Windowed Decoding of Spatially Coupled Codes

Aravind R. Iyengar, *Member, IEEE*, Paul H. Siegel, *Fellow, IEEE*, Rüdiger L. Urbanke, and
Jack Keil Wolf, *Life Fellow, IEEE*

Abstract—**Spatially coupled codes have been of interest recently owing to their superior performance over memoryless binary-input channels. The performance is good both asymptotically, since the belief propagation thresholds approach the Shannon limit, as well as for finite lengths, since degree-2 variable nodes that result in high error floors can be completely avoided. However, to realize the promised good performance, one needs large blocklengths. This in turn implies a large latency and decoding complexity. For the memoryless binary erasure channel, we consider the decoding of spatially coupled codes through a windowed decoder that aims to retain many of the attractive features of belief propagation, while trying to reduce complexity further. We characterize the performance of this scheme by defining thresholds on channel erasure rates that guarantee a target erasure rate. We give analytical lower bounds on these thresholds and show that the performance approaches that of belief propagation exponentially fast in the window size. We give numerical results including the thresholds computed using density evolution and the erasure rate curves for finite-length spatially coupled codes.**

Index Terms—**Belief propagation (BP), erasure channels, iterative decoding, low-density parity-check codes (LDPC), spatial coupling, windowed decoding (WD).**

I. INTRODUCTION

SPARSE graph codes have been of great interest in the coding community for close to two decades, after it was shown that statistical inference techniques on graphical models representing these codes had decoding performance that surpassed that of the best known codes. One class of such codes are low-density parity-check (LDPC) codes, which although

introduced by Gallager in the 1960s [2] were rediscovered in the 1990s after the advent of Turbo Codes [3] and iterative decoding. Luby *et al.* showed [4], [5] that a decoder based on *belief propagation* (BP) [6] had very good performance for these codes over the binary erasure channel (BEC). This superior performance of LDPC codes was shown by Richardson and Urbanke [7] to be true over a broader class of binary-input, memoryless, symmetric-output (BMS) channels. Furthermore, these codes were optimized to approach capacity on many of these BMS channels [8], [9].

The convolutional counterparts of LDPC block codes were first introduced by Felstrom and Zigangirov in [10]. There is considerable literature on the constructions and analysis of these ensembles [11]–[14]. The BP thresholds for these ensembles were reported in [15] and shown to be close to capacity in [16]. In [17], the authors construct regular LDPC convolutional codes based on *protographs* [18] that have BP thresholds close to capacity. In [19], Kudekar *et al.* considered convolutional-like codes which they called spatially coupled codes and showed that the BP thresholds of these codes approached the maximum *a posteriori* (MAP) thresholds of the underlying unstructured ensembles over the BEC. This observation was made for protograph-based generalized LDPC codes in [20]. Evidence for similar results over general BMS channels was given in [21], and proven recently in [22]. Moreover, this phenomenon, termed *threshold saturation*, was shown to be a more generic effect of coupling by showing an improvement in performance of systems based on other graphical models: the random K-SAT, Q-COL problems from computation theory, Curie–Weiss model from statistical mechanics [23], and LDGM and rateless code ensembles [24]. Nonbinary LDPC codes obtained through coupling have also recently been investigated [25].

The good performance of spatially coupled codes is apparent when both the blocklength of individual codes and the coupling length becomes large. However, as either of these parameters becomes large, BP decoding becomes complex. We therefore consider a *windowed decoder* (WD) that exploits the structure of the coupled codes to reduce the decoding complexity while maintaining the advantages of the BP decoder in terms of performance. An additional advantage of the WD is the reduced latency of decoding. The WD scheme studied here is the one used to evaluate the performance of protograph-based codes over erasure channels with and without memory [26]–[28]. The main result of this paper is that the WD thresholds approach the BP thresholds exponentially in the size of the window W. Since the BP thresholds are themselves close to the MAP thresholds for spatially coupled codes, WD thus gives us a way to achieve close to ML performance with complexity reduced further beyond that of the BP decoder.

TABLE I
BP THRESHOLDS $\epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6, \gamma, L)$

$L \backslash \gamma$	2	3	4	5	6
16	0.488079	0.488220	0.489806	0.495671	0.505866
32	0.488079	0.488150	0.488151	0.488164	0.488294
64	0.488078	0.488145	0.488148	0.488149	0.488150
128	0.488075	0.488137	0.488142	0.488144	0.488146

The rest of this paper is organized as follows. Section II gives a brief introduction to spatially coupled codes. In Section III, we discuss the WD scheme. We state here the main result of this paper which we prove in Section IV. We give some numerical results in Section V and conclude in Section VI. Much of the terminology and notation used in the paper is reminiscent of the definitions in [19] and we often refer the reader to this paper.

II. SPATIALLY COUPLED CODES

We describe the (d_l, d_r, γ, L) spatially coupled ensemble that was introduced in [19] in terms of its Tanner graph. There are M variable nodes at each position in $[L] \triangleq \{1, 2, \ldots, L\}$. We will assume that there are $M \frac{d_l}{d_r}$ check nodes at every integer position, but only some of these interact with the variable nodes. The variable (check) nodes at position i constitute the ith *section* of variable (check, resp.) nodes in the code. The L sections of variables are together referred to as the *chain* and L is called the *chain length*. For each of the d_l edges incident on a variable at position i, we first choose a section uniformly at random from the set $\{i, i+1, \ldots, i+\gamma-1\}$, then choose a check uniformly at random from the $M \frac{d_l}{d_r}$ checks in the chosen section, and connect the variable to this check. We refer to the parameter γ as the *coupling length*. It can be shown that this procedure amounts roughly to choosing each of the d_r connections of a check node at position i uniformly and independently from the set $\{i-\gamma+1, i-\gamma+2, \ldots, i\}$. Observe that when $\gamma = 1$, this procedure gives us L copies of the (d_l, d_r)-regular uncoupled ensemble. Since we are interested in coupled ensembles, we will henceforth assume that $\gamma > 1$. Further, we will typically be concerned with this ensemble when $L \gg \gamma$, in which case the *design rate* given by [19]

$$R(d_l, d_r, \gamma, L) = 1 - \frac{d_l}{d_r}\left(1 + O(\frac{\gamma}{L})\right)$$

is close to $1 - \frac{d_l}{d_r}$.

A. BP Performance

In the following, we will briefly state known results that are relevant to this work. See [19] for detailed analysis of the BP performance of spatially coupled codes. The BP performance of the (d_l, d_r, γ, L) spatially coupled ensemble when $M \to \infty$ can be evaluated using *density evolution* (DE). Denote the average erasure probability of a message from a variable node at position i as x_i. We refer to the vector $\underline{x} = (x_1, x_2, \ldots, x_L)$ as the *constellation*.

Definition 1 (BP Forward DE): Consider the BP decoding of a (d_l, d_r, γ, L) spatially coupled code over a BEC with channel erasure rate ϵ. We can write the forward DE (FDE) equation as

follows. Set the initial constellation to be $\underline{x}^{(0)} = (1, 1, \ldots, 1)$ and evaluate the constellations $\{\underline{x}^{(\ell)}\}_{\ell=1}^{\infty}$ according to

$$x_i^{(\ell)} = \begin{cases} 0, \text{if } i \notin [L] \; \forall \; \ell, \\ \epsilon\left(1 - \frac{1}{\gamma}\sum_{j=0}^{\gamma-1}(1 - \frac{1}{\gamma}\sum_{k=0}^{\gamma-1} x_{i+j-k}^{(\ell-1)})^{d_r-1}\right)^{d_l-1}, \\ \text{otherwise.} \end{cases}$$

(1)

This is called the *parallel schedule* of the BP FDE. □

For ease of notation, we will write the equation

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

as

$$x_i = \epsilon g(x_{i-\gamma+1}, \ldots, x_{i+\gamma-1}). \tag{2}$$

It is clear that the function $g(\cdot)$ is monotonic in each of its arguments.

Definition 2 (FP of BP FDE): Consider the parallel schedule of the BP FDE for the (d_l, d_r, γ, L) spatially coupled code over a BEC with erasure rate ϵ. It can be easily seen from the monotonicity of $g(\cdot)$ in (2) that the sequence of constellations $\{\underline{x}^{(\ell)}\}_{\ell=0}^{\infty}$ are ordered as $\underline{x}^{(\ell)} \succeq \underline{x}^{(\ell+1)} \; \forall \; \ell \geq 0$, i.e., $x_i^{(\ell)} \geq x_i^{(\ell+1)} \; \forall \; \ell \geq 0, i \in [L]$ (the ordering is pointwise). Since the constellations are all lower bounded by the all-zero constellation $\underline{0}$, the sequence converges pointwise to a limiting constellation $\underline{x}^{(\infty)}$, called the fixed point (FP) of the FDE. □

It is clear that the FP of FDE $\underline{x}^{(\infty)}$ satisfies

$$x_i^{(\infty)} = \begin{cases} 0, & i \notin [L] \\ \epsilon g(x_{i-\gamma+1}^{(\infty)}, \ldots, x_{i+\gamma-1}^{(\infty)}), & i \in [L]. \end{cases}$$

Definition 3 (BP Threshold): Consider the parallel schedule of the BP FDE for the (d_l, d_r, γ, L) spatially coupled code over a BEC with erasure rate ϵ. The *BP threshold* $\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma, L)$ is defined as the supremum of the channel erasure rates $\epsilon \in [0, 1]$ for which the FP of FDE is the all-zero constellation, i.e., $\underline{x}^{(\infty)} = \underline{0}$. □

Table I gives the BP thresholds evaluated from BP FDE for the $(d_l = 3, d_r = 6, \gamma, L)$ coupled ensemble for a few values of γ and L rounded to the sixth decimal place. The MAP threshold of the underlying (d_l, d_r)-regular ensemble is $\epsilon^{\mathrm{MAP}}(d_l = 3, d_r = 6) \approx 0.488151$. We see from the table that the BP thresholds for (d_l, d_r, γ, L) spatially coupled codes are close to the MAP threshold of the (d_l, d_r)-regular unstructured code ensemble. Note that some of the threshold values in Table I are larger than the MAP threshold of the underlying (d_l, d_r)-regular ensemble. This is because the rates of the spatially coupled ensembles are smaller than the (d_l, d_r)-regular ensemble, and depend on the values of γ and L as stated in the beginning of this section.

It was shown in [19] that the BP thresholds satisfy

$$\lim_{\gamma\to\infty}\lim_{L\to\infty} \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma, L) = \lim_{\gamma\to\infty}\lim_{L\to\infty} \epsilon^{\mathrm{MAP}}(d_l, d_r, \gamma, L)$$
$$= \epsilon^{\mathrm{MAP}}(d_l, d_r).$$

This means that the BP threshold *saturates* to the MAP threshold, and we can obtain MAP performance with the reduced complexity of the BP decoder. Later, when we analyze the WD, we will want to keep the coupling length γ finite and hence will be concerned with the quantity

$$\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) \triangleq \lim_{L \to \infty} \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma, L) \tag{3}$$

as a measure of the performance of the BP decoder. It immediately follows from [19, Th. 10] that

$$\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) \leq \epsilon^{\mathrm{MAP}}(d_l, d_r).$$

III. WINDOWED DECODING

The WD exploits the structure of the spatially coupled codes to break down the BP decoding scheme into a series of suboptimal decoding steps—we tradeoff the performance of the decoder for reduced complexity and decoding latency. When decoding with a window of size W, the WD performs BP over the subcode consisting of the first W sections of the variable nodes and their neighboring check nodes and attempts to decode a subset of symbols (those in the first section) within the window. These symbols that we attempt to decode within a window are referred to as the *targeted symbols*. Upon successful decoding of the targeted symbols (or when a maximum number of iterations have been performed), the window slides over one section and performs BP, attempting to decode the targeted symbols in the window in the new position.

More formally, let $\underline{x}$ be the constellation representing the average erasure probability of messages from variables in each of the Section I through L. Initially, the window consists only of the first W sections in the chain. We will refer to this as the first window configuration, and as the window slides to the right, we will increment the window configuration. In other words, when the window has slid through $(c-1)$ sections to the right (when it consists of sections $c, c+1, \ldots, c+W-1$), it is said to be in the cth window configuration. The cth *window constellation*, denoted $\underline{y}_{\{c\}}$, is the average erasure probability of the variables in the cth window configuration. Thus

$$\underline{y}_{\{c\}} = (y_{1,\{c\}}, y_{2,\{c\}}, \ldots, y_{W,\{c\}}) = (x_c, x_{c+1}, \ldots, x_{c+W-1})$$

for $c \in [L]$, where we assume that $x_c = 0 \; \forall \, c > L$. Thus, the cth window constellation, $\underline{y}_{\{c\}}$, represents the "active" sections within the constellation $\underline{x}$. While referring to the entire constellation after the action of the cth window, we will write $\underline{x}_{\{c\}}$. When the window configuration being considered is clear from the context, with some abuse of notation, we drop the $\{c\}$ from the notation and write $\underline{y} = (y_1, \ldots, y_W)$ to denote the window constellation.

Remark 1 (Note on Notation): When we wish to emphasize the size of the window when we write the constellation, we write $\underline{y}_{\langle W \rangle} = (y_{i,\langle W \rangle}, y_{2,\langle W \rangle}, \ldots, y_{W,\langle W \rangle})$. Note that the window configuration and the window size are specified as subscripts within curly brackets $\{\cdot\}$ and angle brackets $\langle \cdot \rangle$,

respectively. Finally, when the constellation after a particular number of iterations ℓ of DE is to be specified, we write $\underline{y}^{(\ell)} = (y_1^{(\ell)}, y_2^{(\ell)}, \ldots, y_W^{(\ell)})$, where the iteration number appears as a superscript within parentheses $(\cdot)$. Although $\underline{y}_{\{c\},\langle W \rangle}^{(\ell)}$ would be the most general way of specifying the window constellation for the cth window configuration with a window of size W after ℓ iterations of DE, for notational convenience, we will write as few of these parameters as possible based on the relevance to the discussion. □

A. Complexity and Latency

For the BP decoder, the number of iterations required to decode all the symbols in a (d_l, d_r, γ, L) spatially coupled code depends on the channel erasure rate ϵ. Whereas when $\epsilon \in [0, \epsilon^{\mathrm{BP}}(d_l, d_r)]$, this required number of iterations can be fixed to a constant number, and when $\epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma, L)]$, the number of iterations scales as $O(L)$[29]. Therefore, in the waterfall region, the complexity of the BP decoder scales as $O(ML^2)$. For the WD of size W, if we let the number of iterations performed scale as $O(W)$, the overall complexity is of the order $O(MW^2L)$. Thus, for small window sizes $W < \sqrt{L}$, we see that the complexity of the decoder can be reduced. A larger reduction in the complexity is possible if we fix the number of iterations performed within each window.

Another advantage of using the WD is that the decoder only needs to know the symbols in the first W sections of the code to be able to decode the targeted symbols. Therefore, in latency-constrained applications, the decoder can work on the fly, resulting in a latency which is a fraction $\frac{W}{L}$ that of the BP decoder.

B. Asymptotic Performance

The asymptotic performance of the (d_l, d_r, γ, L) spatially coupled ensemble with WD can be analyzed using DE as was done for the BP decoder. We will consider the performance of the ensemble with $M \to \infty$ when the transmission happens over a BEC with channel erasure rate $\epsilon \in [0, 1]$. Further, we will assume that for each window configuration, infinite rounds of message passing are performed.

Definition 4 (WD FDE): Consider the WD of a (d_l, d_r, γ, L) spatially coupled code over a BEC with channel erasure rate ϵ with a window of size W. We can write the FDE equation as follows. Set the initial constellation $\underline{x}_{\{0\}}$ according to

$$x_{i,\{0\}} = \begin{cases} 1, & i \in [L] \\ 0, & i \notin [L]. \end{cases}$$

For every window configuration $c = 1, 2, \ldots, L$, let

$$\underline{y}_{\{c\}}^{(0)} = (x_{c,\{c-1\}}, x_{c+1,\{c-1\}}, \ldots, x_{c+W-1,\{c-1\}})$$

and evaluate the sequence of window constellations $\{\underline{y}_{\{c\}}^{(\ell)}\}_{\ell=1}^{\infty}$ using the update rule

$$y_{i,\{c\}}^{(\ell)} = \epsilon g(y_{i-\gamma+1,\{c\}}^{(\ell-1)}, \ldots, y_{i+\gamma-1,\{c\}}^{(\ell-1)}), i \in [W]$$

where for every ℓ, for indices $i \notin [W]$, we set $y_{i,\{c\}}^{(\ell)} = x_{c+i-1,\{c-1\}}$ on the right-hand side. Then, set $\underline{x}_{\{c\}}$ as

$$x_{i,\{c\}} = \begin{cases} x_{i,\{c-1\}}, & i \neq c \\ y_{1,\{c\}}^{(\infty)}, & i = c. \end{cases}$$

$\square$

Discussion: Note that the constellation $\underline{x}_{\{c\}}$ keeps track of the erasure probabilities of targeted symbols of all window configurations up to the cth, followed by erasure probability of 1 for the variables in sections $c+1$ through L, and zeros for sections outside this range. As defined, $\underline{x}_{\{c\}}$ discards all information obtained by running the WD in its cth configuration apart from the values corresponding to the targeted symbols. In practice, it is more efficient to define

$$x_{i,\{c\}} = \begin{cases} x_{i,\{c-1\}}, & i \notin \{c, c+1, \dots, c+W-1\} \\ y_{i-c+1,\{c\}}^{(\infty)}, & \text{otherwise.} \end{cases}$$

In the sequel, we will stick to Definition 4. We do this for two reasons: first, discarding some information between two window configurations can only perform worse than retaining all the information; and second, this assumption makes the analysis simpler since we then have $\underline{y}_{\{c\}}^{(0)} = \underline{1} \ \forall \ c \in [L]$. $\square$

Definition 4 implicitly assumes that the limiting window constellations $\underline{y}_{\{c\}}^{(\infty)}$ exist. The following guarantees that the updates for $x_{i,\{c\}}$ are well defined.

Lemma 5 (cth Window Configuration FP of FDE): Consider the WD FDE of a (d_l, d_r, γ, L) spatially coupled code over a BEC with erasure rate ϵ with a window of size W. Then, the limiting window constellation $\underline{y}_{\{c\}}^{(\infty)}$ exists for each $c \in [L]$. We refer to this constellation as the cth window configuration FP of FDE.

Proof: As noted earlier, $\underline{y}_{\{c\}}^{(0)} = \underline{1} \ \forall \ c \in [L]$, and $\underline{y}_{\{c\}}^{(0)} = \underline{1} \succeq \underline{\epsilon} \succeq \underline{y}_{\{c\}}^{(1)}$. By induction, from the monotonicity of $g(\cdot)$, this implies that $\underline{y}_{\{c\}}^{(\ell)} \succeq \underline{y}_{\{c\}}^{(\ell+1)} \ \forall \ \ell \geq 0$. Since these constellations are lower bounded by $\underline{0}$, the cth window configuration FP of FDE $\underline{y}_{\{c\}}^{(\infty)}$ exists for every $c \in [L]$. $\blacksquare$

The cth window configuration FP of FDE therefore satisfies

$$y_{i,\{c\}}^{(\infty)} = \begin{cases} x_{c+i-1,\{c-1\}}, & i \notin [W] \\ \epsilon g(y_{i-\gamma+1,\{c\}}^{(\infty)}, \dots, y_{i+\gamma-1,\{c\}}^{(\infty)}), & i \in [W] \end{cases} \quad (4)$$

for every $c \in [L]$. Since the $\underline{x}_{\{0\}}$ vector has nonzero values by definition, from the continuity of the WD FDE equations, so do the vectors $\underline{x}_{\{c\}} \ \forall \ c$. Hence, $\underline{0}$ cannot satisfy (4), i.e., $\underline{0}$ cannot be the cth window configuration FP of FDE. Therefore, $\underline{y}_{\{c\}}^{(\infty)} \succ \underline{0} \ \forall \ c \in [L]$. This means that WD can never reduce the erasure probability of the symbols of a spatially coupled code to zero, although it can be made arbitrarily small by using a large enough window. Therefore, an acceptable target erasure rate δ forms a part of the description of the WD. We say that the WD is successful when $\underline{x}_{\{L\}} \preceq \underline{\delta}$.

Lemma 6 (Maximum of $\underline{x}_{\{L\}}$): The vector $\underline{x}_{\{L\}}$ obtained at the end of WD FDE satisfies $x_{i-1,\{L\}} \leq x_{i,\{L\}} \ \forall \ i \in [L - W - \gamma + 2]$. Moreover, $\exists \ \hat{x} \in [0, 1]$ independent of L such that $x_{i,\{L\}} \leq \hat{x} \ \forall \ i$.

Proof: By definition, $x_{i,\{L\}} = y_{1,\{i\}}^{(\infty)} \ \forall \ i \in [L]$. The claim is true for $i = 1$ since $x_{1,\{L\}} = y_{1,\{1\}}^{(\infty)} \geq 0 = x_{0,\{L\}}$. For the ith window configuration, it is clear from Definition 4 that $\underline{y}_{\{i-1\}}^{(0)} \preceq \underline{y}_{\{i\}}^{(0)}, i \in [L - W - \gamma + 2]$. By induction, from the monotonicity of $g(\cdot)$, it follows that $y_{1,\{i-1\}}^{(\infty)} \leq y_{1,\{i\}}^{(\infty)}$ for i in this range.

For $i > L - W - \gamma + 2$, the aforementioned claim is not valid because we defined $x_{j,\{c\}} = 0$ for $j > L$ and we cannot make use of the monotonicity of $g(\cdot)$ since some arguments (corresponding to sections up to the Lth section) are increasing and others (corresponding to the sections beyond the Lth section) decreasing. Nevertheless, we can still claim that $x_{i,\{L\}} \leq x_{i,\{\infty\}} \ \forall \ i \in \mathbb{N}$ where $\underline{x}_{\{\infty\}}$ is the vector of erasure probabilities obtained after WD for a spatially coupled code with an infinite chain length, i.e., $L = \infty$. For $L = \infty$, the sequence $\{x_{i,\{\infty\}}\}$ is nondecreasing and since the $x_{i,\{\infty\}}$ are probabilities, they are in the bounded, closed interval $[0, 1]$. Consequently, the limit $\lim_{i \to \infty} x_{i,\{\infty\}}$ exists in the interval $[0, 1]$, and $\lim_{i \to \infty} x_{i,\{\infty\}} = \sup_i x_{i,\{\infty\}} \overset{\triangle}{=} \hat{x}$. $\blacksquare$

As a consequence of Lemma 6, we can say that the WD is successful when $\hat{x} \leq \delta$. This definition of the success of WD is independent of the chain length L and allows us to compare the performance of WD to that of the BP decoder through the thresholds defined in (3). Note that although the upper bound for $\hat{x}$ in Lemma 6 is a trivial bound, we will in the following give conditions when $\hat{x}$ can be made smaller than an arbitrarily chosen δ, thereby characterizing the WD thresholds.

Definition 7 (WD Thresholds): Consider the WD of a (d_l, d_r, γ, L) spatially coupled code over a BEC of erasure rate ϵ with a window of size W. The WD threshold $\epsilon^{\mathrm{WD}}(d_l, d_r, \gamma, W, \delta)$ is defined as the supremum of channel erasure rates ϵ for which $\hat{x} \leq \delta$. $\square$

Discussion: Since we defined the WD threshold based on $\hat{x}$, it is clear that this is independent of the chain length L. On the other hand, if we used $\max_{i \in [L]} x_{i,\{L\}} \leq \delta$ as the condition for success of the WD in our definition, we would obtain an L-dependent threshold. But $\hat{x}$ denotes the "worst case" remanent erasure probability after WD, and imposing constraints on $\hat{x}$ therefore guarantees good performance for codes with any L.

Note that keeping $\hat{x} \leq \delta$ is sufficient to guarantee an *a posteriori* erasure probability p_e smaller than δ because

$$p_e = \epsilon \left(\frac{\hat{x}}{\epsilon} \right)^{\frac{d_l}{d_l - 1}} = \hat{x} \left(\frac{\hat{x}}{\epsilon} \right)^{\frac{1}{d_l - 1}} \leq \hat{x} \leq \delta.$$

$\square$

We will now state the main result in this paper and prove it in the following section.

Theorem 8 (WD Threshold Bound): Consider WD of the (d_l, d_r, γ, L) spatially coupled ensemble over the BEC. Then, for a target erasure rate $\delta < \delta_*$, there exists a positive integer

$W_{\min}(\delta)$ such that when the window size $W \geq W_{\min}(\delta)$, the WD threshold satisfies

$$\epsilon^{\mathrm{WD}}(d_l, d_r, \gamma, W, \delta) \geq \Big(1 - \frac{d_l d_r}{2}\delta^{\frac{d_l-2}{d_l-1}}\Big)$$
$$\times \Big(\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) - e^{-\frac{1}{\mathsf{B}}(\frac{W}{\gamma-1} - \mathsf{A}\ln\ln\frac{\mathsf{D}}{\delta} - \mathsf{C})}\Big). \tag{5}$$

Here, $\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}$, and δ_* are strictly positive constants that depend only on the ensemble parameters d_l, d_r, and γ. ∎

Theorem 8 says that the WD thresholds approach the BP threshold $\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)$ defined in (3) at least exponentially fast in the ratio of the size of the window W to the coupling length γ for a fixed target erasure probability $\delta < \delta_*$. Moreover, the sensitivity of the bound to changes in δ is small in the exponent in (5) owing to the $\log\log\frac{1}{\delta}$ factor, but larger in the first term in the product on the right-hand side of (5) where it is roughly linear in δ. However, since we intend to set δ to be very small, e.g., 10^{-15}, the first term does not influence the bound heavily. The requirement that $W \geq W_{\min}(\delta)$ is necessary to keep the term within parentheses in the exponent nonnegative. Therefore, the minimum window size required, $W_{\min}(\delta)$, also depends on the constants A, C, and D and, in turn, on the ensemble parameters d_l, d_r, and γ.

The bound guaranteed by Theorem 8 is fairly loose. Numerical results suggest that the minimum window size $W_{\min}(\delta)$ is actually much smaller than the bound obtained from analysis (cf., Section IV). DE also reveals that for a fixed window size, the WD thresholds are much closer to the BP threshold than the bound obtained from Theorem 8.

We note here that the gap between analytical results and numerical experiments is mainly due to the reliance on bounding the DE function in (2) using the counterpart for regular unstructured LDPC ensembles, which proves to be easier to handle than the multivariate (2) (see, e.g., the bound in (11)). However, the scaling of the WD thresholds with the window size and the target erasure probability seems to be as dictated by the bound in (5), suggesting that Theorem 8 captures the essence of the WD algorithm.

Table II gives the WD thresholds obtained through FDE for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ spatially coupled ensemble for different target erasure rates δ and different window sizes W. These thresholds have been rounded to the sixth decimal point. A few comments are in order. As can be seen from the table, the thresholds are close to $\epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6, \gamma = 3) \approx 0.488137$ even for window sizes that are much smaller than the $W_{\min}(\delta)$ obtained analytically, e.g., $W = 16$. Moreover, the WD thresholds are more sensitive to changes in δ for small window sizes where the bound in Theorem 8 is not valid. It is obvious that the thresholds decrease as δ is decreased. Also note that for a fixed target erasure rate, the window size can be made large enough to make the WD thresholds close to the BP threshold.

IV. PERFORMANCE ANALYSIS

In this section, we prove Theorem 8 in steps. First, we analyze the performance of the first window configuration. We will characterize the first window configuration FP of FDE.

TABLE II
WD THRESHOLDS $\epsilon^{\mathrm{WD}}(d_l = 3, d_r = 6, \gamma = 3, W, \delta)$

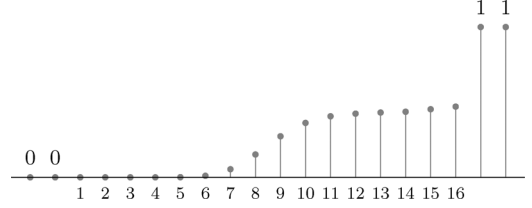| $W\backslash\delta$	10^{-6}	10^{-12}	10^{-18}
4	0.068403	0.000772	0.000008
8	0.472992	0.390749	0.254339
16	0.487504	0.487504	0.487504

Fig. 1. First window configuration FP of FDE for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ ensemble with a window of size $W = 16$ for $\epsilon = 0.48812$. The left and the right boundaries are fixed at 0 and 1, respectively. The sections within the window are indexed from 1 to $W = 16$. The first section has a FP erasure probability $y_{1,\{1\}}^{(\infty)} \approx 2 \times 10^{-15}$.

We will establish that for the variables in the first section of the window, the FP erasure probability can be made small at least double exponentially in the size of the window. We will show that this is possible for all channel erasure rates smaller than a certain ϵ, which we will call the first window threshold $\epsilon^{\mathrm{FW}}(d_l, d_r, \gamma, W, \delta)$, provided the window size is larger than a certain minimum size.

Once we have this, we consider the performance of the cth window configuration for $1 < c \leq L$. In this case also, we will show that the FP erasure probability of the first section within the window is guaranteed to decay double exponentially in the window size. As for the first window configuration, this result holds provided the window size is larger than a certain minimal size and this time the minimal size is slightly larger than the minimal size required for the first window configuration. Moreover, such a result is true for channel erasure rates smaller than a value which is itself smaller than the first window threshold, and this value will be our lower bound for the WD threshold.

A. First Window Configuration

From Definition 4, FDE for the first window configuration amounts to the following. Set $\underline{y}_{\{1\}}^{(0)} = \underline{1}$ and evaluate the sequence of window constellations $\{\underline{y}_{\{1\}}^{(\ell)}\}_{\ell=1}^{\infty}$ according to

$$y_{i,\{1\}}^{(\ell)} = \begin{cases} 0, & i \leq 0 \\ \epsilon g(y_{i-\gamma+1,\{1\}}^{(\ell-1)}, \ldots, y_{i+\gamma-1,\{1\}}^{(\ell-1)}), & i \in [W] \\ 1, & i > W. \end{cases} \tag{6}$$

Since $\underline{y}_{\{1\}}^{(0)}$ is nondecreasing, i.e., $y_{i,\{1\}}^{(0)} \leq y_{i+1,\{1\}}^{(0)} \ \forall\, i$, so is the first window configuration FP, $\underline{y}_{\{1\}}^{(\infty)}$, by induction and monotonicity of $g(\cdot)$.

Fig. 1 shows the first window configuration FP of FDE for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ ensemble with a window of size $W = 16$ for a channel erasure rate $\epsilon = 0.48812$.

The scheduling scheme used in the definition of the window configuration FPs is what is called the *parallel* schedule. In general, we can consider a scheduling scheme where, in each step, a subset of the sections within the window are updated. We say that such an arbitrary scheduling scheme is admissible if every

section is updated infinitely often with the correct boundary conditions, i.e., with the correct values set at the left and the right ends of the window. It is easy to see from the standard argument of nested computation trees (see, e.g., [19]) that the FP is independent of the scheduling scheme.

We know that the first window configuration FP of FDE, $\underline{y}_{\{1\}}^{(\infty)}$, is nondecreasing, i.e., $y_{i,\{1\}}^{(\infty)} \leq y_{i+1,\{1\}}^{(\infty)} \ \forall \ i$. The following shows the ordering of the FP values of individual sections in windows of different sizes. With the understanding that we are considering only the first window configuration in this section, we will drop the window configuration number from the notation for window constellations throughout this section for convenience.

Lemma 9 (FPs and Window Size): Let $\underline{y}_{\langle W \rangle}$ and $\underline{y}_{\langle W+1 \rangle}$ denote the first window configuration FPs of FDE with windows of sizes W and $W+1$, respectively, for $\epsilon \in [0, 1]$. Then

$$y_{i,\langle W \rangle} \geq y_{i,\langle W+1 \rangle} \geq y_{i-1,\langle W \rangle}$$

where $y_{i,\langle W \rangle}$ denotes the FP erasure probability of the ith section in a window of size W.

Proof: Consider the following schedule. Set $\underline{y}_{\langle W+1 \rangle}^{(0)} = (\underline{y}_{\langle W \rangle}, 1)$ and evaluate the sequence of window constellations $\{\underline{y}_{\langle W+1 \rangle}^{(\ell)}\}$ according to (6). Clearly, we have

$$\underline{y}_{\langle W+1 \rangle}^{(1)} \preceq (\underline{y}_{\langle W \rangle}, \epsilon) \preceq (\underline{y}_{\langle W \rangle}, 1) = \underline{y}_{\langle W+1 \rangle}^{(0)}$$

so that the sequence $\{\underline{y}_{\langle W+1 \rangle}^{(\ell)}\}$ is pointwise nonincreasing by induction. We claim that this schedule is admissible. This is true because the DE updates are first performed infinitely many times over the first W sections to obtain $\underline{y}_{\langle W+1 \rangle}^{(0)}$, and then over all the $W+1$ sections infinitely many times again. Therefore, the updates are performed over all sections infinitely often with the correct boundary conditions. The limiting FP must hence be exactly $\underline{y}_{\langle W+1 \rangle}$ and the first inequality in the statement of the lemma holds. Intuitively, this is true because in going from W to $W+1$ and checking the ith section, we have moved further away from the right end of the window (where $y_i = 1$) while remaining at the same distance from the left end (where $y_i = 0$).

To prove the second inequality, consider the following schedule. Set $\underline{y}_{i,\langle W \rangle}^{(0)} = \underline{y}_{i+1,\langle W+1 \rangle}, i = 1, \ldots, W$ and evaluate the sequence of constellations $\{\underline{y}_{\langle W \rangle}^{(\ell)}\}$ according to (6). Since $\underline{y}_{0,\langle W \rangle}^{(0)} = 0 \leq \underline{y}_{1,\langle W+1 \rangle}$, we must have $\underline{y}_{\langle W \rangle}^{(1)} \preceq \underline{y}_{\langle W \rangle}^{(0)}$ and by induction the sequence of constellations thus obtained is also pointwise nonincreasing. Again we claim that the aforementioned schedule is admissible. This is true because we first update all W sections within the window and also the zeroth section infinitely often, and then set the boundary condition that the zeroth section also has all variables completely known. In all, every section within the window gets updated infinitely often with the correct boundary conditions. The limiting FP must hence be exactly $\underline{y}_{\langle W \rangle}$ and the second inequality claimed in the statement of the lemma follows. As in the previous case, this is intuitively true because in going from the $(i+1)$th section with window size $W+1$ to the ith section with window

size W, we have moved closer to the left end of the window while maintaining the distance from the right end. ∎

We now give some bounds on the FP erasure probabilities of individual sections within a window.

Lemma 10 (Bounds on FP): Consider the WD of the (d_l, d_r, γ, L) ensemble with a window of size W over a channel with erasure rate ϵ and $d_l \geq 3$. The first window configuration FP $\underline{y}$ satisfies

$$y_i \geq \left(\epsilon (\frac{\gamma-1}{2\gamma})^{d_l-1} \right)^{\frac{(d_l-1)^j-1}{d_l-2}} y_{i+j}^{(d_l-1)^j}$$

$$y_i \leq \epsilon \left(1 - \alpha_k (1 - y_{i+k})^{d_r-1} \right)^{d_l-1}$$

for $i \in [1, W], j \in [0, W+1-i], k \in [0, \gamma-1]$, where $\alpha_k = (1 - \frac{(\gamma-k-1)(\gamma-k)}{2\gamma^2})^{d_r-1}$. ∎

We relegate the proof to Appendix I. The following shows that once the FP erasure probability of a section within the window is smaller than a certain value, it decays very quickly as we move further to the left in the window.

Lemma 11 (Double Exponential Tail of the FP): Consider WD of the (d_l, d_r, γ, L) ensemble with a window of size W over a channel with erasure rate $\epsilon \in (0, 1)$. Let $d_l \geq 3$ and let $\underline{y}$ be the first window configuration FP of FDE. If there exists an $i \in [W]$ such that $y_i < \delta_0 \triangleq \left((d_r-1)^{\frac{d_l-1}{d_l-2}} \right)^{-1}$, then

$$y_{i-j(\gamma-1)} \leq \Psi e^{-\psi(d_l-1)^j}$$

where $\Psi = \delta_0 \epsilon^{\frac{-1}{d_l-2}}$ and $\psi = \ln(\frac{\Psi}{\delta_0}) = \frac{1}{d_l-2} \ln \frac{1}{\epsilon} > 0$.

Proof: Since the FP is nondecreasing, we have

$$
\begin{aligned}
y_{i-(\gamma-1)} &= \epsilon g(y_{i-2(\gamma-1)}, \ldots, y_i) \\
&\leq \epsilon g(y_i, y_i, \ldots, y_i) \\
&= \epsilon (1 - (1 - y_i)^{d_r-1})^{d_l-1} \quad (7) \\
&\leq \epsilon ((d_r-1)y_i)^{d_l-1}
\end{aligned}
$$

which can be applied recursively to obtain

$$
\begin{aligned}
y_{i-j(\gamma-1)} &\leq \epsilon^{\frac{(d_l-1)^j-1}{d_l-2}} (d_r-1)^{\frac{d_l-1}{d_l-2}((d_l-1)^j-1)} y_i^{(d_l-1)^j} \\
&< \epsilon^{\frac{(d_l-1)^j-1}{d_l-2}} (d_r-1)^{\frac{d_l-1}{d_l-2}((d_l-1)^j-1)} \delta_0^{(d_l-1)^j} \\
&\triangleq \Psi e^{-\psi(d_l-1)^j}
\end{aligned}
\quad (8)
$$

where Ψ and ψ are as defined in the statement. It is worthwhile to note that δ_0 is a lower bound on the *breakout value* for the (d_l, d_r)-regular ensemble [30]. The emergence of the breakout value in this context is not entirely unexpected since it is known that for the (d_l, d_r)-regular ensemble, the erasure probability decays double exponentially in the number of iterations below the breakout value, and in case of spatially coupled ensembles, the counterpart for the number of iterations is the number of sections (cf. (7)). ∎

We now show that the FP erasure probability of a message from a variable node in the first section, y_1, can be made small by increasing the window size W for any $\epsilon < \epsilon^{BP}(d_l, d_r, \gamma)$. Assuming that the window size is "large enough," we will count

the number of sections, starting from the right, that have a FP erasure probability larger than a small δ for a channel erasure rate $\epsilon = \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) - \Delta\epsilon$.

Definition 12 (Transition Width): Consider WD of a (d_l, d_r, γ, L) spatially coupled code over a BEC of erasure rate ϵ. Let $\underline{y}$ be the first window configuration FP of FDE. Then, we define the transition width $\tau(\epsilon, \delta)$ of $\underline{y}$ as

$$\tau(\epsilon, \delta) = |\{i \in [W] : \delta < y_i \leq 1\}|.$$

$\square$

Note that from the definition of the transition width, it depends on the window size W. We first upper bound $\tau(\epsilon, \delta) \leq \hat{\tau}(\epsilon, \delta)$ such that the upper bound is independent of the window size W, and then claim from Lemma 9 that by employing a window whose size is larger than $\hat{\tau}(\epsilon, \delta)$, we can guarantee $y_1 \leq \delta$.

Definition 13 (First Window Threshold): Consider WD of the (d_l, d_r, γ, L) spatially coupled ensemble with a window of size W over a BEC with erasure rate ϵ. The first window threshold $\epsilon^{\mathrm{FW}}(d_l, d_r, \gamma, W, \delta)$ is defined as the supremum of channel erasure rates for which the first window configuration FP of FDE $\underline{y}$ satisfies $y_1 \leq \delta$. $\square$

From Definitions 12 and 13, we can see that by ensuring that $W > \hat{\tau}(\epsilon, \delta)$, we can bound $\epsilon^{\mathrm{FW}}(d_l, d_r, \gamma, W, \delta) \geq \epsilon$.

Proposition 14 (Maximum Transition Width): Consider the first window configuration FP of FDE $\underline{y}$ for the (d_l, d_r, γ, L) spatially coupled ensemble with a window of size $W < L$ for $\epsilon \in [\frac{\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) + \epsilon^{\mathrm{BP}}(d_l, d_r)}{2}, \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)) \triangleq \mathscr{E}$. Then

$$\tau(\epsilon, \delta) \leq (\gamma - 1)\left(\mathsf{A} \ln \ln \frac{\mathsf{D}}{\delta} + \mathsf{B} \ln \frac{1}{\Delta\epsilon} + \hat{\mathsf{C}}\right) \triangleq \hat{\tau}(\epsilon, \delta)$$

provided $\delta \leq \delta_0$. Here, $\Delta\epsilon = \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) - \epsilon$, and $\mathsf{A}, \mathsf{B}, \hat{\mathsf{C}}, \mathsf{D}$, and δ_0 are strictly positive constants that depend only on the ensemble parameters d_l, d_r, and γ. ∎

The proof is given in Appendix II. This means that the smallest window size that guarantees $y_1 \leq \delta$ for a channel erasure rate $\frac{\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) + \epsilon^{\mathrm{BP}}(d_l, d_r)}{2}$ is

$$\hat{W}_{\min}(\delta) = \left\lfloor (\gamma - 1)\left(\mathsf{A} \ln \ln \frac{\mathsf{D}}{\delta} + \mathsf{B} \ln \frac{1}{\Delta\epsilon_{\max}} + \hat{\mathsf{C}}\right)\right\rfloor + 1$$
$$> \hat{\tau}\left(\frac{\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) + \epsilon^{\mathrm{BP}}(d_l, d_r)}{2}, \delta\right)$$

where $\Delta\epsilon_{\max} = \frac{\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) - \epsilon^{\mathrm{BP}}(d_l, d_r)}{2}$. When $W \geq \hat{W}_{\min}(\delta)$, we have

$$\epsilon^{\mathrm{FW}}(d_l, d_r, \gamma, W, \delta) \geq \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)$$
$$- e^{-\frac{1}{\mathsf{B}}\left(\frac{W}{\gamma - 1} - \mathsf{A} \ln \ln \frac{\mathsf{D}}{\delta} - \hat{\mathsf{C}}\right)}. \quad (9)$$

Discussion: We restricted $\epsilon \in \mathscr{E}$ in Proposition 14 to obtain constants that are independent of ϵ. As can be seen from the proof of the proposition, these constants are dependent on ϵ, unless each is optimized in the range $\mathscr{E}$. As we let the minimum ϵ in $\mathscr{E}$ approach $\epsilon^{\mathrm{BP}}(d_l, d_r)$, the constants in the expression for

$\hat{\tau}(\epsilon, \delta)$ blow up and the upper bound will be useless. It is therefore necessary to keep the minimum of $\epsilon \in \mathscr{E}$ strictly larger than $\epsilon^{\mathrm{BP}}(d_l, d_r)$ and the value chosen in the above was motivated by our intent to ensure that the first window threshold was closer to $\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)$ than to $\epsilon^{\mathrm{BP}}(d_l, d_r)$. Note that the increase in the upper bound for $\tau(\epsilon, \delta)$ with decrease in ϵ is purely an artifact of the upper bounding technique we have employed, i.e., it is obvious that as we decrease ϵ, $\tau(\epsilon, \delta)$ also decreases. $\square$

B. *c*th *Window Configuration,* $1 < c \leq L$

We now evaluate the performance of the WD scheme when the window has slid a certain number of sections from the left end of the code. We arrive at conditions under which $\hat{x}$ is guaranteed to be smaller than δ while operating with a window of size W. We start by establishing a property of $\hat{x}$.

Lemma 15 (FP Equation Involving $\hat{x}$): Consider the function $\Omega(\underline{y})$ where

$$\Omega(y_i) = \begin{cases} \pi(y_i), & i < 1 \\ \epsilon g(\pi(y_{i-\gamma+1}), \ldots, \pi(y_{i+\gamma-1})), & i \in [W] \\ 1, & i > W \end{cases}$$

where

$$\pi(y_i) = \begin{cases} y_1, & i < 1 \\ y_i, & i \in [W] \\ 1, & i > W. \end{cases}$$

Then, there exists a solution $\underline{\omega}$ to the equation $\underline{y} = \Omega(\underline{y})$ such that $\omega_1 = \hat{x}$. Moreover, $\underline{\omega}$ is the smallest such constellation, i.e., if $\underline{\hat{\omega}} = \Omega(\underline{\hat{\omega}})$, then $\underline{\hat{\omega}} \succeq \underline{\omega}$.

Proof: We have

$$\hat{x} = x_{\infty, \{\infty\}} = y_{1, \{\infty\}}^{(\infty)}$$
$$\overset{(4)}{=} \epsilon g(y_{-\gamma+2, \{\infty\}}^{(\infty)}, \ldots, y_{0, \{\infty\}}^{(\infty)}, y_{1, \{\infty\}}^{(\infty)}, \ldots, y_{\gamma, \{\infty\}}^{(\infty)})$$
$$\overset{(4)}{=} \epsilon g(\underbrace{x_{\infty, \{\infty\}}, \ldots, x_{\infty, \{\infty\}}}_{\gamma}, y_{2, \{\infty\}}^{(\infty)}, \ldots, y_{\gamma, \{\infty\}}^{(\infty)})$$
$$= \epsilon g(\underbrace{\hat{x}, \ldots, \hat{x}}_{\gamma}, y_{2, \{\infty\}}^{(\infty)}, \ldots, y_{\gamma, \{\infty\}}^{(\infty)}).$$

Hence, if we define $\underline{\omega}$ as follows:

$$\omega_i = \begin{cases} \hat{x} = y_{1, \{\infty\}}^{(\infty)}, & i \leq 1 \\ y_{i, \{\infty\}}^{(\infty)}, & i > 1 \end{cases}$$

then it is clear that $\underline{\omega} = \pi(\underline{\omega}) = \Omega(\underline{\omega})$.

Note that any FP $\underline{\hat{\omega}}$ of the function $\Omega(\cdot)$ has to satisfy $\underline{\hat{\omega}} \succeq \underline{y}_{\{1\}}^{(\infty)}$ for the same channel erasure rate $\epsilon \in [0, 1]$ from the monotonicity of $g(\cdot)$. In particular, $\underline{\omega} \succeq \underline{y}_{\{1\}}^{(\infty)}$. From the continuity of the DE equations in Definition 4, it follows that $\underline{\omega}$ is the least solution to the equation $\underline{y} = \Omega(\underline{y})$, since it is the limiting constellation of the sequence of nondecreasing constellations $\{\underline{y}_{\{n\}}^{(\infty)}\}_{n=1}^{\infty}$. ∎

We defer the proof of the following proposition to Appendix III, which is the central argument in the proof of Theorem 8. Using the bound on the maximum transition width from Proposition 14, we obtain an upper bound on $\hat{x}$ for
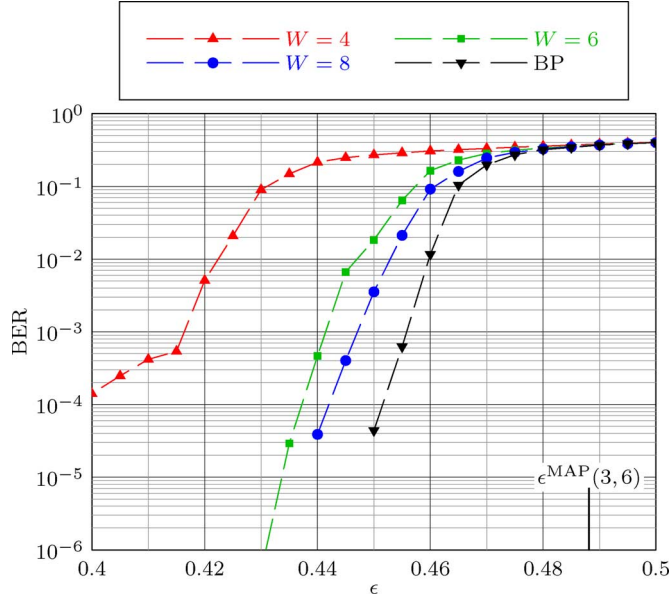
Fig. 2. Bit erasure probability of the $(d_l = 3, d_r = 6, \gamma = 3, L = 64)$ spatially coupled code with $M = 1024$ achieved with a WD of window sizes $W = 4, 6$, and 8.

a given window size W and erasure rate $\epsilon \in \mathscr{E}$. From this, we arrive at a lower bound for ϵ that guarantees $\hat{x} \leq \delta$ when δ is an arbitrarily chosen value smaller than δ_* (which depends only on d_l, d_r) and the window size is larger than $W_{\min}(\delta)$ (which depends only on the code parameters d_l, d_r, γ, and δ). This gives us our lower bound on the WD threshold.

Proposition 16 (WD and FW Thresholds): Consider WD of the (d_l, d_r, γ, L) spatially coupled ensemble with a window of size $W \geq W_{\min}(\delta) = \hat{W}_{\min}(\delta) + \gamma - 1$ over a BEC with erasure rate ϵ. Then, we have

$$\epsilon^{\text{WD}}(d_l, d_r, \gamma, W, \delta) \geq \left(1 - \frac{d_l d_r}{2} \delta^{\frac{d_l-2}{d_l-1}}\right)$$
$$\times \epsilon^{\text{FW}}(d_l, d_r, \gamma, W - \gamma + 1, \delta)$$

provided $\delta < \delta_* = \left(\frac{2}{d_l d_r}\right)^{\frac{d_l-1}{d_l-2}}$, where $\epsilon^{\text{FW}}(d_l, d_r, \gamma, W, \delta)$ is the first window threshold. ∎

From Proposition 16 and (9), we immediately have that

$$\epsilon^{\text{WD}}(d_l, d_r, \gamma, W, \delta) \geq \left(1 - \frac{d_l d_r}{2} \delta^{\frac{d_l-2}{d_l-1}}\right)$$
$$\times \left(\epsilon^{\text{BP}}(d_l, d_r, \gamma) - e^{-\frac{1}{\delta}\left(\frac{W-\gamma+1}{\gamma-1} - A \ln \ln \frac{D}{\delta} - \hat{C}\right)}\right)$$

provided $W \geq W_{\min}(\delta)$. By making the substitution $\mathsf{C} = \hat{\mathsf{C}} + 1$, we see that this proves Theorem 8.

V. EXPERIMENTAL RESULTS

In this section, we give results obtained by simulating WD of finite-length spatially coupled codes over the BEC. The code used for simulation was generated randomly by fixing the parameters $M = 1024$, $d_l = 3$, $d_r = 6$, with coupling length $\gamma = 3$ and chain length $L = 64$. The blocklength of the code was hence $n = ML = 65\,536$ and the rate was $R \approx 0.484375$.
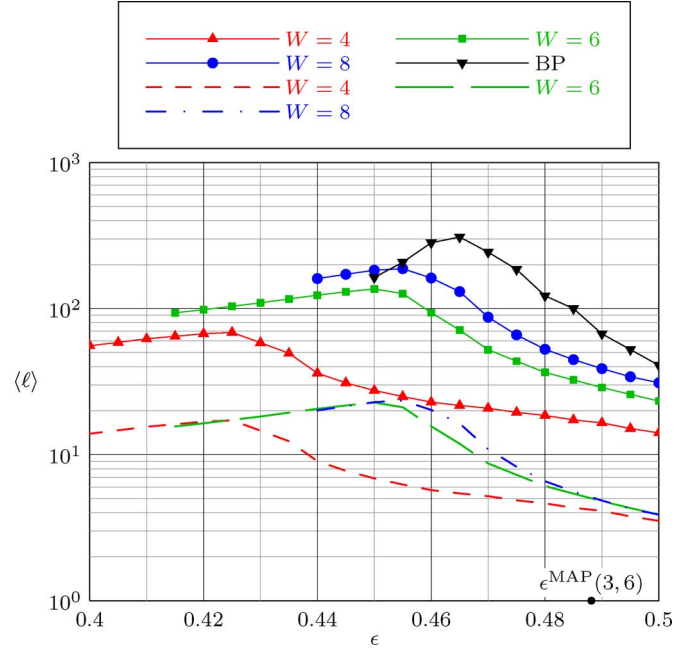

Fig. 3. Average number of iterations $\langle \ell \rangle$ for BP and WD as a function of the channel erasure rate is shown for each window size in solid lines. For the WD, we show in dashed lines, the average number of iterations required within each window configuration.

From Table I, the BP threshold for the ensemble to which this code belongs is $\epsilon^{\text{BP}}(d_l = 3, d_r = 6, \gamma = 3, L = 64) \approx 0.487514$.

Fig. 2 shows the bit erasure rates achieved by using windows of length $W = 4, 6, 8$, i.e., the number of bits within each window was $WM = 4096, 6144$, and 8192, respectively. From the figure, it is clear that good performance can be obtained for a wide range of channel erasure rates even for small window lengths, e.g., $W = 6, 8$. In performing the simulations above, we let the decoders (BP and WD) run for as many iterations as possible, until the decoder could solve for no further bits. For the WD, this meant that within each window configuration, the decoder was allowed to run until it could solve no further bits within the window. Fig. 3 plots the average number of iterations for the BP decoder and the average number of iterations within each window configuration times the chain length (which corresponds to the average number of iterations) for the WD. We can see that for randomly chosen spatially coupled codes, a modest reduction in complexity is possible by using the WD in the waterfall region. Interestingly, the average number of iterations required per window configuration is independent of the chain length below certain channel erasure rates. The number of iterations required decreases beyond a certain value of ϵ because for these higher erasure rates, the decoder is no longer able to decode and gets stuck quickly. Although the smaller window sizes have a large reduction in complexity and a decent BER performance, the block erasure rate performance can be fairly bad, e.g., for the window of size 4, the block erasure rate was 1 in the range of erasure rates considered in Fig. 2. However, the block erasure rate improves drastically with increasing window size—for the window of size 8, the block erasure rate at $\epsilon = 0.44$ was $\approx 6.3 \times 10^{-4}$.

The aforementioned illustration suggests that for good performance with reduced complexity via WD, careful code design is necessary. For a certain variety of spatially coupled codes—protograph-based LDPC convolutional codes—certain design rules for good performance with WD were given in [28], and ensembles with good performance for a wide range of window sizes (including window sizes as small as γ) over erasure channels with and without memory were constructed. For these codes constructed using PEG [31] and ACE [32] techniques, not only can the error floor be lowered but also the performance of a medium-sized WD with fixed number of iterations can be made to be very close to that of the BP decoder [28]. It is for such codes that the windowed decoder is able to attain very good performance with significant reduction in complexity and decoding latency.

VI. CONCLUSION

We considered a WD scheme for decoding spatially coupled codes that has smaller complexity and latency than the BP decoder. We analyzed the asymptotic performance limits of such a scheme by defining WD thresholds for meeting target erasure rates. We gave a lower bound on the WD thresholds and showed that these thresholds are guaranteed to approach the BP threshold for the spatially coupled code at least exponentially in the window size. Through DE, we showed that, in fact, the WD thresholds approach the BP threshold much faster than is guaranteed by the lower bound proved analytically. Since the BP thresholds for spatially coupled codes are themselves close to the MAP threshold, WD gives us an efficient way to trade off complexity and latency for decoding performance approaching the optimal MAP performance. Since the MAP decoder is capacity-achieving as the degrees of variables and checks are increased, similar performance is achievable through a WD scheme for a target erasure floor.

Through simulations, we showed that WD is a viable scheme for decoding finite-length spatially coupled codes and that even for small window sizes, good performance is attainable for a wide range of channel erasure rates. However, the complexity reduction for randomly constructed spatially coupled codes is not as significant as that obtained for protograph-based LDPC convolutional codes with a large girth. Thus, characterizing good spatially coupled codes within the ensemble of randomly coupled codes is a question that remains.

The WD scheme was analyzed here for the BEC and, therefore, the superior performance of these codes and the low complexity and latency of the WD scheme make these attractive for applications in coding over upper layers of the internet protocol. Furthermore, the same scheme can be employed for decoding spatially coupled codes over any channel. However, for channels that introduce errors apart from erasures, the WD scheme can suffer from error propagation. This effect would be similar to what occurs in decoding convolutional codes using a Viterbi decoder with a fixed traceback length. Analysis of the WD scheme and providing performance guarantees over such channels will play a key role in making spatially coupled codes and the WD scheme practical.

APPENDIX I
PROOF OF LEMMA 10

For the lower bound, we have

$$
\begin{aligned}
y_i &= \epsilon g(y_{i-\gamma+1}, \ldots, y_{i+\gamma-1}) \\
&\geq \epsilon g(\underbrace{0, \ldots, 0}_{\gamma}, \underbrace{y_{i+1}, \ldots, y_{i+1}}_{\gamma-1}) \\
&\overset{(a)}{\geq} \epsilon \left(\frac{\gamma-1}{2\gamma} y_{i+1} \right)^{d_l-1}
\end{aligned}
$$

where (a) follows from the fact [19, Lem. 24(iii)] that

$$
g(y_{i-\gamma+1}, \ldots, y_{i+\gamma-1}) \geq \bar{y}_i^{d_l-1}
$$

where $\bar{y}_i = \frac{1}{\gamma^2} \sum_{j,k=0}^{\gamma-1} y_{i+j-k}$. Applying this bound recursively for $y_i, y_{i+1}, \ldots, y_{i+j-1}$, we get

$$
y_i \geq \left(\epsilon(\frac{\gamma-1}{2\gamma})^{d_l-1} \right)^{\frac{(d_l-1)^j-1}{d_l-2}} y_{i+j}^{(d_l-1)^j} \overset{\triangle}{=} \Phi e^{-\phi_j(d_l-1)^j}
$$

where $\Phi = \left(\epsilon(\frac{\gamma-1}{2\gamma})^{d_l-1} \right)^{\frac{-1}{d_l-2}} \geq 1$ and $\phi_j = \ln\left(\frac{\Phi}{y_{i+j}} \right) \geq 0$. When $i + j = W + 1$, since $y_{W+1} = 1$

$$
y_i \geq \Phi e^{-\phi(d_l-1)^{W+1-i}}, \quad i = 1, 2, \ldots, W
$$

with $\phi = \phi_{W+1-i} = \ln \Phi$.

For the upper bound

$$
\begin{aligned}
y_i &\leq \epsilon g(\underbrace{y_{i+k}, \ldots, y_{i+k}}_{\gamma+k}, \underbrace{1, \ldots, 1}_{\gamma-k-1}) \\
&\overset{(b)}{\leq} \epsilon \left(1 - \alpha_k(1 - y_{i+k})^{d_r-1} \right)^{d_l-1} \\
&\overset{\triangle}{=} f_k(\epsilon, y_{i+k})
\end{aligned}
$$

for $k \in [0, \gamma-1]$, where $\alpha_k = (1 - \frac{(\gamma-k-1)(\gamma-k)}{2\gamma^2})^{d_r-1}$. Here, (b) follows from [19, Lem. 24(i)]

$$
g(y_{i-\gamma+1}, \ldots, y_{i+\gamma-1}) \leq (1 - (1 - \bar{y}_i)^{d_r-1})^{d_l-1}.
$$

Note that for $k = \gamma - 1$, $f_{\gamma-1}(\epsilon, x) = f(\epsilon, x)$, the FDE update equation for the (d_l, d_r)-regular ensemble. This proves the Lemma. We now discuss the utility and limitations of the upper bounds derived here.

Fig. 4 plots the bounds $f_k(\epsilon, y_{i+k})$ for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ ensemble for two values of ϵ, one below and the other above the BP threshold $\epsilon^{\mathrm{BP}}(d_l, d_r)$. As is clear from the figure, the tightest bounds are obtained for $k = \gamma - 1$. Note that the bound when $k = 0$ can be recursively computed to obtain a universal upper bound y_{ub} on all the window constellation points y_i for a given (d_l, d_r, γ, L) ensemble, given by the FP of the equation

$$
y = f_0(\epsilon, y) = \epsilon \left(1 - (\frac{\gamma+1}{2\gamma})^{d_r-1}(1-y)^{d_r-1} \right)^{d_l-1}
$$

which is plotted in Fig. 5. As can be seen from the plot, these upper bounds are only marginally tighter than the trivial upper bound of ϵ. In general, we can write $y_W \leq y_{\mathrm{ub}}$ and use the other upper bounds $f_k(\cdot, \cdot)$ to obtain better bounds for other sections
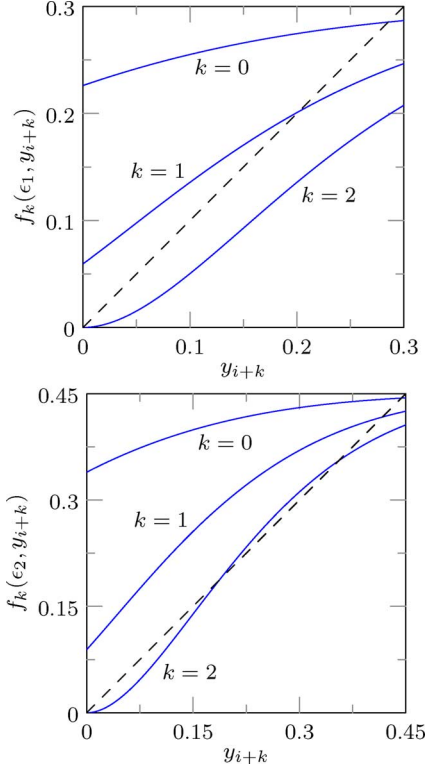
Fig. 6. Upper bound, $\hat{y}_1 \geq y_1$, for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ ensemble with a window of size $W = 9$. The channel erasure rate $\epsilon = 0.3$. Note that $c = 4, d = 0, y_{\mathrm{ub}} \approx 0.285464$, and $\hat{y}_1 = f_{24}(0.3, y_{\mathrm{ub}}) \approx 0.0343947$.

Fig. 4. Upper bounds $f_k(\epsilon, y_{i+k})$ for two values of ϵ: $\epsilon_1 = 0.3 < \epsilon^{\mathrm{BP}}(d_l, d_r) \approx 0.4294 < \epsilon_2 = 0.45$ for the $(d_l = 3, d_r = 6, \gamma = 3, L)$ ensemble.
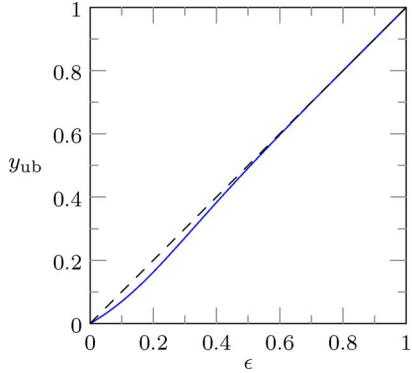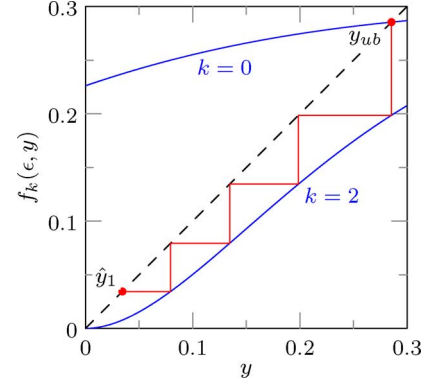
where $c = \left\lfloor \frac{W-1}{\gamma-1} \right\rfloor$, $d = W - 1 - c(\gamma - 1)$. This bounding is particularly useful when $\epsilon \leq \epsilon^{\mathrm{BP}}(d_l, d_r)$ when the FP of the $f_{\gamma-1}(\cdot, \cdot)$ upper bound is zero. It is sometimes possible that $f_{(\gamma-1)^c}(\epsilon, y_{\mathrm{ub}}) \leq f_{d,(\gamma-1)^c}(\epsilon, y_{\mathrm{ub}})$, in which case we can retain the tighter upper bound $f_{(\gamma-1)^c}(\epsilon, y_{\mathrm{ub}})$. Fig. 6 shows an example of the upper bound on y_1 graphically. As a consequence of this upper bound, as $W \to \infty$, we have that $y_1 \to 0$ for $\epsilon \leq \epsilon^{\mathrm{BP}}(d_l, d_r)$. However, for $\epsilon > \epsilon^{\mathrm{BP}}(d_l, d_r)$, these upper bounds are not very useful since the FP of the $f_{\gamma-1}(\cdot, \cdot)$ upper bound is nonzero (cf., Fig. 4).

APPENDIX II
PROOF OF PROPOSITION 14

In the following, we will use some results from [19] summarized below. We define

$$h(y) \stackrel{\triangle}{=} f(\epsilon, y) - y$$

where

$$f(\epsilon, y) = \epsilon (1 - (1 - y)^{d_r - 1})^{d_l - 1}$$

Fig. 5. Universal upper bounds y_{ub} on the constellation points y_i as a function of ϵ for the $(d_l = 3, d_r = 6, \gamma = 2, L)$ ensemble. These bounds are only marginally tighter than the straightforward upper bound ϵ. Also, the bounds are nondecreasing in γ.

the DE update equation for randomized (d_l, d_r)-regular ensembles. For $\epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$, the equation $h(y) = 0$ has exactly three roots in the interval $[0, 1]$, given by $0, y_u(\epsilon)$ and $y_s(\epsilon)$. Between 0 and $y_u(\epsilon)$, $h(y)$ is negative, attaining a unique minimum at $y_{\min}(\epsilon)$. Between $y_u(\epsilon)$ and $y_s(\epsilon)$, $h(y)$ is positive, attaining a unique maximum at $y_{\max}(\epsilon)$. Beyond $y_s(\epsilon)$, $h(y)$ is negative again. Between 0 and $y_{\min}(\epsilon)$, $h(y)$ is upper bounded by a line through the origin with slope

$$-\mu_1(\epsilon) \stackrel{\triangle}{=} \frac{h(y_{\min}(\epsilon))}{y_{\min}(\epsilon)}$$

as follows. In the sequel, we shall write $f_{k_1, k_2, \ldots, k_c}(\epsilon, y)$ to denote $f_{k_1}(\epsilon, f_{k_2}(\epsilon, \ldots f_{k_c}(\epsilon, y)))$ and similarly define

$$f_{k^c}(\epsilon, y) \stackrel{\triangle}{=} \underbrace{f_k(\epsilon, f_k(\epsilon, \ldots, f_k(\epsilon, y)))}_{c}.$$

Thus, for $j = c(\gamma - 1) + d, 0 \leq c, 0 \leq d < \gamma - 1$, we can write $y_i \leq f_{d,(\gamma-1)^c}(\epsilon, y_{i+j})$. The FP value of the erasure probability of a variable node in the first section y_1 can therefore be bounded in terms of the window size W as

$$y_1 \leq f_{d,(\gamma-1)^c}(\epsilon, y_{\mathrm{ub}})$$

i.e., the line $l(y) = -\mu_1(\epsilon)y$. Between $y_{\min}(\epsilon)$ and $y_u(\epsilon)$, $h(y)$ is upper bounded by a line passing through $(y_u(\epsilon), 0)$ with a slope

$$\mu_2(\epsilon) \stackrel{\triangle}{=} \min\left\{ \frac{-h(y_{\min}(\epsilon))}{y_u(\epsilon) - y_{\min}(\epsilon)}, h'(y_u(\epsilon)) \right\}.$$
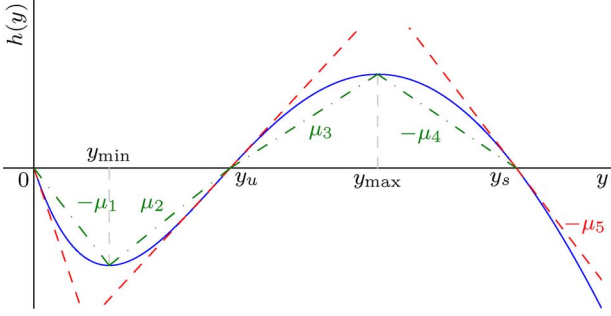
Fig. 7. Plot of $h(y)$ (in solid blue) for the $(d_l = 3, d_r = 6)$ ensemble for $\epsilon = 0.47$ illustrating the properties stated previously. We have dropped the dependence of all the parameters on ϵ from the notation. The tangents at 0, y_u, and y_s are shown as dashed red lines. The other lines used in bounding $h(y)$ are shown as dash-dotted green lines. The μ_i's, $i = 1, \ldots, 5$, are shown in the same color as the lines, whose absolute values of slopes they represent, that bound $h(y)$ in various regions.

Between $y_u(\epsilon)$ and $y_{\max}(\epsilon)$, $h(y)$ is lower bounded by a line through $(y_u(\epsilon), 0)$ with a slope

$$\mu_3(\epsilon) \overset{\triangle}{=} \min\left\{\frac{h(y_{\max}(\epsilon))}{y_{\max}(\epsilon) - y_u(\epsilon)}, h'(y_u(\epsilon))\right\}.$$

Between $y_{\max}(\epsilon)$ and $y_s(\epsilon)$, $h(y)$ is lower bounded by the line through $(y_s(\epsilon), 0)$ with slope

$$-\mu_4(\epsilon) \overset{\triangle}{=} \max\left\{\frac{-h(y_{\max}(\epsilon))}{y_s(\epsilon) - y_{\max}(\epsilon)}, h'(y_s(\epsilon))\right\}.$$

Beyond $y_s(\epsilon)$, $h(y)$ is upper bounded by the line through $(y_s(\epsilon), 0)$ with slope

$$-\mu_5(\epsilon) \overset{\triangle}{=} h'(y_s(\epsilon)).$$

Each of the $\mu_i(\epsilon)$'s, $i = 1, \ldots, 5$, defined previously is strictly positive for ϵ in the specified range. For a general ϵ, we will drop the dependence of each of these parameters on ϵ from the notation. When $\epsilon = \epsilon^* \overset{\triangle}{=} \epsilon^{\mathrm{MAP}}(d_l, d_r)$, the corresponding parameters are themselves shown with ∗'s. These properties of $h(y)$ are illustrated in Fig. 7. We can lower bound $y_{\min}$ as $y_{\min} \geq \frac{1}{d_l^2 d_r^2}$, and the slope μ_1 as $\mu_1 \geq \frac{1}{8d_r^2} \overset{\triangle}{=} \tilde{\mu}_1 \ \forall \ \epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$.

Further, $|h'(y)| \leq d_l d_r \ \forall \ y \in [0, 1]$. We have $h'(0) = h'(1) = -1$, $h'(y_{\min}) = h'(y_{\max}) = 0$. $h''(0) = h''(\hat{y}) = h''(1) = 0$, where

$$\hat{y} = 1 - \left(\frac{d_r - 2}{d_l d_r - d_l - d_r}\right)^{\frac{1}{d_r - 1}}$$

and

$$h''(y) \begin{cases} > 0, & y \in (0, \hat{y}) \\ < 0, & y \in (\hat{y}, 1). \end{cases}$$

From Rolle's Theorem, $y_{\min} \leq \hat{y} \leq y_{\max}$. We first give some simple bounds for the μ_i's defined earlier which will be useful in the proof.

Lemma 17 (μ_4, μ_5 Bounds): For $\epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$, we have $0 < \mu_4 \leq \mu_5 < 1$.

Proof: Since $h''(y) < 0$ for $y \in (y_{\max}, 1)$, $h'(y)$ monotonically decreases in this interval. Thus, $0 < \mu_5 = -h'(y_s) < 1$. From the mean value theorem, we have $h'(\xi) = -\frac{h(y_{\max})}{y_s - y_{\max}} \geq$
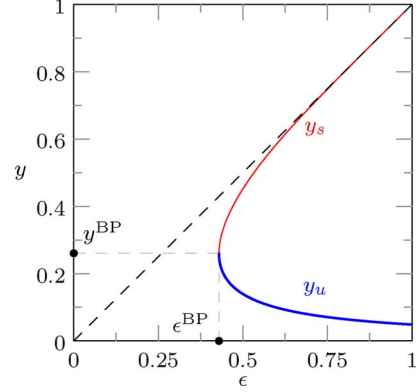
$h'(y_s)$ for some $\xi \in [y_{\max}, y_s]$ so that $0 < \mu_4 = \frac{h(y_{\max})}{y_s - y_{\max}} \leq -h'(y_s) = \mu_5 < 1$. ∎

The values y_u and y_s are referred to as the *unstable* and *stable* FPs of DE for the (d_l, d_r)-regular ensemble, respectively. This is because both these values satisfy $h(y) = 0$ or $y = f(\epsilon, y) = \epsilon(1 - (1 - y)^{d_r - 1})^{d_l - 1}$. The ϵ for which the FP is $y \in [0, 1]$ is given by

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

The BP threshold is hence the smallest value of $\epsilon(y)$, i.e., $\epsilon^{\mathrm{BP}}(d_l, d_r) = \min\{\epsilon(y), y \in [0, 1]\}$. The value of y that achieves this minimum is denoted y^{BP}. Then, $\forall \ \epsilon \in [\epsilon^{\mathrm{BP}}(d_l, d_r), 1]$, the unstable and stable FPs are given by

$$y_u(\epsilon) = y \in [0, y^{\mathrm{BP}}] : \epsilon(y) = \epsilon$$
$$y_s(\epsilon) = y \in [y^{\mathrm{BP}}, 1] : \epsilon(y) = \epsilon. \tag{10}$$

Fig. 8 plots these stable and unstable FPs. The reason why y_s is called the stable FP (and y_u the unstable FP) can be explained through Fig. 8. For $\epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), 1]$, when the FDE updates are performed, the value y monotonically decreases from 1 and converges to the first solution of the equation $h(y) = 0$, which happens to be $y_s(\epsilon)$ for ϵ in this range. Therefore, performing BP always results in the FP y_s and hence the adjective "stable." Similarly for y_u, which is a solution never reached through BP, it can be shown that a small perturbation from the value of y_u will result in convergence to either y_s or 0. Therefore, y_u's are "unstable" FPs.

We can define the derivatives y'_s and y'_u of y_s and y_u, respectively, with respect to ϵ for $\epsilon \in (\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$. It is easy to see that y'_s is monotonically decreasing and y'_u is monotonically increasing in ϵ. For details and proofs of the aforementioned properties, see [19, Appendix II] and [33].

We are now ready to prove the proposition. Note that when W is smaller than the claimed upper bound on the transition width, the claim is trivially true, i.e., the transition width cannot be longer than the window size. However, in this case, we cannot guarantee $y_1 \leq \delta$. Hence, we will assume that W is larger than the bound. In the following, we will often use the bound

$$f(\epsilon, y_{i-\gamma+1}) \leq y_i \leq f(\epsilon, y_{i+\gamma-1}).$$

Fig. 8. Unstable and stable FPs of DE for the $(d_l = 3, d_r = 6)$-regular ensemble as given by (10). $y_u(\epsilon)$ is shown as the thick blue curve and $y_s(\epsilon)$ as the thin red curve. By definition, $y_s(\epsilon) \geq y^{\mathrm{BP}}$ and $y_u(\epsilon) \leq y^{\mathrm{BP}}$. y^{BP} and $\epsilon^{\mathrm{BP}} \equiv \epsilon^{\mathrm{BP}}(d_l = 3, d_r = 6)$ are also shown. Note that $y_s(\epsilon^{\mathrm{BP}}) = y^{\mathrm{BP}} = y_u(\epsilon^{\mathrm{BP}})$.

We now define a schedule that results in a FP window constellation that dominates the FP of the parallel schedule, $\underline{y}$, for a channel erasure rate $\epsilon \in \mathscr{E}$. We then upper bound the actual transition width by the transition width of the dominating FP. We generate the dominating FP in steps.

1) Set $\underline{y}^{(0)} = \underline{1}$ and evaluate the sequence of window constellations $\{\underline{y}^{(\ell)}\}$ according to (6), but with the boundary conditions

$$y_i^{(\ell)} = \begin{cases} y_1^{(\ell)}, & i \leq 0 \\ 1, & i > W. \end{cases}$$

We have the FP in this case, $\underline{y}^A$, satisfying $\underline{y}^A \succeq \underline{y}$ by induction. Further

$$y_1^A = \epsilon g(\underbrace{y_1^A, \ldots, y_1^A}_{\gamma}, y_2^A, \ldots, y_\gamma^A) \tag{11}$$

$$\geq \epsilon g(\underbrace{y_1^A, \ldots, y_1^A}_{2\gamma-1}) = f(\epsilon, y_1^A)$$

so that $y_1^A \geq y_s$. Note that $y_1^A \leq y_u$ cannot happen since, starting from 1, y_1^A will equal the first solution of (11), which from the continuity of the DE equations is guaranteed to have a solution no smaller than y_s. Starting from the right end, we now count the number of sections until $y_i^A \leq y_s + \Upsilon_A$ where we choose $\Upsilon_A = \frac{y_s^* - y_s}{2}$. Recall that the $*$-ed values correspond to $\epsilon^* = \epsilon^{\mathrm{MAP}}(d_l, d_r)$. We first observe that

$$y_{W-\gamma+1}^A \leq \epsilon g(y_W^A, \ldots, y_W^A) = f(\epsilon, y_W^A).$$

Hence

$$y_W^A - y_{W-\gamma+1}^A \geq y_W^A - f(\epsilon, y_W^A)$$
$$= -h(y_W^A) \geq \mu_5(y_W^A - y_s)$$

which implies that $y_{W-\gamma+1}^A - y_s \leq (1 - \mu_5)(y_W^A - y_s)$. From similar reasoning, we can show that

$$y_{W-m(\gamma-1)}^A - y_s \leq (1 - \mu_5)(y_{W-(m-1)(\gamma-1)}^A - y_s)$$
$$\leq (1 - \mu_5)^m(y_W^A - y_s)$$
$$\leq (1 - \mu_5)^m(y_{\mathrm{ub}} - y_s).$$

Since $\mu_5 < 1$ from Lemma 17, the aforementioned difference is decreasing in m. From the definition of y_{ub} (note that this upper bound is valid even for the boundary conditions specified here) in the proof of Lemma 10, it is easy to see that $y_{\mathrm{ub}} \geq y_s$ so that the right-hand side of the aforementioned chain of inequalities is nonnegative. Thus, $y_{W-m(\gamma-1)}^A \leq y_s + \Upsilon_A$ if

$$m \geq \left\lceil \frac{\ln \frac{y_{\mathrm{ub}} - y_s}{\Upsilon_A}}{\ln \frac{1}{1-\mu_5}} \right\rceil.$$

Let $\tilde{\mu}_5 = \min\{\mu_5, \epsilon \in \mathscr{E}\}$. Then, we can write from the mean value theorem

$$\Upsilon_A = \frac{y_s^* - y_s}{2} = \frac{\epsilon^* - \epsilon}{2} y_s'(\hat{\epsilon})$$

for some $\hat{\epsilon} \in [\epsilon, \epsilon^*]$. We can lower bound this as

$$\Upsilon_A \geq \frac{\epsilon^* - \epsilon}{2} y_s'(\epsilon^*) \geq \frac{\Delta\epsilon}{2} y_s'(\epsilon^*)$$

where the first inequality follows from the fact that y_s' is decreasing in ϵ in the interval $(\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$ and the second from $\epsilon^* \geq \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)$. Therefore, this width is no more than

$$(\gamma - 1)\left(\frac{\ln \frac{2(y_{\mathrm{ub}} - y^{\mathrm{BP}})}{y_s'(\epsilon^*)\Delta\epsilon}}{\ln \frac{1}{1-\tilde{\mu}_5}} + 1\right)$$

sections, since $y^{\mathrm{BP}} \leq y_s$.

2) From the definition of Υ_A, we have $y_s + \Upsilon_A = y_s^* - \Upsilon_A$. Let i_A be the largest index for which $y_{i_A}^A \leq y_s^* - \Upsilon_A$. Set $\underline{y}^{(0)} = \underline{y}^A$ and evaluate the sequence of window constellations $\{\underline{y}^{(\ell)}\}$ according to (6) performing the updates only for those sections with indices $i < i_A$. Further, perform the updates for the channel erasure rate ϵ^* since we only require an upper bound on the transition width. We set the left end of the window to perform these updates to 0, i.e., $y_i = 0 \ \forall i \leq 0$. Let $\underline{y}^B$ denote the FP window constellation at the end of this procedure. By induction, we have $y_i^B \leq y_s^* - \Upsilon_A \ \forall i \leq i_A$. Also, $y_{i_A-1}^B \geq f(\epsilon^*, y_{i_A-\gamma}^B)$ so that

$$y_{i_A-1}^B - y_{i_A-\gamma}^B \geq h(y_{i_A-\gamma}^B) \geq \mu_4^*(y_s^* - y_{i_A-\gamma}^B)$$

where the last inequality assumes that $y_{i_A-\gamma}^B \geq y_{\mathrm{max}}^*$. Note that there is no loss of generality in this assumption, for if it were not true, we have that the number of sections with FP values y_i^B between y_{max}^* and y_s^* is smaller than the upper bound we derive in the following. The aforementioned inequality implies that

$$y_s^* - y_{i_A-\gamma}^B \geq \frac{y_s^* - y_{i_A-1}^B}{1 - \mu_4^*}.$$

Similarly, it can be shown that as long as $y_{i_A-1-m(\gamma-1)}^B \geq y_{\mathrm{max}}^*$

$$y_s^* - y_{i_A-1-m(\gamma-1)}^B \geq \frac{y_s^* - y_{i_A-1-(m-1)(\gamma-1)}^B}{1 - \mu_4^*}$$

and by induction

$$y_s^* - y_{i_A-1-m(\gamma-1)}^B \geq \frac{y_s^* - y_{i_A-1}^B}{(1 - \mu_4^*)^m} \geq \frac{\Upsilon_A}{(1 - \mu_4^*)^m}.$$

Note that since $\mu_4^* < 1$ from Lemma 17, the aforementioned difference is increasing in m. Thus, there are no more than

$$(\gamma - 1)\left(\frac{\ln \frac{2(y_s^* - y_{\mathrm{max}}^*)}{y_s'(\epsilon^*)\Delta\epsilon}}{\ln \frac{1}{1-\mu_4^*}} + 1\right)$$

sections with $y_i^B \in [y_{\mathrm{max}}^*, y_s^*]$.

3) Let i_{max} be the largest index i such that $y_i^B \leq y_{\mathrm{max}}^*$. We define $\Upsilon_B = \frac{y_u - y_u^*}{2}$ and count the number of sections with

FP values y_i^B between $y_u^* + \Upsilon_B$ and $y_{\max}^*$. Since $y_{i_{\max}}^B \geq f(\epsilon^*, y_{i_{\max}-(\gamma-1)}^B)$, we have

$$y_{i_{\max}}^B - y_{i_{\max}-(\gamma-1)}^B \geq h(y_{i_{\max}-(\gamma-1)}^B)$$
$$\geq \mu_3^*(y_{i_{\max}-(\gamma-1)}^B - y_u^*),$$

where we again assume without loss of generality that $y_{i_{\max}-(\gamma-1)}^B \geq y_u^*$. The aforementioned inequality implies that

$$y_{i_{\max}-(\gamma-1)}^B - y_u^* \leq \frac{y_{i_{\max}}^B - y_u^*}{1+\mu_3^*}.$$

Again by induction

$$y_{i_{\max}-m(\gamma-1)}^B - y_u^* \leq \frac{y_{i_{\max}}^B - y_u^*}{(1+\mu_3^*)^m}$$

as long as $y_{i_{\max}-m(\gamma-1)}^B \geq y_u^*$. Since $\mu_3^* > 0$, the aforementioned difference is decreasing in m, and consequently, $y_{i_{\max}-m(\gamma-1)}^B \leq y_u^* + \Upsilon_B$ if

$$m \geq \left\lceil \frac{\ln \frac{y_{\max}^* - y_u^*}{\Upsilon_B}}{\ln(1+\mu_3^*)} \right\rceil.$$

Writing

$$\Upsilon_B = \frac{y_u - y_u^*}{2} = -\frac{\epsilon^* - \epsilon}{2} y_u'(\breve{\epsilon})$$

from the mean value theorem for some $\breve{\epsilon} \in [\epsilon, \epsilon^*]$, we can bound this as

$$\Upsilon_B \geq -\frac{\epsilon^* - \epsilon}{2} y_u'(\epsilon^*) \geq -\frac{\Delta\epsilon}{2} y_u'(\epsilon^*)$$

where the first inequality follows because $-y_u'$ is decreasing in ϵ in the interval $(\epsilon^{\mathrm{BP}}(d_l, d_r), 1)$ and the second because $\epsilon^* \geq \epsilon^{\mathrm{BP}}(d_l, d_r, \gamma)$. This implies that there are no more than

$$(\gamma-1)\left(\frac{\ln \frac{2(y_{\max}^* - y_u^*)}{-y_u'(\epsilon^*)\Delta\epsilon}}{\ln(1+\mu_3^*)} + 1\right)$$

sections with FP values between $y_u^* + \Upsilon_B$ and $y_{\max}^*$.

4) From the definition of Υ_B, we have $y_u^* + \Upsilon_B = y_u - \Upsilon_B$. Let i_B be the largest index i such that $y_i^B \leq y_u - \Upsilon_B$. Set $\underline{y}^{(0)} = \underline{y}^B$ and evaluate $\{\underline{y}^{(\ell)}\}$ according to (6) performing the updates only for sections with indices $i \leq i_B$ with channel erasure rate ϵ. Again, we set the left end of the window to 0 while performing the updates. Denote the FP obtained at the end of this procedure as $\underline{y}^C$. Clearly, $y_i^C \leq y_u - \Upsilon_B \ \forall \ i \leq i_B$. Since $y_{i_B-\gamma+1}^C \leq f(\epsilon, y_{i_B}^C)$, we have

$$y_{i_B}^C - y_{i_B-\gamma+1}^C \geq -h(y_{i_B}^C) \geq \mu_2(y_u - y_{i_B}^C)$$

so that

$$y_u - y_{i_B-\gamma+1}^C \geq (1+\mu_2)(y_u - y_{i_B}^C).$$

Here, we assume that $y_{i_B}^C \geq y_{\min}$ in order to obtain an upper bound on the number of sections in the range

$[y_{\min}, y_u]$. From similar reasoning as earlier, as long as $y_{i_B-m(\gamma-1)}^C \geq y_{\min}$

$$y_u - y_{i_B-m(\gamma-1)}^C \geq (1+\mu_2)(y_u - y_{i_B-(m-1)(\gamma-1)}^C)$$

and by induction

$$y_u - y_{i_B-m(\gamma-1)}^C \geq (1+\mu_2)^m (y_u - y_{i_B}^C)$$
$$\geq (1+\mu_2)^m \Upsilon_B.$$

Since $\mu_2 > 0$, the aforementioned difference is increasing in m. By letting $\tilde{\mu}_2 = \min\{\mu_2, \epsilon \in \mathscr{E}\}$ and noting that

$$y_u - y_{\min} \leq y^{\mathrm{BP}} - y_{\min} \leq y^{\mathrm{BP}} - \frac{1}{d_l^2 d_r^2}$$

we have that there are no more than

$$(\gamma-1)\left(\frac{\ln \frac{2(y^{\mathrm{BP}} - \frac{1}{d_l^2 d_r^2})}{-y_u'(\epsilon^*)\Delta\epsilon}}{\ln(1+\tilde{\mu}_2)} + 1\right)$$

sections with FP values in the interval $[y_{\min}, y_u]$.

5) Let i_C be the largest index i such that $y_i^C \leq y_{\min}$. Proceeding as earlier, we have

$$y_{i_C}^C - y_{i_C-\gamma+1}^C \geq -h(y_{i_C}^C) \geq \mu_1 y_{i_C}^C$$

and by induction

$$y_{i_C-m(\gamma-1)}^C \leq (1-\mu_1)^m y_{i_C}^C.$$

Thus, between δ_0 and $y_{\min}$, there are no more than

$$(\gamma-1)\left(\frac{\ln \frac{y^{\mathrm{BP}}}{\delta_0}}{\ln \frac{1}{1-\tilde{\mu}_1}} + 1\right)$$

sections with FP values in the interval $[\delta_0, y_{\min}]$, since $y_{\min} \leq y^{\mathrm{BP}}$ and $\mu_1 \geq \tilde{\mu}_1$.

6) Let i_D be the largest index i such that $y_i^C \leq \delta_0$. From Lemma 11, we know that the tail decays double exponentially for $i \leq i_D$. From (8), we have $y_{i_D-m(\gamma-1)}^C \leq \Psi e^{-\psi(d_l-1)^m} \leq \tilde{\Psi} e^{-\tilde{\psi}(d_l-1)^m}$, where

$$\Psi = \delta_0 \epsilon^{-\frac{1}{d_l-2}}$$
$$\leq \delta_0 \left(\frac{\epsilon^{\mathrm{BP}}(d_l, d_r, \gamma) + \epsilon^{\mathrm{BP}}(d_l, d_r)}{2}\right)^{-\frac{1}{d_l-2}} \triangleq \tilde{\Psi}$$

and

$$\psi = \frac{1}{d_l-2}\ln\frac{1}{\epsilon} \geq \frac{1}{d_l-2}\ln\frac{1}{\epsilon^{\mathrm{MAP}}(d_l, d_r)} \triangleq \tilde{\psi}.$$

Thus, there are no more than

$$(\gamma-1)\left(\frac{1}{\ln(d_l-1)}\ln\ln\frac{\tilde{\Psi}}{\delta} + \frac{\ln\frac{1}{\tilde{\psi}}}{\ln(d_l-1)} + 1\right)$$

sections with $y_i^C \in [\delta, \delta_0]$.

Finally, collecting all these terms, we conclude that the transition width of the FP obtained from the procedure highlighted in the steps 1) through 6) is upper bounded by

$$\hat{\tau}(\epsilon, \delta) = (\gamma-1)\left(\mathsf{A}\ln\ln\frac{\mathsf{D}}{\delta} + \mathsf{B}\ln\frac{1}{\Delta\epsilon} + \hat{\mathsf{C}}\right)$$

where the constants A, B, and D are as follows:

$$A = \frac{1}{\ln(d_l - 1)}$$

$$B = \frac{1}{\ln(1 + \tilde{\mu}_2)} + \frac{1}{\ln(1 + \mu_3^*)} + \frac{1}{\ln(\frac{1}{1-\mu_4^*})} + \frac{1}{\ln(\frac{1}{1-\tilde{\mu}_5})}$$

$$D = \tilde{\Psi}.$$

The constant $\hat{C}$ is as given in (12) shown at the bottom of the page. Note that these constants depend only on the ensemble parameters d_l, d_r, and γ. Since it is clear that the FP obtained through the procedure in steps 1) through 6) earlier dominates pointwise the first window configuration FP of FDE with a window of size W for channel erasure rate ϵ, we can guarantee that the transition width is upper bounded by the aforementioned expression. This completes the proof.

APPENDIX III
PROOF OF PROPOSITION 16

We start with the first window configuration FP of FDE when the channel erasure rate is ϵ and show that this FP dominates the cth window configuration FP of FDE for every c for a smaller channel erasure rate $\upsilon \leq \epsilon$. To prove this, it suffices to show that the FP $\underline{\omega}$ defined in Lemma 15 for channel erasure rate υ is dominated pointwise by the first window configuration FP for channel erasure rate ϵ. This establishes υ as being a lower bound on the WD threshold ϵ^{WD}.

Set $\underline{y}^{(0)} = \underline{y}_{\{1\}}^{(\infty)}$, the first window configuration FP of FDE for channel erasure rate ϵ. Evaluate $\{\underline{y}^{(\ell)}\}_{\ell=1}^{\infty}$ according to

$$\underline{y}^{(\ell)} = \Omega(\underline{y}^{(\ell-1)})$$

where $\Omega(\cdot)$ is as defined in Lemma 15, but for channel erasure rate $\upsilon = \epsilon - \Delta\epsilon$. Then, the following are true:

$$y_i^{(0)} = \begin{cases} \epsilon g(\underbrace{0,\ldots,0}_{\gamma-i}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}), & 1 \leq i < \gamma \\ \epsilon g(y_{i-\gamma+1}^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}), & \gamma \leq i \leq W \end{cases}$$

and

$$y_i^{(1)} = \begin{cases} \upsilon g(\underbrace{y_1^{(0)}, \ldots, y_1^{(0)}}_{\gamma-i}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}), & 1 \leq i < \gamma \\ \upsilon g(y_{i-\gamma+1}^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}), & \gamma \leq i \leq W. \end{cases}$$

For $\gamma \leq i \leq W$

$$y_i^{(1)} = \frac{\upsilon}{\epsilon} y_i^{(0)} = \frac{\epsilon - \Delta\epsilon}{\epsilon} y_i^{(0)} \leq y_i^{(0)}. \tag{13}$$

Let us write

$$g_i(\sigma, y_1, \ldots, y_{i+\gamma-1}) \overset{\triangle}{=} g(\underbrace{\sigma, \ldots, \sigma}_{\gamma-i}, y_1, \ldots, y_{i+\gamma-1})$$

and

$$G_i(\epsilon, \sigma, y_1, \ldots, y_{i+\gamma-1}) \overset{\triangle}{=} \epsilon g_i(\sigma, y_1, \ldots, y_{i+\gamma-1})$$

for $1 \leq i < \gamma$. For i in this range, consider

$$\begin{aligned} y_i^{(0)} - y_i^{(1)} &= G_i(\epsilon, 0, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \\ &\quad - G_i(\upsilon, y_1^{(0)}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \\ &= \Big[G_i(\epsilon, 0, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \\ &\quad - G_i(\upsilon, 0, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \Big] \\ &\quad - \Big[G_i(\upsilon, y_1^{(0)}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \\ &\quad - G_i(\upsilon, 0, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \Big] \\ &\overset{(a)}{=} \Delta\epsilon \frac{\partial G_i}{\partial\epsilon}\Big|_{\xi,\sigma=0,\underline{y}^{(0)}} - y_1^{(0)} \frac{\partial G_i}{\partial\sigma}\Big|_{\upsilon,\zeta,\underline{y}^{(0)}} \end{aligned} \tag{14}$$

where $\xi \in [\upsilon, \epsilon]$ and $\zeta \in [0, y_1^{(0)}]$. Here, (a) follows from the mean value theorem. We have

$$\frac{\partial G_i}{\partial\epsilon}\Big|_{\xi,\sigma=0,\underline{y}^{(0)}} = g_i(\sigma, y_1, \ldots, y_{i+\gamma-1})\Big|_{\xi,\sigma=0,\underline{y}^{(0)}} = \frac{y_i^{(0)}}{\epsilon}.$$

Since $\frac{\partial G_i}{\partial\sigma} = \epsilon\frac{\partial g_i}{\partial\sigma}$, we focus on g_i. When expanded out, the expression for g_i can be written as

$$\begin{aligned} g_i = \Big[1 - \frac{1}{\gamma}\Big((\alpha_{i,1} - \frac{\gamma-i}{\gamma}\sigma)^{d_r-1} + \cdots \\ + (\alpha_{i,\gamma-i} - \frac{1}{\gamma}\sigma)^{d_r-1} + \alpha_{i,\gamma-i+1}^{d_r-1} + \cdots + \alpha_{i,\gamma}^{d_r-1} \Big) \Big]^{d_l-1} \end{aligned}$$

where

$$\alpha_{i,j+1} = \begin{cases} 1 - \frac{1}{\gamma}\sum_{c=1}^{i+j} y_c, & 0 \leq j \leq \gamma-i-1 \\ 1 - \frac{1}{\gamma}\sum_{c=i+j-\gamma+1}^{i+j} y_c, & \gamma-i \leq j \leq \gamma-1. \end{cases}$$

Clearly, $0 \leq \alpha_{i,j+1} \leq 1 \ \forall \ 1 \leq i < \gamma, 0 \leq j \leq \gamma-1$. Therefore

$$\begin{aligned} \frac{\partial g_i}{\partial\sigma} &= \frac{(d_l-1)(d_r-1)}{\gamma} g_i^{\frac{d_l-2}{d_l-1}} \Big[(\alpha_{i,1} - \frac{\gamma-i}{\gamma}\sigma)^{d_r-2}\frac{\gamma-i}{\gamma} \\ &\quad + \cdots + (\alpha_{i,\gamma-i} - \frac{1}{\gamma}\sigma)^{d_r-2}\frac{1}{\gamma} \Big] \\ &\overset{(a)}{\leq} \frac{d_l d_r}{\gamma^2} g_i^{\frac{d_l-2}{d_l-1}} \frac{(\gamma-i)(\gamma-i+1)}{2} \leq \frac{d_l d_r}{2} g_i^{\frac{d_l-2}{d_l-1}}. \end{aligned}$$

$$\hat{C} = \frac{\ln\frac{1}{\psi}}{\ln(d_l-1)} + \frac{\ln\frac{y^{\mathrm{BP}}}{\delta_0}}{\ln\frac{1}{1-\tilde{\mu}_1}} + \frac{\ln\frac{2(y^{\mathrm{BP}} - \frac{1}{d_l^2 d_r^2})}{-y_u'(\epsilon^*)}}{\ln(1+\tilde{\mu}_2)} + \frac{\ln\frac{2(y_{\max}^* - y_u^*)}{-y_u'(\epsilon^*)}}{\ln(1+\mu_3^*)} + \frac{\ln\frac{2(y_s^* - y_{\max}^*)}{y_s'(\epsilon^*)}}{\ln\frac{1}{1-\mu_4^*}} + \frac{\ln\frac{2(y_{\mathrm{ub}} - y^{\mathrm{BP}})}{y_s'(\epsilon^*)}}{\ln\frac{1}{1-\tilde{\mu}_5}} + 6 \tag{12}$$

Here, (a) holds because $0 \leq (\alpha_{i,j+1} - \frac{\gamma-i-j}{\gamma}\sigma) \leq 1 \ \forall \ 0 \leq j \leq \gamma - i - 1$. This implies that

$$
\begin{aligned}
\left.\frac{\partial G_i}{\partial \sigma}\right|_{v,\zeta,\underline{y}^{(0)}} &\leq \frac{d_l d_r}{2} g_i^{\frac{d_l-2}{d_l-1}} \epsilon \Big|_{v,\zeta,\underline{y}^{(0)}} \\
&= \frac{d_l d_r}{2} v^{\frac{1}{d_l-1}} \\
&\quad \times \left(vg(\underbrace{\zeta, \ldots, \zeta}_{\gamma-i}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \right)^{\frac{d_l-2}{d_l-1}} \\
&\overset{(b)}{\leq} \frac{d_l d_r}{2} v^{\frac{1}{d_l-1}} \\
&\quad \times \left(vg(\underbrace{y_1^{(0)}, \ldots, y_1^{(0)}}_{\gamma-i}, y_1^{(0)}, \ldots, y_{i+\gamma-1}^{(0)}) \right)^{\frac{d_l-2}{d_l-1}} \\
&= \frac{d_l d_r}{2} v^{\frac{1}{d_l-1}} \left(y_i^{(1)} \right)^{\frac{d_l-2}{d_l-1}} \\
&\overset{(c)}{\leq} \frac{d_l d_r}{2} v^{\frac{1}{d_l-1}} \left(y_\gamma^{(1)} \right)^{\frac{d_l-2}{d_l-1}} \leq \frac{d_l d_r}{2} \left(y_\gamma^{(1)} \right)^{\frac{d_l-2}{d_l-1}} \\
&\overset{(13)}{\leq} \frac{d_l d_r}{2} y_\gamma^{(0) \frac{d_l-2}{d_l-1}}.
\end{aligned}
$$

Here, the inequality labeled (b) is true because $\zeta \leq y_1^{(0)}$, (c) follows from the observation that $y_i^{(1)} \leq y_{i+1}^{(1)}$, $i \geq 1$, which is in turn true since $\underline{y}^{(0)}$ and $\pi(\underline{y}^{(0)})$ were nondecreasing. Substituting back in (14), we have for $1 \leq i < \gamma$

$$
y_i^{(0)} - y_i^{(1)} \geq \Delta\epsilon \frac{y_1^{(0)}}{\epsilon} - y_1^{(0)} \frac{d_l d_r}{2} y_\gamma^{(0) \frac{d_l-2}{d_l-1}}.
$$

Thus, if $\frac{\Delta\epsilon}{\epsilon} \geq \frac{d_l d_r}{2} y_\gamma^{(0) \frac{d_l-2}{d_l-1}}$, $y_i^{(0)} \geq y_i^{(1)} \ \forall \ i \geq 1$, and hence $\pi(\underline{y}^{(0)})$ dominates $\underline{y}^{(1)}$ pointwise. Recall that

$$
\pi(y_i) = \begin{cases} y_1, & i < 1 \\ y_i, & i \geq 1. \end{cases}
$$

It therefore follows by induction that the limiting constellation $\underline{y}^{(\infty)}$ exists, and is also dominated by $\pi(\underline{y}^{(0)})$. It is clear that $\underline{y}^{(\infty)}$ satisfies

$$
\underline{y}^{(\infty)} = \Omega(\underline{y}^{(\infty)}).
$$

From Lemma 15, $\underline{y}^{(\infty)} \succeq \underline{\omega}$ and hence $y_1^{(\infty)} \geq \hat{x}$.

If the window size is chosen to be $W \geq \hat{W}_{\min}(\delta) + \gamma - 1 \overset{\triangle}{=} W_{\min}(\delta)$, then for the first window, we can guarantee $y_\gamma^{(0)} \leq \delta$ for some $\delta < \delta_0$ for all channel erasure rates smaller than $\epsilon^{\mathrm{FW}} \equiv \epsilon^{\mathrm{FW}}(d_l, d_r, \gamma, W - \gamma + 1, \delta)$. From the aforementioned argument, it follows that we can ensure $\hat{x} \leq \delta$ for all erasure rates smaller than $\epsilon^{\mathrm{FW}}\left(1 - \frac{d_l d_r}{2} \delta^{\frac{d_l-2}{d_l-1}}\right)$. As long as

$$
\delta < \delta_* \overset{\triangle}{=} \left(\frac{2}{d_l d_r}\right)^{\frac{d_l-1}{d_l-2}} < \left(\frac{1}{d_r-1}\right)^{\frac{d_l-1}{d_l-2}} = \delta_0
$$

this erasure rate is a nontrivial lower bound on the WD threshold ϵ^{WD}.

ACKNOWLEDGMENT

REFERENCES

[1] A. R. Iyengar, P. H. Siegel, R. L. Urbanke, and J. K. Wolf, "Windowed decoding of spatially coupled codes," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.–Aug. 31–5, 2011, pp. 2552–2556.

[2] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 23–26, 1993, vol. 2, pp. 1064–1070.

[4] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[5] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.

[7] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[8] A. Amraoui, "Asymptotic and finite-length optimization of LDPC codes," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland, 2006.

[9] A. Amraoui and R. Urbanke, LDPCOpt 2012 [Online]. Available: http://ipgdemos.epfl.ch/ldpcopt/

[10] A. J. Felstrom and K. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2181–2191, Sep. 1999.

[11] K. Engdahl and K. S. Zigangirov, "On the theory of low-density convolutional codes I," *Probl. Peredachi Inf.*, vol. 35, pp. 12–27, 1999.

[12] K. Engdahl, M. Lentmaier, and K. Zigangirov, "On the theory of low-density convolutional codes," in *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes* Transl.:Lecture Notes in Computer Science. Berlin, Germany: Springer-Verlag, 1999, vol. 1719, , pp. 77–86.

[13] M. Lentmaier, D. V. Truhachev, and K. S. Zigangirov, "To the theory of low-density convolutional codes. II," *Probl. Inf. Transmiss.*, vol. 37, pp. 288–306, 2001.

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

[15] A. Sridharan, M. Lentmaier, D. J. Costello, and K. S. Zigangirov, "Convergence analysis of a class of LDPC convolutional codes for the erasure channel," in *Proc. 42nd Annu. Allerton Conf. Commun., Control Comput.*, Monticello, IL, Sep.–Oct. 29–1, 2004, pp. 953–962.

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

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

[18] J. Thorpe, Low-density parity-check (LDPC) codes constructed from protographs California Inst. Technol., Pasadena42-154, 2003.

[19] S. Kudekar, T. Richardson, and R. L. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.

[20] M. Lentmaier and G. Fettweis, "On the thresholds of generalized LDPC convolutional codes based on protographs," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, Jun. 13–18, 2010, pp. 709–713.

[21] S. Kudekar, C. Measson, T. J. Richardson, and R. L. Urbanke, "Threshold saturation on BMS channels via spatial coupling," *CoRR*, vol. abs/1004.3742, 2010.

[22] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *CoRR*, vol. abs/1201.2999, 2012.

[23] H. Hasani, N. Macris, and R. Urbanke, "Coupled graphical models and their thresholds," in *Proc. IEEE Inf. Theory Workshop*, Dublin, Ireland, Aug.-Sep. 30–3, 2010, pp. 1–5.

[24] V. Aref and R. Urbanke, "Universal rateless codes from coupled LT codes," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 277–281.

[25] H. Uchikawa, K. Kasai, and K. Sakaniwa, "Terminated LDPC convolutional codes over $GF(2^p)$," *CoRR*, vol. abs/1010.0060, 2010.

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

[27] A. R. Iyengar, M. Papaleo, G. Liva, P. H. Siegel, J. K. Wolf, and G. E. Corazza, "Protograph-based LDPC convolutional codes for correlated erasure channels," in *Proc. IEEE Int. Conf. Commun.*, Cape Town, South Africa, May 2010, pp. 1–6.

[28] A. R. Iyengar, M. Papaleo, P. H. Siegel, J. K. Wolf, A. Vanelli-Coralli, and G. E. Corazza, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2303–2320, Apr. 2012.

[29] P. Olmos and R. Urbanke, "Scaling behavior of convolutional LDPC ensembles over the BEC," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.–Aug. 31–5, 2011, pp. 1816–1820.

[30] M. Lentmaier, D. Truhachev, K. Zigangirov, and D. Costello, "An analysis of the block error probability performance of iterative decoding," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3834–3855, Nov. 2005.

[31] X.-Y. Hu, E. Eleftheriou, and D. Arnold, "Regular and irregular progressive edge-growth Tanner graphs," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 386–398, Jan. 2005.

[32] T. Tian, C. Jones, J. Villasenor, and R. Wesel, "Selective avoidance of cycles in irregular LDPC code construction," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1242–1247, Aug. 2004.

[33] T. Richardson and R. Urbanke, *Modern Coding Theory*.   New York: Cambridge Univ. Press, 2008.

Aravind R. Iyengar (S'09–M'12) received his B.Tech degree in Electrical Engineering from the Indian Institute of Technology Madras, Chennai, in 2007; his M.S. and Ph.D. degrees in Electrical Engineering from the University of California in San Diego, La Jolla, where he was affiliated with the Center for Magnetic Recording Research, in 2009 and 2012 respectively. He is currently with Qualcomm Technologies Inc., Santa Clara, where he is involved in the design of baseband modems. In 2006, he was a visiting student intern at the École Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), Cergy, France. He was a visiting doctoral student at the Communication Theory Laboratory at the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland in 2010. His research interests are in the areas of information and coding theory, and in signal processing and wireless communications.

A. R. Iyengar was the recipient of the Sheldon Schultz Prize for Excellence in Graduate Student Research at the University of California, San Diego in 2012.

Paul H. Siegel (M'82–SM'90–F'97) received the S.B. and Ph.D. degrees in mathematics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1975 and 1979, respectively.

He held a Chaim Weizmann Postdoctoral Fellowship at the Courant Institute, New York University. He was with the IBM Research Division in San Jose, CA, from 1980 to 1995. He joined the faculty at the University of California, San Diego in July 1995, where he is currently Professor of Electrical and Computer Engineering in the Jacobs School of Engineering. He is affiliated with the Center for Magnetic Recording Research where he holds an endowed chair and served as Director from 2000 to 2011. His primary research interests lie in the areas of information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission.

Prof. Siegel was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and from 2009 to 2011. He was re-elected for another 3-year term in 2012. He served as Co-Guest Editor of the May 1991 Special Issue on "Coding for Storage Devices" of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served the same Transactions as Associate Editor for Coding Techniques from 1992 to 1995, and as Editor-in-Chief from July 2001 to July 2004. He was also Co-Guest Editor of the May/September 2001 two-part issue on "The Turbo Principle: From Theory to Practice" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.

Prof. Siegel was co-recipient, with R. Karabed, of the 1992 IEEE Information Theory Society Paper Award and shared the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award with B.H. Marcus and J.K. Wolf. With J.B. Soriaga and H.D. Pfister, he received the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He holds several patents in the area of coding and detection, and was named a Master Inventor at IBM Research in 1994. He is an IEEE Fellow and a member of the National Academy of Engineering.

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

From 1995 to 1999, he held a position at the Mathematics of Communications Department at Bell Labs. Since November 1999, he has been a faculty member at the School of Computer & Communication Sciences of EPFL, Lausanne, Switzerland, where he is the head of the Communications Theory Lab as well as the head of the Doctoral Program of the School of Computer and Communication Sciences (comprising roughly 250 PhD students).

Dr. Urbanke's research is focused on the analysis and design of coding systems and, more generally, graphical models.

Dr. Urbanke is a recipient of a Fulbright Scholarship. From 2000–2004 he was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and he has been elected in October 2012 to the Board of Governors of IEEE Information Theory Society. He is also currently on the board of the series *Foundations and Trends in Communications and Information Theory*. He is a co-recipient of the IEEE Information Theory Society 2002 Best Paper Award and a co-recipient of the 2011 IEEE Kobayashi Computers and Communications Award. He is co-author of the book *Modern Coding Theory* published by Cambridge University Press.

Jack Keil Wolf (S'54–M'60–F'73–LF'97) received the B.S.E.E. degree from the University of Pennsylvania Philadelphia, in 1956, and the M.S.E., M.A., and Ph.D. degrees from Princeton University, Princeton, NJ, in 1957, 1958, and 1960, respectively. He was the Stephen O. Rice Professor of Electrical and Computer Engineering and a member of the Center for Magnetic Recording Research at the University of California-San Diego, La Jolla. He was a member of the Electrical Engineering Department at New York University from 1963 to 1965, and the Polytechnic Institute of Brooklyn from 1965 to 1973. He was Chairman of the Department of Electrical and Computer Engineering at the University of Massachusetts, Boston, from 1973 to 1975, and he was Professor there from 1973 to 1984. From 1984 to 2011, he was a Professor of Electrical and Computer Engineering and a member of the Center for Magnetic Recording Research at the University of California-San Diego. He also held a part-time appointment at Qualcomm, Inc., San Diego. From 1971 to 1972, he was an NSF Senior Postdoctoral Fellow, and from 1979 to 1980, he held a Guggenheim Fellowship. His most recent research interest was in signal processing for storage systems.

Dr. Wolf was elected to the National Academy of Engineering in 1993. He was the recipient of the 1990 E. H. Armstrong Achievement Award of the IEEE Communications Society and was co-recipient with D. Slepian of the 1975 IEEE Information Theory Group Paper Award for the paper "Noiseless coding for correlated information sources." He shared the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award with B. Marcus and P.H. Siegel for the paper "Finite-State Modulation Codes for Data Storage." He served on the Board of Governors of the IEEE Information Theory Group from 1970 to 1976 and from 1980 to 1986. Dr. Wolf was President of the IEEE Information Theory Group in 1974. He was International Chairman of Committee C of URSI from 1980 to 1983. He was the recipient of the 1998 IEEE Koji Kobayashi Computers and Communications Award, "for fundamental contributions to multi-user communications and applications of coding theory to magnetic data storage devices." In May 2000, he received a UCSD Distinguished Teaching Award. In 2004 Professor Wolf received the IEEE Richard W. Hamming Medal for "fundamental contributions to the theory and practice of information transmission and storage." In 2005 he was elected by the American Academy of Arts and Sciences as a Fellow, and in 2010 was elected as a member of the National Academy of Sciences. He was co-recipient with I.M. Jacobs of the 2011 Marconi Society Fellowship and Prize. Prof. Wolf passed away on May 12, 2011.