these bounds and canceling common terms yields

$$\left(\frac{\epsilon_i(\alpha)}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \geq \left(1-\frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k}\right)^{(d_r-2)} \times \frac{\left(\sum_{j,k=0}^{w-1} x^*_{i-p+j-k}\right)^{\alpha}\left(\sum_{j,k=0}^{w-1} x^*_{i-p+1+j-k}\right)^{1-\alpha}}{\sum_{j,k=0}^{w-1}\left(x^*_{i-p+j-k}\right)^{\alpha}\left(x^*_{i-p+1+j-k}\right)^{1-\alpha}}. \tag{33}$$

Applying Holder's inequality[6] we get

$$\frac{\left(\sum_{j,k=0}^{w-1} x^*_{i-p+j-k}\right)^{\alpha}\left(\sum_{j,k=0}^{w-1} x^*_{i-p+1+j-k}\right)^{1-\alpha}}{\sum_{j=0}^{w-1}\sum_{k=0}^{w-1}\left(x^*_{i-p+j-k}\right)^{\alpha}\left(x^*_{i-p+1+j-k}\right)^{1-\alpha}} \geq 1.$$

Putting everything together, we now get

$$\left(\frac{\epsilon_i(\alpha)}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \geq \left(1-\frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k}\right)^{d_r-2}. \tag{34}$$

By assumption $x_i(\alpha) \leq \gamma$. Again from monotonicity, we have $x_i(\alpha) \geq x_i(1) = x^*_{i-p}$. Thus, $x^*_{i-p} \leq \gamma$. Combining this with Lemma 24 part iii) and (19) in the hypothesis of the theorem, we obtain

$$\frac{(d_r-1)(d_l-1)(1+w^{1/8})}{w} \geq \frac{1}{w^2}\sum_{j,k=0}^{w-1} x^*_{i-p+j-k}.$$

Suppose that $x^*_{i-p+w-\lceil w^{7/8}\rceil} > \frac{1}{w^{1/8}}$. Then, from the above inequality, we conclude that

$$\frac{(d_r-1)(d_l-1)(1+w^{1/8})}{w} \geq \frac{1}{w^2 w^{1/8}}\left(1+2+\cdots+w^{\frac{7}{8}}\right)$$

where we set to zero all the terms smaller than $x^*_{i-p+w-\lceil w^{7/8}\rceil}$. Upper bounding $(1+w^{1/8})$ by $2w^{1/8}$, we get

$$4(d_r-1)(d_l-1) \geq w^{1/2}.$$

But this is contrary to the hypothesis of the theorem, $w > 2^4 d_l^2 d_r^2$. Hence, we must have $x^*_{i-p+w-\lceil w^{7/8}\rceil} \leq \frac{1}{w^{1/8}}$. Therefore

$$\frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k} \leq \frac{1}{w}\left(\frac{w-\lceil w^{7/8}\rceil}{w^{1/8}} + \lceil w^{7/8}\rceil + 1\right)$$

where we replace $x^*_{i-p+1},\ldots,x^*_{i-p+w-\lceil w^{7/8}\rceil}$ by $\frac{1}{w^{1/8}}$ and the remaining $\lceil w^{7/8}\rceil + 1$ values by 1. Thus, we have

$$\frac{1}{w}\sum_{k=0}^{w-1} x^*_{i-p+w-k} \leq \frac{4}{w^{1/8}}.$$

Using $w \geq 2^{16}$ and combining everything, we get

$$\left(\frac{\epsilon_i(\alpha)}{\epsilon^*}\right)^{\frac{1}{d_l-1}} \geq \left(1-\frac{4}{w^{1/8}}\right)^{d_r-2}.$$

v) Area under EXIT curve[7]: Consider the set of M variable nodes at position i, $i \in [-L, L]$. We want to compute their associated EXIT integral, i.e., we want to compute $\int_0^1 h_i(\alpha)d\epsilon_i(\alpha)$. We use the technique introduced in [2]. We consider the set of M computation trees of height 2 rooted in all variable nodes at position i, $i \in [-L, L]$. For each such computation tree there are d_l check nodes and $1 + d_l(d_r - 1)$ variable nodes. Each of the leaf variable nodes of each computation tree has a certain position in the range $[i-w+1, i+w-1]$. These positions differ for each computation tree. For each computation tree, assign to its root node the channel value $\epsilon_i(\alpha)$, whereas each leaf variable node at position k "sees" the channel value $x_k(\alpha)$.

In order to compute $\int_0^1 h_i(\alpha)d\epsilon_i(\alpha)$, we proceed as follows. We apply the standard area theorem [13, Th. 3.81], to the M simple codes represented by these M computation trees. Each such code has length $1 + d_l(d_r - 1)$ and d_l (linearly independent) check nodes. As we will discuss shortly, the standard area theorem tells us the value of the sum of the $1 + d_l(d_r - 1)$ individual EXIT integrals associated to a particular code. This sum consists of the EXIT integral of the root node as well as the $d_l(d_r - 1)$ EXIT

[6]For any two n-length real sequences $(a_0, a_1, \ldots, a_{n-1})$ and $(b_0, b_1, \ldots, b_{n-1})$ and two real numbers $p, q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, Holder's inequality asserts that

$$\sum_{k=0}^{n-1} |a_k b_k| \leq \left(\sum_{k=0}^{n-1} |a_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=0}^{n-1} |b_k|^q\right)^{\frac{1}{q}}.$$

[7]A slightly more involved proof shows that the area under the EXIT curve (or more precisely, the value of the EXIT integral) is *equal* to the design rate, assuming that the design rate is defined in an appropriate way (see the discussion on page 5). For our purpose it is sufficient, however, to determine the area up to bounds of order w/L. This simplifies the expressions and the proof.

integrals of the leaf nodes. Assume that we can determine the contributions of the EXIT integrals of the leaf nodes for each computation tree. In this case, we can subtract the average such contribution from the sum and determine the average EXIT integral associated to the root node. In the ensuing argument, we consider a fixed instance of a computation tree rooted in i. We then average over the randomness of the ensemble. For the root node the channel value stays the same for all instances, namely, $\epsilon_i(\alpha)$ as given in Definition 28 of the EXIT curve. Hence, for the root node the average, over the ensemble, is taken only over the EXIT value. Then, exchanging the integral (w.r.t. α) and the average and using the fact that each edge associated to the root node behaves independently, we conclude that the average EXIT integral associated to the root node is equal to $\int_0^1 h_i(\alpha)d\epsilon_i(\alpha)$, the desired quantity. Let us now discuss this program in more detail.

For $i \in [-L+w-1, L-w+1]$, we claim that the average sum of the EXIT integrals associated to any such computation tree is equal to $1+d_l(d_r-2)$. This is true since for i in this range, the positions of all leaf nodes are in the range $[-L, L]$. Now applying the area theorem[8] one can conclude that the average sum of all the $1+d_l(d_r-1)$ EXIT integrals associated to the tree code equals the number of variable nodes minus the number of check nodes: $1+d_l(d_r-1)-d_l = 1+d_l(d_r-2)$.

For $i \in [-L, -L+w-2] \cup [L-w+2, L]$, the situation is more complicated. It can happen that some of the leaf nodes of the computation tree see a perfect channel for all values α since their position is outside $[-L, L]$. These leaf nodes are effectively not present in the code and we should remove them before counting. Although it would not be too difficult to determine the exact average contribution for such a root variable node we only need bounds—the average sum of the EXIT integrals associated to such a root node is at least 0 and at most $1+d_l(d_r-2)$.

We summarize: If we consider all computation trees rooted in all variable nodes in the range $[-L, L]$ and apply the standard area theorem to each such tree, then the total average contribution is at least $M(2L-2w+3)(1+d_l(d_r-2))$ and at most $M(2L+1)(1+d_l(d_r-2))$. From these bounds, we now have to subtract the contribution of all the leaf nodes of all the computation trees and divide by M in order to determine bounds on $\sum_{i=-L}^{L}\int_0^1 h_i(\alpha)d\epsilon_i(\alpha)$.

Consider the expected contribution of the $d_l(d_r-1)$ EXIT integrals of each of the M computation trees rooted at i, $i \in [-L+w-1, L-w+1]$. We claim that this contribution is equal to $Md_l(d_r-1)^2/d_r$. For computation trees rooted in $i \in [-L, -L+w-2] \cup [L-w+2, L]$, on the other hand, this contribution is at least 0 and at most $Md_l(d_r-1)$.

Let us start with computation trees rooted in i, $i \in [-L+w-1, L-w+1]$. Fix i. It suffices to consider in detail one "branch" of a computation tree since the EXIT integral is an expected value and expectation is linear. By assumption the root node is at position i. It is connected to a check node, let us say at position j, $j \in [i, i+w-1]$, where the choice is made uniformly at random. In turn, this check node has (d_r-1) children. Let the positions of these children be $k_1, \ldots, k_{d_r-1}$, where all these indices are in the range $[k-w+1, k]$, and all choices are independent and are made uniformly at random.

Consider now this check node in more detail and apply the standard area theorem to the corresponding parity-check code of length d_r. The message from the root node is $x_i(\alpha)$, whereas the messages from the leaf nodes are $x_{k_l}(\alpha)$, $l = 1, \ldots, d_r-1$, respectively. We know from the standard area theorem applied to this parity-check code of length d_r that the sum of the d_r EXIT integrals is equal to d_r-1. So the average contribution of one such EXIT integral is $(d_r-1)/d_r$, and the average of (d_r-1) randomly chosen such EXIT integrals is $(d_r-1)^2/d_r$. Recalling that so far we only considered one out of d_l branches and that there are M computation trees, the total average contribution of all leaf nodes of all computation trees rooted in i should therefore be $Md_l(d_r-1)^2/d_r$.

Let us now justify why the contribution of the leaf nodes is equal to the "average" contribution. Label the d_r edges of the check node from 1 to d_r, where "1" labels the root node. Further, fix j, the position of the check node. As we have seen, we get the associated channels $(i, k_1, \ldots, k_{d_r-1})$ if we root the tree in position i, connect to check node j, and then connect further to $k_1, \ldots, k_{d_r-1}$. This particular realization of this branch happens with probability w^{-d_r} (given that we start in i) and the expected number of branches starting in i that have exactly the same "type" $(i, k_1, \ldots, k_{d_r-1})$ equals $Md_lw^{-d_r}$. Consider a permutation of $(i, k_1, \ldots, k_{d_r-1})$ and keep j fixed. To be concrete, let us say we consider the permutation $(k_3, i, k_2, \ldots, k_1)$. This situation occurs if we root the tree in k_3, connect to check node j, and then connect further to $i, k_2, \ldots, k_1$. Again, this happens with probability w^{-d_r} and the expected number of such branches is $Md_lw^{-d_r}$. It is crucial to observe that all permutations of $(i, k_1, \ldots, k_{d_r-1})$ occur with equal probability in these computation trees and that all the involved integrals occur for computation graphs that are rooted in a position in the range $[-L, L]$. Therefore, the "average" contribution of the (d_r-1) leaf nodes is just a fraction $(d_r-1)/d_r$ of the total contribution, as claimed. Here, we have used a particular notion of "average." We have averaged not only over various computation trees rooted at position i but also over computation trees rooted, let us say, in position k_l, $l = 1, \ldots d_r-1$. Indeed, we have averaged over an equivalence class given by all permutations of $(i, k_1, \ldots, k_{d_r-1})$, with j, the position of the check node held fixed. Since $i \in [-L+w-1, L-w+1]$, all these quantities are also in the range $[-L, L]$, and so they are included in our consideration.

[8]To be precise, the proof of the area theorem given in [13, Th. 3.81] assumes that the channel value of the root node, call it $\epsilon_i(\alpha)$, stays within the range $[0, 1]$. This does not apply in our setting; for $\alpha \to 0$, $\epsilon_i(\alpha)$ becomes unbounded. Nevertheless, it is not hard to show, by explicitly writing down the sum of all EXIT integrals, using integration by parts and finally using the fact that $(\underline{x}(\alpha), \underline{e}(\alpha))$ is an FP, that the result still applies in this more general setting.

It remains to justify the "average" contributions that we get for computation trees rooted in $i \in [-L, -L+w-2] \cup [L-w+2, L]$. The notion of average is the same as we have used it above. Even though we are talking about averages, for each computation tree, it is clear that the contribution is nonnegative since all the involved channel values $x_k(\alpha)$ are increasing functions in α. This proves that the average contribution is nonnegative. Further, the total uncertainty that we remove by each variable leaf node is at most 1. This proves the upper bound.

We can now summarize. We have

$$\begin{aligned}
&\frac{\sum_{i=-L}^{L} \int_0^1 h_i(\alpha) d\epsilon_i(\alpha)}{2L+1} \\
&\quad \le 1 - \frac{d_l}{d_r} + \frac{2(w-1)}{2L+1} \frac{d_l (d_r-1)^2}{d_r} \\
&\quad \le 1 - \frac{d_l}{d_r} + \frac{w}{L} d_l d_r \\
&\frac{\sum_{i=-L}^{L} \int_0^1 h_i(\alpha) d\epsilon_i(\alpha)}{2L+1} \\
&\quad \ge 1 - \frac{d_l}{d_r} - \frac{2(w-1)}{2L+1} \left(1 + d_l(d_r-1) - \frac{d_l}{d_r} \right) \\
&\quad \ge 1 - \frac{d_l}{d_r} - \frac{w}{L} d_l d_r.
\end{aligned}$$

vi) Bound on ϵ^*: Consider the EXIT function constructed according to Definition 28. Recall that the EXIT value at position $i \in [-L, L]$ is defined by

$$h_i(\alpha) = \left(g(x_{i-w+1}(\alpha), \ldots, x_{i+w-1}(\alpha)) \right)^{\frac{d_l}{d_l - 1}} \tag{35}$$

and the area under the EXIT curve is given by

$$A(d_l, d_r, w, L) = \int_0^1 \frac{1}{2L+1} \sum_{i=-L}^{L} h_i(\alpha) d\epsilon_i(\alpha). \tag{36}$$

As we have just seen, this integral is close to the design rate $R(d_l, d_r, \mathrm{w}, L)$, and from Lemma 3, we know that this design rate converges to $1 - d_l/d_r$ for any fixed w when L tends to infinity.

The basic idea of the proof is the following. We will show that $A(d_l, d_r, w, L)$ is also "close" to $1 - \frac{d_l}{d_r} + p^{\mathrm{MAP}}(x(\epsilon^*))$, where $p^{\mathrm{MAP}}(\cdot)$ is the polynomial defined in Lemma 4. In other words, $x(\epsilon^*)$ must be "almost" a zero of $p^{\mathrm{MAP}}(\cdot)$. But $p^{\mathrm{MAP}}(\cdot)$ has only a single positive root and this root is at $\epsilon^{\mathrm{MAP}}(d_l, d_r)$.

More precisely, we first find upper and lower bounds on $A(d_l, d_r, w, L)$ by splitting the integral (36) into four phases. We will see that the main contribution to the area comes from the first phase and that this contribution is close to $1 - \frac{d_l}{d_r} - p^{\mathrm{MAP}}(x(\epsilon^*))$. For all other phases, we will show that the contribution can be bounded by a function, which does not depend on $(\epsilon^*, \underline{x}^*)$ and which tends to 0 if let w and L tend to infinity.

For $i = \{1, 2, 3, 4\}$, define T_i as

$$T_i = \int_{\frac{4-i}{4}}^{\frac{5-i}{4}} \frac{2}{2L+1} \sum_{i=-L+w-1}^{-w+1} h_i(\alpha) d\epsilon_i(\alpha).$$

Further, let

$$\begin{aligned}
T_5 &= \int_0^1 \frac{2}{2L+1} \sum_{i=-L}^{-L+w-2} h_i(\alpha) d\epsilon_i(\alpha) \\
T_6 &= \int_0^1 \frac{1}{2L+1} \sum_{i=-w+2}^{w-2} h_i(\alpha) d\epsilon_i(\alpha).
\end{aligned}$$

Clearly, $A(d_l, d_r, w, L) = T_1 + T_2 + T_3 + T_4 + T_5 + T_6$. We claim that for $w > \max\{2^4 d_l^2 d_r^2, 2^{16}\}$

$$\begin{aligned}
T_1 &= 1 - \frac{d_l}{d_r} - p^{\mathrm{MAP}}(x_0^*) \\
-d_l d_r \left(x_0^* - x_{-L}^* \right) \le T_2 &\le d_r \left(x_0^* - x_{-L}^* \right) \\
-w^{-\frac{1}{8}} - \frac{2 d_r d_l^2}{w^{\frac{7}{8}} \left(1 - 4 w^{-\frac{1}{8}} \right)^{d_r} \epsilon^{\mathrm{BP}}(d_l, d_r)} \le T_3 &\le 4 d_l d_r w^{-\frac{1}{8}} \\
-d_l d_r x_{-L'+L}^* \le T_4 &\le d_r x_{-L'+L}^* \\
-\frac{d_l w}{L} \le T_5 &\le \frac{w}{L} \\
-\frac{d_l w}{L} \le T_6 &\le \frac{w}{L}.
\end{aligned}$$

If we assume these bounds for a moment, and simplify the expressions slightly, we see that for $w > \max\{2^{16}, 2^4 d_l^2 d_r^2\}$

$$\begin{aligned}
&\left| A(d_l, d_r, w, L) - 1 + \frac{d_l}{d_r} + p^{\mathrm{MAP}}(x_0^*) \right| \\
&\quad \le 4 d_l d_r w^{-\frac{1}{8}} + \frac{2 w d_l}{L} \\
&\qquad + d_l d_r \left(x_{-L'+L}^* + x_0^* - x_{-L}^* \right) \\
&\qquad + \frac{2 d_r d_l^2}{\left(1 - 4 w^{-\frac{1}{8}} \right)^{d_r} \epsilon^{\mathrm{BP}}(d_l, d_r)} w^{-\frac{7}{8}}.
\end{aligned}$$

Now using the bound in part v) on the area under the EXIT curve, we get

$$|p^{\mathrm{MAP}}(x_0^*)| \le c_1(d_l, d_r, w, L)$$

where

$$\begin{aligned}
c_1(d_l, d_r, w, L) = &\, 4 d_l d_r w^{-\frac{1}{8}} + \frac{2 w d_l}{L} + \frac{w d_l d_r}{L} \\
&+ d_l d_r \left(x_{-L'+L}^* + x_0^* - x_{-L}^* \right) \\
&+ \frac{2 d_r d_l^2}{\left(1 - 4 w^{-\frac{1}{8}} \right)^{d_r} \epsilon^{\mathrm{BP}}(d_l, d_r)} w^{-\frac{7}{8}}.
\end{aligned}$$

From this we can derive a bound on ϵ^* as follows. Using Taylor's expansion, we get

$$p^{\mathrm{MAP}}(x_0^*) = p^{\mathrm{MAP}}(x_s(\epsilon^*)) + \left(x_0^* - x_s(\epsilon^*) \right) \left(p^{\mathrm{MAP}}(\eta) \right)'$$

where $(p^{\mathrm{MAP}}(x))'$ denotes the derivative w.r.t. x and $\eta \in (x_0^*, x_s(\epsilon^*))$. From Lemma 4, one can verify that $|(p^{\mathrm{MAP}}(x))'| \le 2d_l d_r$ for all $x \in [0,1]$. Thus

$$|p^{\mathrm{MAP}}(x_s(\epsilon^*))| \le 2d_l d_r |x_0^* - x_s(\epsilon^*)| + c_1(d_l, d_r, w, L).$$

Now using $p^{\mathrm{MAP}}(x_s(\epsilon^{\mathrm{MAP}})) = 0$ and the fundamental theorem of calculus, we have

$$p^{\mathrm{MAP}}(x_s(\epsilon^*)) = -\int_{x_s(\epsilon^*)}^{x_s(\epsilon^{\mathrm{MAP}})} (p^{\mathrm{MAP}}(x))' dx.$$

Further, for a (d_l, d_r)-regular ensemble, we have

$$(p^{\mathrm{MAP}}(x))' = \left(1-(1-x)^{d_r-1}\right)^{d_l} \epsilon'(x)$$

where we recall that $\epsilon(x) = x/(1-(1-x)^{d_r-1})^{d_l-1}$. Next, from Lemma 23, we have that $\epsilon^* > \epsilon^{\mathrm{BP}}$. Thus, $x_s(\epsilon^*) > x^{\mathrm{BP}}$. Also, $\epsilon^{\mathrm{MAP}} > \epsilon^{\mathrm{BP}}$. As a consequence, $(1-(1-x)^{d_r-1})^{d_l} \ge (1-(1-x^{\mathrm{BP}})^{d_r-1})^{d_l}$ and $\epsilon'(x) \ge 0$ for all x in the interval of the above integral.
Combining everything, we get

$$\begin{aligned}
&|p^{\mathrm{MAP}}(x_s(\epsilon^*))| \\
&\quad \ge (1-(1-x^{\mathrm{BP}})^{d_r-1})^{d_l} \left| \int_{x_s(\epsilon^*)}^{x_s(\epsilon^{\mathrm{MAP}})} \epsilon'(x) dx \right| \\
&\quad = (1-(1-x^{\mathrm{BP}})^{d_r-1})^{d_l} \left| \epsilon(x_s(\epsilon^{\mathrm{MAP}})) - \epsilon(x_s(\epsilon^*)) \right|.
\end{aligned}$$

Define

$$\begin{aligned}
c(d_l, d_r, w, L) &= 4d_l d_r w^{-\frac{1}{8}} + \frac{2wd_l}{L} + \frac{wd_l d_r}{L} \\
&+ d_l d_r \left(x^*_{-L'+L} + x_0^* - x^*_{-L}\right) + \frac{2d_r d_l^2}{\left(1-4w^{-\frac{1}{8}}\right)^{d_r}} w^{-\frac{7}{8}}.
\end{aligned}$$

Then, using $\epsilon(x_s(\epsilon^*)) = \epsilon^*$ and $\epsilon(x_s(\epsilon^{\mathrm{MAP}})) = \epsilon^{\mathrm{MAP}}(d_l, d_r)$, the final result is

$$\begin{aligned}
&|\epsilon^{\mathrm{MAP}}(d_l, d_r) - \epsilon^*| \\
&\quad \le \frac{2d_l d_r |x_0^* - x_s(\epsilon^*)| + c(d_l, d_r, w, L)}{\epsilon^{\mathrm{BP}}(d_l, d_r)\left(1-(1-x^{\mathrm{BP}})^{d_r-1}\right)^{d_l}} \\
&\quad \overset{(a)}{=} \frac{2d_l d_r |x_0^* - x_s(\epsilon^*)| + c(d_l, d_r, w, L)}{x^{\mathrm{BP}}(d_l, d_r)\left(1-(1-x^{\mathrm{BP}})^{d_r-1}\right)} \\
&\quad \overset{(b)}{\le} \frac{2d_l d_r |x_0^* - x_s(\epsilon^*)| + c(d_l, d_r, w, L)}{\left(x^{\mathrm{BP}}(d_l, d_r)\right)^2} \\
&\quad \overset{\text{Lemma 7}}{\le} \frac{2d_l d_r |x_0^* - x_s(\epsilon^*)| + c(d_l, d_r, w, L)}{\left(1-(d_l-1)^{-\frac{1}{d_r-2}}\right)^2}.
\end{aligned}$$

To obtain (a), we use that x^{BP} is an FP of standard DE for channel parameter ϵ^{BP}. Also, we use $(1-(1-x^{\mathrm{BP}})^{d_r-1}) \ge x^{\mathrm{BP}}(d_l, d_r)$ to get (b).
It remains to verify the bounds on the six integrals. Our strategy is the following. For $i \in [-L+w-1, -w+1]$, we evaluate the integrals directly in phases i), ii), and iii), using the general bounds on the quantities $\epsilon_i(\alpha)$. For the boundary points, i.e., for $i \in [-L, -L+w-2]$ and $i \in [-w+2, 0]$, as well as for all the positions in phase iv), we use the following crude but handy bounds, valid for $0 \le \alpha_1 \le \alpha_2 \le 1$:

$$\begin{aligned}
\int_{\alpha_1}^{\alpha_2} h_i(\alpha) d\epsilon_i(\alpha) &\le h_i(\alpha_2)\epsilon_i(\alpha_2) - h_i(\alpha_1)\epsilon_i(\alpha_1) \\
&\le x_i(\alpha_2)(g(x_{i-w+1}(\alpha_2), \\
&\qquad \ldots, x_{i+w-1}(\alpha_2))^{\frac{1}{d_l-1}} \\
&\le x_i(\alpha_2) \le 1 \qquad (37) \\
\int_{\alpha_1}^{\alpha_2} h_i(\alpha) d\epsilon_i(\alpha) &\ge -\int_{\alpha_1}^{\alpha_2} \epsilon_i(\alpha) dh_i(\alpha) \\
&\ge -d_l \left\{ (h_i(\alpha_2))^{\frac{1}{d_l}} - (h_i(\alpha_1))^{\frac{1}{d_l}} \right\} \\
&\ge -d_l (h_i(\alpha_2))^{\frac{1}{d_l}} \ge -d_l. \qquad (38)
\end{aligned}$$

To prove (37), use integration by parts to write

$$\begin{aligned}
&\int_{\alpha_1}^{\alpha_2} h_i(\alpha) d\epsilon_i(\alpha) \\
&\qquad = \int_{\alpha_1}^{\alpha_2} d(h_i(\alpha)\epsilon_i(\alpha)) - \int_{\alpha_1}^{\alpha_2} \epsilon_i(\alpha) dh_i(\alpha).
\end{aligned}$$

Now note that $\epsilon_i(\alpha) \ge 0$ and that $h_i(\alpha)$ is an increasing function in α by construction. The second term on the right-hand side of the above equality is therefore negative and we get an upper bound if we drop it. We get the further bounds by inserting the explicit expressions for h_i and ϵ_i and by noting that x_i as well as g are upper bounded by 1. To prove (38), we also use integration by parts, but now we drop the first term. Since $h_i(\alpha)$ is an increasing function in α and it is continuous, it is invertible. We can therefore write the integral in the form $\int_{h_i(\alpha_1)}^{h_i(\alpha_2)} \epsilon_i(h) dh$. Now note that $\epsilon_i(h) h = x_i(h) g^{\frac{1}{d_l-1}}(h) = x_i(h) h^{\frac{1}{d_l}} \le h^{\frac{1}{d_l}}$, where we used the fact that $h = g^{\frac{d_l}{d_l-1}}$ [recall the definition of $g(\ldots)$ from (35)]. This shows that $\epsilon_i(h) \le h^{\frac{1-d_l}{d_l}}$. We conclude that

$$\begin{aligned}
\int_{\alpha_1}^{\alpha_2} \epsilon_i(\alpha) dh_i(\alpha) &\le \int_{h_i(\alpha_1)}^{h_i(\alpha_2)} h^{\frac{1-d_l}{d_l}} dh \\
&= d_l \left\{ h_i(\alpha_2)^{\frac{1}{d_l}} - h_i(\alpha_1)^{\frac{1}{d_l}} \right\} \\
&\le d_l h_i(\alpha_2)^{\frac{1}{d_l}} \le d_l.
\end{aligned}$$

The bounds on T_4, T_5, and T_6 are straightforward applications of (38) and (37). For example, to prove that $T_6 \le \frac{w}{L}$, note that there are $2w-3$ positions that are involved. For each position, we know from (37) that the integral is upper bounded by 1. The claim now follows since $\frac{2w-3}{2L-1} \le \frac{w}{L}$. Using (38) leads to the lower bound. Exactly the same line of reasoning leads to both the bounds for T_5.
For the upper bound on T_4, we use the second inequality in (37). We then bound $x_i(\alpha) \le 1$ and use $h_i(\ldots)^{\frac{1}{d_l}} = g(\ldots)^{\frac{1}{d_l-1}}$; cf., (35). Next, we bound each term in the sum by the maximum term. This maximum is $h_0\left(\frac{1}{4}\right)^{\frac{1}{d_l}}$. This term can further be upper bounded by $1-(1-x^*_{-L'+L})^{d_r-1} \le d_r x^*_{-L'+L}$. Indeed, replace all the x

values in $h_0\left(\frac{1}{4}\right)$ by their maximum, $x^*_{-L'+L}$. The lower bound follows in a similar way using the penultimate inequality in (38).

Let us continue with T_1. Note that for $\alpha \in [3/4, 1]$ and $i \in [-L+w-1, -w+1]$, $\epsilon_i(\alpha) = \frac{x_i(\alpha)}{(1-(1-x_i(\alpha))^{d_r-1})^{d_l-1}}$ and that $h_i(\alpha) = (1-(1-x_i(\alpha))^{d_r-1})^{d_l}$. A direct calculation shows that

$$\begin{aligned} T_1 &= \int_{\frac{3}{4}}^{1} h_i(\alpha) d\epsilon_i(\alpha) \\ &= p^{\mathrm{MAP}}(1) - p^{\mathrm{MAP}}\left(x_i\left(\frac{3}{4}\right)\right) \\ &= 1 - \frac{d_l}{d_r} - p^{\mathrm{MAP}}\left(x_0\left(\frac{3}{4}\right)\right) \\ &= 1 - \frac{d_l}{d_r} - p^{\mathrm{MAP}}\left(x_0^*\right). \end{aligned}$$

Let us now compute bounds on T_2. Using (37), we get

$$\begin{aligned} T_2 &\le \frac{2}{2L+1} \\ &\quad \times \sum_{i=-L+w-1}^{-w+1} \left(h_i(3/4)\epsilon_i(3/4) - h_i(1/2)\epsilon_i(1/2)\right) \\ &\le \left\{x_0^*\left(1-(1-x_0^*)^{d_r-1}\right)\right. \\ &\qquad \left. -x^*_{-L}\left(1-(1-x^*_{-L})^{d_r-1}\right)\right\} \\ &\le d_r\left(x_0^* - x^*_{-L}\right). \end{aligned}$$

To obtain the second inequality, we use $\epsilon_i(\alpha)h_i(\alpha) = x_i(\alpha)(h_i(\alpha))^{\frac{1}{d_l}}$. Using the second inequality of (38), we lower bound T_2 as follows. We have

$$\begin{aligned} T_2 &\ge -\frac{2d_l}{2L+1} \sum_{i=-L+w-1}^{-w+1} \left(h_i(3/4)^{\frac{1}{d_l}} - h_i(1/2)\right)^{\frac{1}{d_l}} \\ &\ge -d_l\left\{\left(1-x^*_{-L}\right)^{d_r-1} - (1-x_0^*)^{d_r-1}\right\} \\ &\ge -d_l d_r\left(x_0^* - x^*_{-L}\right). \end{aligned}$$

To obtain the second inequality, we use $h_i(3/4) = (1-(1-x_0^*)^{d_r-1})^{d_l}$ and $h_i(1/2) \ge (1-(1-x^*_{-L})^{d_r-1})^{d_l}$. It remains to bound T_3. For $i \in [-L+w-1, -w+1]$, consider

$$\int_{\frac{1}{2}}^{\frac{1}{4}} d(h_i(\alpha)\epsilon_i(\alpha)) = \epsilon^*\left(h_i\left(\frac{1}{2}\right) - h_i\left(\frac{1}{4}\right)\right) \quad (39)$$

where we have made use of the fact that for $\alpha = \frac{1}{4}$ and $\alpha = \frac{1}{2}$, $\epsilon_i(\alpha) = \epsilon^*$. To get an upper bound on T_3 write

$$\begin{aligned} &\int_{\frac{1}{4}}^{\frac{1}{2}} \epsilon_i(\alpha)) dh_i(\alpha) \\ &\ge \epsilon^*\left(1-\frac{4}{w^{\frac{1}{8}}}\right)^{(d_r-2)(d_l-1)} \left(h_i\left(\frac{1}{2}\right) - h_i\left(\frac{1}{4}\right)\right). \end{aligned}$$

Here we have used the lower bounds on $\epsilon_i(\alpha)$ in phase iii) from Theorem 30 and the fact that $w > \max\{2^{16}, 2^4 d_l^2 d_r^2\}$. Again using integration by parts, and upper bounding both ϵ^* and $(h_i(1/2) - h_i(1/4))$ by 1, we conclude that

$$\begin{aligned} \int_{\frac{1}{2}}^{\frac{1}{4}} h_i(\alpha) d\epsilon_i(\alpha) &\le 1 - \left(1-\frac{4}{w^{1/8}}\right)^{(d_r-2)(d_l-1)} \\ &\le 4 d_r d_l w^{-1/8}. \end{aligned}$$

Note that the right-hand side is independent of i so that this bound extends directly to the sum, i.e.,

$$T_3 \le 4 d_r d_l w^{-1/8}.$$

For the lower bound we can proceed in a similar fashion. We first apply integration by parts. Again using (39), the first term corresponding to the total derivative can be written as

$$\frac{2}{2L+1} \sum_{-L+w-1}^{-w+1} \epsilon^*\left(h_i\left(\frac{1}{2}\right) - h_i\left(\frac{1}{4}\right)\right).$$

We write the other term in the integration by parts as follows. For every section number $i \in [-L+w-1, -w+1]$, let β_i correspond to the smallest number in $\left[\frac{1}{4}, \frac{1}{2}\right]$ such that $x_i(\beta_i) > \gamma$. Recall the definition of γ from part iv) of Theorem 30. If for any section number i, $x_i\left(\frac{1}{2}\right) > \gamma$, then β_i is well defined and $x_i(\alpha) > \gamma$ for all $\alpha \in \left[\beta_i, \frac{1}{2}\right]$. Indeed, this follows from the continuity and the monotonicity of $x_i(\alpha)$ w.r.t. α. On the other hand, if $x_i\left(\frac{1}{2}\right) \le \gamma$, we set $\beta_i = \frac{1}{2}$. Then, we can write the second term as

$$\begin{aligned} &\frac{-2}{2L+1} \\ &\quad \times \sum_{-L+w-1}^{-w+1} \left(\int_{\frac{1}{4}}^{\beta_i} \epsilon_i(\alpha) dh_i(\alpha) + \int_{\beta_i}^{\frac{1}{2}} \epsilon_i(\alpha) dh_i(\alpha)\right). \end{aligned}$$

We now lower bound the two integrals as follows. For $\alpha \in \left[\beta_i, \frac{1}{2}\right]$, we use the upper bound on $\epsilon_i(\alpha)$ valid in phase iii) from Theorem 30. This gives us the lower bound

$$\begin{aligned} &\frac{-2}{2L+1} \\ &\quad \times \sum_{-L+w-1}^{-w+1} \epsilon^*\left(1+\frac{1}{w^{1/8}}\right)\left(h_i\left(\frac{1}{2}\right) - h_i\left(\frac{1}{4}\right)\right) \end{aligned}$$

where above we used the fact that $h_i(\beta_i) \ge h_i\left(\frac{1}{4}\right)$.

For $\alpha \in \left[\frac{1}{4}, \beta_i\right]$, we use the universal bound $-d_l h_i(\beta_i)^{\frac{1}{d_l}}$ [on $\int_{\frac{1}{4}}^{\beta_i} \epsilon_i(\alpha) dh_i(\alpha)$] stated in (38). Since $1/4 \le \beta_i \le 1/2$, using the lower bound on $\epsilon_i(\beta_i) \ge \epsilon^*(1-4w^{-1/8})^{(d_r-2)(d_l-1)}$ [in phase iii) of Theorem 30], we get

$$\begin{aligned} -d_l h_i(\beta_i)^{\frac{1}{d_l}} &= -d_l\left(\frac{x_i(\beta_i)}{\epsilon_i(\beta_i)}\right)^{\frac{1}{d_l-1}} \\ &\ge -d_l\left(\frac{\gamma^{\frac{1}{d_l-1}}}{\epsilon^{\mathrm{BP}}(d_l,d_r)\left(1-4w^{-\frac{1}{8}}\right)^{d_r}}\right). \end{aligned}$$

Above we use $\epsilon^* \geq \epsilon^{\mathrm{BP}}(d_l, d_r)$, replace $(d_r - 2)$ by d_r and $\left(\epsilon^{\mathrm{BP}}(d_l, d_r)\right)^{\frac{1}{d_l - 1}}$ by $\epsilon^{\mathrm{BP}}(d_l, d_r)$. Putting everything together

$$T_3 \geq 1 - \left(1 + \frac{1}{w^{1/8}}\right) - d_l \left(\frac{\gamma^{\frac{1}{d_l - 1}}}{\epsilon^{\mathrm{BP}} \left(1 - 4w^{-\frac{1}{8}}\right)^{d_r}} \right)$$
$$= -w^{-\frac{1}{8}} - d_l \left(\frac{\gamma^{\frac{1}{d_l - 1}}}{\epsilon^{\mathrm{BP}} \left(1 - 4w^{-\frac{1}{8}}\right)^{d_r}} \right).$$

Since $\gamma^{\frac{1}{d_l - 1}} \leq \frac{2 d_r d_l}{w^{\frac{7}{8}}}$, the final result is

$$T_3 = -w^{-\frac{1}{8}} - \frac{2 d_r d_l^2}{w^{\frac{7}{8}} \left(1 - 4w^{-\frac{1}{8}}\right)^{d_r} \epsilon^{\mathrm{BP}}(d_l, d_r)}.$$

ACKNOWLEDGMENT

The authors would like to thank N. Macris for his help in choosing the title and sharing his insights and the reviewers for their thorough reading and numerous suggestions. They would also like to thank D. J. Costello, Jr., P. Vontobel, and A. R. Iyengar for their many comments and very helpful feedback on an earlier draft. Last but not least, they would like to thank G. D. Forney, Jr. for handling this paper.

REFERENCES

[1] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "Life above threshold: From list decoding to area theorem and MSE," in *Proc. IEEE Inf. Theory Workshop*, San Antonio, TX, 2004 [Online]. Available: http://arxiv.org/abs/cs/0410028

[2] C. Méasson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5277–5307, Dec. 2008.

[3] R. M. Tanner, "Error-correcting coding system," U.S. Patent #4,295,218, 1981.

[4] R. M. Tanner, "Convolutional codes from quasi-cyclic codes: A link between the theories of block and convolutional codes," Univ. California, Santa Cruz, CA, Tech. Rep. UCSC-CRL-87-21, 1987.

[5] A. J. Felström and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2181–2190, Sep. 1999.

[6] K. Engdahl and K. S. Zigangirov, "On the theory of low density convolutional codes I," *Problemy Peredachi Informatsii*, vol. 35, no. 4, pp. 295–310, 1999.

[7] K. Engdahl, M. Lentmaier, and K. S. Zigangirov, "On the theory of low-density convolutional codes," in *Proc. 13th Int. Symp. Appl. Algebra Algebraic Algorithms Error-Correcting Codes*, London, U.K., 1999, pp. 77–86.

[8] M. Lentmaier, D. V. Truhachev, and K. S. Zigangirov, "To the theory of low-density convolutional codes. II," *Probl. Inf. Transm.*, vol. 37, no. 4, pp. 288–306, 2001.

[9] R. M. Tanner, D. Sridhara, A. Sridharan, T. E. Fuja, and D. J. Costello, Jr., "LDPC block and convolutional codes based on circulant matrices," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2966–2984, Dec. 2004.

[10] A. Sridharan, M. Lentmaier, D. J. Costello Jr., and K. S. Zigangirov, "Convergence analysis of a class of LDPC convolutional codes for the erasure channel," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, Oct. 2004, pp. 953–962.

[11] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello Jr., "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.

[12] M. Lentmaier, A. Sridharan, K. S. Zigangirov, and D. J. Costello, Jr., "Terminated LDPC convolutional codes with thresholds close to capacity," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 1372–1376.

[13] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[14] M. Lentmaier and G. P. Fettweis, "On the thresholds of generalized LDPC convolutional codes based on protographs," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, 2010, pp. 709–713.

[15] D. G. M. Mitchell, A. E. Pusane, K. S. Zigangirov, and D. J. Costello, Jr, "Asymptotically good LDPC convolutional codes based on protographs," in *Proc. IEEE Int. Symp. Inf. Theory*, Toronto, ON, Canada, Jul. 2008, pp. 1030–1034.

[16] M. Lentmaier, G. P. Fettweis, K. S. Zigangirov, and D. J. Costello, Jr., "Approaching capacity with asymptotically regular LDPC codes," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, 2009, pp. 173–177.

[17] R. Smarandache, A. Pusane, P. Vontobel, and D. J. Costello, Jr, "Pseudo-codewords in LDPC convolutional codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 1364–1368.

[18] R. Smarandache, A. Pusane, P. Vontobel, and D. J. Costello, Jr, "Pseudocodeword performance analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2577–2598, Jun. 2009.

[19] M. Papaleo, A. Iyengar, P. Siegel, J. Wolf, and G. Corazza, "Windowed erasure decoding of LDPC convolutional codes," in *Proc. IEEE Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010, pp. 78–82.

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

[21] J. Thorpe, "Low-density parity-check (LDPC) codes constructed from protographs," Jet Propulsion Lab., Pasadena, CA, INP Progress Rep. 42-154, 2003.

[22] D. Divsalar, S. Dolinar, and C. Jones, "Constructions of protograph LDPC codes with linear minimum distance," in *Proc. IEEE Int. Symp. Inf. Theory*, Seattle, WA, Jul. 2006, pp. 664–668.

[23] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, Jr, "New families of LDPC block codes formed by terminating irregular protograph-based LDPC convolutional codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, Jun. 2010, pp. 824–828.

[24] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, "The generalized area theorem and some of its consequences," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4793–4821, Nov. 2009.

[25] T. Richardson and R. Urbanke, "Multi-edge type LDPC codes," in *The Workshop Honoring Prof. Bob McEliece on his 60th Birthday*, Pasadena, CA, 2002.

[26] M. Lentmaier, D. G. M. Mitchell, G. P. Fettweis, and D. J. Costello, Jr., "Asymptotically good LDPC convolutional codes with AWGN channel thresholds close to the Shannon limit," in *Proc. 6th Int. Symp. Turbo Codes Iterative Inf. Process.*, Sep. 2010, pp. 324–328.

[27] S. Kudekar, C. Méasson, T. Richardson, and R. Urbanke, "Threshold saturation on BMS channels via spatial coupling," in *Proc. 6th Int. Symp. Turbo Codes Iterative Inf. Process.*, Sep. 2010, pp. 309–313.

Shrinivas Kudekar (S'06–M'10) received the B.Tech. degree from the Electrical Engineering Department, Indian Institute of Technology, Bombay, India, in 2003 and the Ph.D. degree from the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, in June 2009.

Currently, he is pursuing his postdoctoral research at the New Mexico Consortium and CNLS, Los Alamos National Laboratory, Los Alamos, NM.

Thomas J. Richardson (S'01–SM'02–F'07) received the Ph.D. degree from the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, in 1990. His undergraduate and Masters work was done at the University of Toronto, Toronto, ON, Canada.

From 1990 to 2000, he was a member of the Mathematical Research Center in Bell Laboratories, Lucent Technologies (originally AT& T Bell Labs). From 2000 until 2006, he was at Flarion Technologies where he held the title of Vice President and Chief Scientist. Since 2006, he has been employed by Qualcomm Inc., Bridgewater, NJ, through the acquisition of Flarion.

Dr. Richardson is a corecipient of the 2002 IEEE Information Theory Society Best Paper Award and a corecipient of the 2011 IEEE Koji Kobayashi Award.

Rüdiger L. Urbanke received the Diplomingenieur degree from the Vienna Institute of Technology, Vienna, Austria, in 1990 and the M.S. and Ph.D. degrees in electrical engineering from Washington University in St. Louis, St. Louis, MO, in 1992 and 1995, respectively.

From 1995 to 1999, he held a position at the Mathematics of Communications Department, Bell Labs. Since November 1999, he has been a faculty member at the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. He is a coauthor of the book *Modern Coding Theory* (Cambridge, U.K.: Cambridge Univ. Press, 2008).

Dr. Urbanke is a recipient of a Fulbright Scholarship. From 2000 to 2004, he was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and he is currently on the board of the series *Foundations and Trends in Communications and Information Theory*. He is a corecipient of the IEEE Information Theory Society 2002 Best Paper Award, and a corecipient of the 2011 IEEE Koji Kobayashi Award.