On Trapping Sets for Repeat Accumulate Accumulate Codes

Jörg Kliewer*, Christian Koller†, Alexandre Graell i Amat‡, and Daniel J. Costello, Jr.‡

*Klipsch School of Electrical and Computer Engineering,
New Mexico State University, Las Cruces, NM 88003, USA
Email: jkliewer@nmsu.edu

†Electronics Department,
Institut TELECOM-TELECOM Bretagne, 29238 Brest, France
Email: alexandre.graell@telecom-bretagne.eu

‡Department of Electrical Engineering,
University of Notre Dame, Notre Dame, IN 46556, USA
Email: {ckoller, costello.2}@nd.edu

***Abstract*—The serial concatenation of a repetition code with two or more accumulators has the advantage of a simple encoder structure. Furthermore, the resulting ensemble is asymptotically good and exhibits minimum distance growing linearly with block length. For low-density parity-check codes, the notion of trapping sets has been introduced to estimate the performance of these codes under non-maximum likelihood decoding. We briefly address asymptotic expressions for the normalized minimum trapping distance for the Gallager-Zyablov-Pinsker bit flipping decoding algorithm. Then we consider belief propagation decoding and present a closed form finite length ensemble average trapping set enumerator for repeat accumulate accumulate codes by creating a trellis representation of trapping sets. For this case, we also obtain asymptotic expressions and evaluate them numerically.**

I. Introduction

The serial concatenation of constituent codes combined with iterative decoding has led to powerful code constructions which can perform close to the Shannon limit. One of the simplest of such constructions is the repeat-accumulate (RA) code [1], which consists only of a repetition code, an interleaver, and an accumulator. One advantage of these codes is their very low encoding complexity of $O(1)$ and a moderate decoding complexity which makes them well suited for power-limited environments, e.g., in sensor networks. As a drawback, these codes are not asymptotically good [2–5], i.e., their minimum distance does not grow linearly with block length, which affects the performance in the error-floor regime. However, it has been shown in [5], [6] that, by adding an additional interleaver and accumulator, the resulting repeat accumulate accumulate (RAA) code ensemble is asymptotically good for rates equal to $1/3$ or smaller. While linear minimum distance growth with block length is useful in assessing the error-floor behavior under maximum-likelihood (ML) decoding, it does not provide an accurate characterization of the low bit error rate (BER) performance regime for practically feasible iterative decoding schemes.

For LDPC codes under belief propagation (BP) decoding, the height of the error floor has been linked to so called *near codewords* [7]. This concept was generalized to *trapping sets* in [8], which are associated with substructures in the Tanner graph of a code that may cause an iterative message passing decoder to fail. For certain LDPC codes, small trapping sets, rather than the minimum distance of the code, dominate the error floor regime. An (a, b) trapping set for a given parity check matrix and its corresponding Tanner graph is defined in [8] as a collection of a variable nodes for which the subgraph induced by the a variable nodes contains b check nodes with odd degree. The intuition behind this definition is that, if $a > b$, potentially unreliable messages within the induced subgraph may not be corrected by reliable messages from outside the trapping set due to the relatively small number of odd degree check nodes within the subgraph.

In the following, we extend these results to RAA code ensembles and give an overview of trapping set characterizations for such ensembles under both bit flipping and BP decoding [9]. In particular, we first focus on communication over a binary symmetric channel (BSC). For decoding, we employ a variant of the majority-logic (bit flipping) decoding algorithm introduced by Gallager [10] and Zyablov and Pinsker [11], resp., for LDPC codes on a factor graph representation of the RAA code. We show that by using this decoding approach an estimate of the minimum trapping distance can be obtained.

Second, we focus on BP decoding, which can be employed on the factor graph of RAA codes in the same way as for LDPC codes, albeit usually with a different message passing schedule and fewer iterations. Consequently, the decoder

This work was partly supported by NSF grants CCF05-15012 and CCF08-30666, NASA grant NNX07AK536, and the Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme.

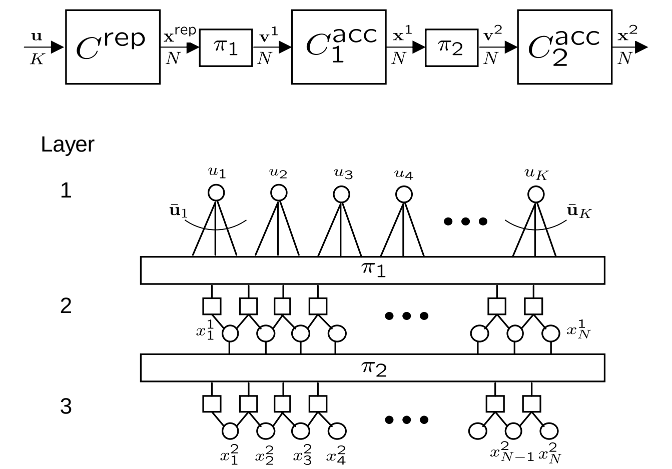

Fig. 1. Block diagram and factor graph of an RAA encoder ($q = 3$). The circles represent variable nodes and the boxes check nodes, respectively.

should also be susceptible to trapping sets. By using the above trapping set definition from [8], we address the enumeration of subgraphs in RAA codes and derive a closed form trapping set enumerator (TSE) for general (a, b) trapping sets. Further, we present asymptotic expressions for the TSE.

II. System setup

The encoder structure of an RAA code $\mathcal{C}_{\text{RAA}}$ of rate R is shown in Fig. 1. It is a serial concatenation of a repetition code C^{rep} of rate $R = R_{\text{rep}} = 1/q$ and two identical rate-1, memory-1, accumulate codes C_l^{acc}, $l = \{1, 2\}$, with generator polynomials $g(D) = 1/(1 + D)$, connected by interleavers π_1 and π_2. For the repetition code C^{rep}, we denote the binary input sequence of length K by $\mathbf{u} = [u_1, \ldots, u_K]$ and the binary output sequence of length N by $\mathbf{x}^{\text{rep}} = [x_1^{\text{rep}}, \ldots, x_N^{\text{rep}}]$. Likewise, for encoder C_l^{acc}, $\mathbf{v}^l = [v_1^l, \ldots, v_N^l]$ and $\mathbf{x}^l = [x_1^l, \ldots, x_N^l]$ denote the input sequence and the codeword, respectively, where both are of length N. Fig. 1 also shows the factor graph for the RAA code with $q = 3$, where the circles represent variable nodes and the boxes check nodes, respectively. The same quantities are displayed as in the encoder structure. In addition, the subset of the q values from $\mathbf{x}^{\text{rep}}$ which share edges with variable nodes u_k is denoted by $\bar{\mathbf{u}}_k = [\bar{u}_{k,1}, \bar{u}_{k,2}, \ldots, \bar{u}_{k,q}]$.

III. Bit flipping decoding of RAA codes

In the following, we consider transmitting RAA codewords over the BSC and describe a variant of the Gallager-Zyablov-Pinsker bit flipping decoding algorithm for LDPC codes [11], which is carried out on the factor graph of the code in Fig. 1. We define the sequence $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_K]$, where

$$\eta_k \triangleq \frac{w_H(\bar{\mathbf{u}}_k)}{q}, \tag{1}$$

and the syndrome symbols

$$s_k \triangleq \begin{cases} 0 & \text{if} \quad \eta_k = 0 \text{ or } \eta_k = 1, \\ 1 & \text{otherwise.} \end{cases} \tag{2}$$

In the first decoding iteration, those symbols in the received vector $\mathbf{x}^2$ for which flipping reduces the syndrome weight $w_H(\boldsymbol{s})$, with $\boldsymbol{s} = [s_1, \ldots, s_K]$, are marked. After flipping those symbols simultaneously, the resulting flipped sequence is considered as a new received sequence $\mathbf{x}^2$, and the next decoding iteration starts. The sequence $\mathbf{x}^2$ is called a *trapping set* if there are no candidate symbols available that lead to a reduction of the (non-zero) syndrome weight.

We now consider the case when the transmitted information sequence $\mathbf{u}$ of length K is the all-zero sequence, and we define the trapping distance of a trapping set in $\mathbf{x}^2$ as its Hamming weight $w_H(\mathbf{x}^2)$. The parameter d_{trap} is called the *minimum trapping distance* if there are no trapping sets with trapping distance $d_0 < d_{\text{trap}}$, and there is at least one trapping set with trapping distance d_{trap}.

An estimate of the minimum trapping distance can now be derived by computing ensemble-average weight spectra for trapping sets. Consider a corrupted received sequence $\mathbf{x}^2$ at the decoder input, and define the vector $\boldsymbol{\nu} \triangleq (\nu_0, \nu_1, \ldots, \nu_q)$, where each ν_ℓ counts the number of occurrences of the values ℓ/q, $\ell = 0, 1, \ldots, q$, in the sequence $\boldsymbol{\eta}$. The vector $\boldsymbol{\nu}$ satisfies the following relations:

$$\sum_{\ell=0}^{q} \nu_\ell = K \quad \text{and} \quad \sum_{\ell=0}^{q} \ell\, \nu_\ell = w_H(\mathbf{u}) = \omega, \tag{3}$$

We can now obtain the ensemble-average input-output weight enumerator (IOWE) for a given $\boldsymbol{\eta}$ and the weights $\omega = w_H(\mathbf{u})$, $d_1 = w_H(\mathbf{x}^1)$, and $d = w_H(\mathbf{x}^2)$ as follows.

Theorem 1 ([9])**.** *The ensemble-average IOWE $\bar{A}_{d,d_1,\omega}$ for an RAA code ensemble with rate $R = 1/q$ and a given vector $\boldsymbol{\nu} = (\nu_0, \nu_1, \ldots, \nu_q)$ equals*

$$\bar{A}_{d,d_1,\omega}(\boldsymbol{\nu}) = \frac{\binom{d_1 - 1}{\lceil \frac{\omega}{2} \rceil - 1}\binom{N - d_1}{\lfloor \frac{\omega}{2} \rfloor}\binom{d - 1}{\lceil \frac{d_1}{2} \rceil - 1}\binom{N - d}{\lfloor \frac{d_1}{2} \rfloor}}{\binom{N}{\omega}\binom{N}{d_1}} \cdot \binom{K}{\nu_0, \nu_1, \ldots, \nu_q} \cdot \frac{(q!)^K}{\prod_{\ell=0}^{q} (\ell!\,(q-\ell)!)^{\nu_\ell}}. \tag{4}$$

An intuitive condition for the absence of a trapping set can be given as

$$\nu_1 - q\,\nu_q > 0. \tag{5}$$

The idea behind this is as follows. If a candidate symbol x_n^2 is flipped, we assume that the syndrome weight can be decreased only for the case $\eta_k = 1/q$, whereas the syndrome weight can only increase for $\eta_k = 1$. In all other cases we assume that the syndrome weight will stay constant. Assuming that (5) holds, the ensemble average weight enumerator (WE) for received sequences of weight d that constitute trapping sets is now given as

$$\bar{A}_d = \sum_{\omega=1}^{N} \sum_{d_1=1}^{N} \sum_{\substack{\boldsymbol{\nu}\,:\, \sum_{\ell=0}^{q} \nu_\ell = K \\ \sum_{\ell=0}^{q} \ell\, \nu_\ell = \omega \\ \nu_1 - q\, \nu_q \le 0}} \bar{A}_{d,d_1,\omega}(\boldsymbol{\nu}). \tag{6}$$

Now, using the normalized trapping weight $\rho \triangleq d/N$, we define the function

$$r_{\text{trap}}(\rho) \triangleq \lim_{N\to\infty} \sup \frac{1}{N} \ln \bar{A}_{\rho N}, \tag{7}$$

where $\alpha \triangleq \omega/N$ and $\beta \triangleq d_1/N$ are normalized weights, s, λ are auxiliary parameters, and the supremum is taken over all remaining variables.

Theorem 2 ([9]). *Given the condition* (5)*, the function* $r_{\textit{trap}}(\rho)$ *can be upper bounded by*

$$\begin{aligned} r_{\textit{trap}}(\rho) &\le \hat{r}_{\textit{trap}}(\rho) \\ &= \sup_{0<\alpha<0.5,\, 0<\beta<0.5} \left(f(\alpha,\beta,\rho) + \inf_{s,\,\lambda<0} g(\alpha,s,\lambda) \right). \end{aligned} \tag{8}$$

The function $f(\alpha,\beta,\rho)$ is given as

$$\begin{aligned} f(\alpha,\beta,\rho) = -H(\beta) - H(\rho) + H\left(\frac{\beta-\alpha/2}{1-\alpha}\right)(1-\alpha) + \\ \alpha \ln 2 + H\left(\frac{\rho-\beta/2}{1-\beta}\right)(1-\beta) + \beta \ln 2, \end{aligned} \tag{9}$$

whereas $g(\alpha,s,\lambda)$ depends on q. For $q=3$, it is given as

$$g(\alpha,s,\lambda) = \frac{1}{3} \ln\left(1 + 3\, e^{s+\lambda} + 3\, e^{2\, s} + e^{3\,(s-\lambda)}\right) - s\alpha. \tag{10}$$

Fig. 2 shows the resulting lower bound ρ'_{trap} on the asymptotic normalized minimum trapping distance for $q \ge 3$, obtained by numerically evaluating the zero crossing in (8). Assuming that (5) holds, the result suggests that the minimum trapping distance $d_{\text{trap}} = \rho_{\text{trap}}\, N$ grows linearly with block length, even though the growth rate coefficients are very small.

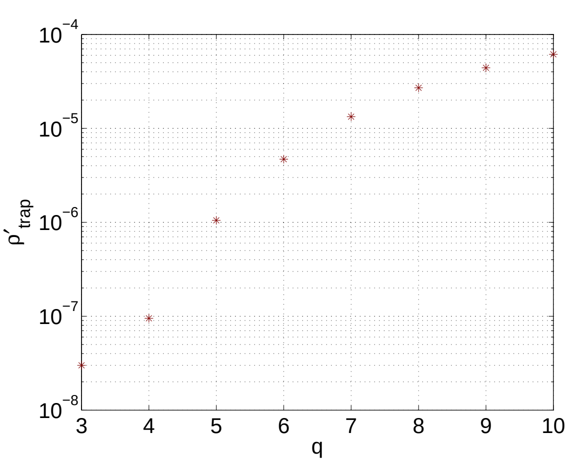

Fig. 2. Lower bound ρ'_{trap} on the normalized asymptotic trapping distance for the RAA code ensemble versus the inverse rate $q = 1/R$.

IV. Trapping set enumerators

In this section we consider BP decoding on the graph in Fig. 1. To this end, let $\bar{A}_{a,b}^{\mathcal{C}_{\text{RAA}}}$ be the TSE of an RAA code ensemble, i.e., the average number of (a,b) trapping sets according to the definition in [8]. With reference to Fig. 1, we denote by w the number of information bits that participate in an (a,b) trapping set of $\mathcal{C}_{\text{RAA}}$. Also, let a_1^i, a_1^o, and b_1 be the number of variable nodes corresponding to input bits, the number of variable nodes corresponding to code bits, and the number of unsatisfied checks, respectively, of code C_1^{acc} involved in an (a,b) trapping set of $\mathcal{C}_{\text{RAA}}$. Likewise, we define a_2^i, a_2^o, and b_2 for code C_2^{acc}. In the factor graph, the accumulator input bits a_i^1 and a_i^2 are not directly depicted, but are represented by the output bits of the next higher level. In addition, we define the enumerators of the component codes $A_{w,qw}^{C^{\text{rep}}}$ and $A_{a_l^i,a_l^o,b_l}^{C_l^{\text{acc}}}$, for $l = \{1,2\}$.

Since there are no check nodes in layer 1 of the factor graph, $A_{w,qw}^{C^{\text{rep}}}$ is the IOWE of the repetition code, giving the number of codewords in C^{rep} of input weight w and output weight qw, while $A_{a_l^i,a_l^o,b_l}^{C_l^{\text{acc}}}$ is the input-output trapping set enumerator (IOTSE) of code C_l^{acc}, denoting the number of trapping sets in C_l^{acc} consisting of a_l^i input variable nodes (i.e., variable nodes corresponding to information bits for code $\mathcal{C}_l^{\text{acc}}$), a_l^o output variable nodes (i.e., variable nodes corresponding to code bits), and b_l unsatisfied checks.

With these definitions, the ensemble average TSE $\bar{A}_{a,b}^{\mathcal{C}_{\text{RAA}}}$ can be computed using the uniform interleaver concept [12] as

$$\begin{aligned} \bar{A}_{a,b}^{\mathcal{C}_{\text{RAA}}} &= \sum_{\substack{w,a_1^o,a_2^o:\; w+a_1^o+a_2^o=a \\ b_1,b_2:\; b_1+b_2=b}} \frac{A_{w,qw}^{C^{\text{rep}}} A_{qw,a_1^o,b_1}^{C_1^{\text{acc}}} A_{a_1^o,a_2^o,b_2}^{C_2^{\text{acc}}}}{\binom{N}{qw}\binom{N}{a_1^o}} \\ &= \sum_{\substack{w,a_1^o,a_2^o:\; w+a_1^o+a_2^o=a \\ b_1,b_2:\; b_1+b_2=b}} \bar{A}_{w,a_1^o,b_1,a_2^o,b_2}^{\mathcal{C}^{\text{RAA}}}, \end{aligned} \tag{11}$$

where $\bar{A}_{w,a_1^o,b_1,a_2^o,b_2}^{\mathcal{C}^{\text{RAA}}}$ is called the ensemble-average conditional TSE.

In the following, we address the computation of the IOTSE $A_{a_l^i,a_l^o,b_l}^{C_l^{\text{acc}}}$ of an accumulate code by considering an equivalent trellis representation of trapping sets in the factor graph, shown in Fig. 3. Each edge between two trellis states is labeled with a binary 3-tuple $s_i/c/s_o$, where s_i denotes the input symbol, s_o denotes the output symbol, and c is 1 if the check node in the corresponding equivalent factor graph representation is unsatisfied. The IOTSE of the accumulate code can now be computed by considering a trellis consisting of N concatenated trellis sections like the one in Fig. 3 and enumerating all possible paths.

Theorem 3 ([13]). *Let* (a^i, a^o, b) *be a trapping set with* a^i *information variable nodes,* a^o *code variable nodes, and* b *unsatisfied checks. The input-output trapping set enumerator (IOTSE) for the rate-1, memory-one, convolutional encoder*

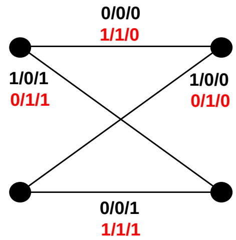

Fig. 3. Extended trellis section.

C^{acc} *with generator polynomial* $g(D) = 1/(1+D)$*, terminated to the all-zero state at the end of the trellis, and with input and output block length* N*, can be given in closed form as*

$$A_{a^i,a^o,b} = \sum_m \sum_n \binom{N-a^o}{m}\binom{a^o-1}{m-1} \cdot \cdot \binom{a^o-m}{\frac{a^i+b}{2}-n-m}\binom{N-a^o-m}{n}\binom{2m}{\frac{a^i-b}{2}+m}, \tag{12}$$

where m *and* n *must satisfy the constraints*

$$\begin{aligned} m &\geq \frac{|a^i-b|}{2}, \quad m \leq \min\{a^o, N-a^o\} \\ n &\geq \frac{a^i+b}{2} - a^o, \quad n \leq N - a^o - m. \end{aligned} \tag{13}$$

In order to determine the asymptotic behavior, we define the normalized logarithmic asymptotic TSE $r^{\mathcal{C}}(\alpha,\beta)$ of a code ensemble $\mathcal{C}$, analogous to (7), as

$$r^{\mathcal{C}}(\alpha,\beta) = \lim_{N\to\infty} \sup \frac{\ln \bar{A}^{\mathcal{C}}_{a,b}}{N}, \tag{14}$$

where $\alpha = a/N$ and $\beta = b/N$.

By inserting (11) into (14) and considering Theorem 3, the resulting expression for the asymptotic TSE can be numerically evaluated. Following [14], we keep the ratio $\Delta = \beta/\alpha$ of unsatisfied check nodes and erroneous variable nodes constant and compute $r(\alpha, \Delta\alpha)$ for varying values of α. In Fig. 4 the unsatisfied checks in the RAA code ensemble are equally distributed among the middle and inner accumulator, i.e., $\beta_1 = \beta_2 = \beta/2$. For $\Delta = 0$, when no unsatisfied checks are present in the factor graph, $r(\alpha, 0)$ exhibits a zero stretch in the beginning and turns positive when the number of codewords starts to grow exponentially in N with increasing α.

In Fig. 5 we display the influence of the repetition factor q on the shape of the asymptotic TSE. The case of $\beta_1 = \beta$ is shown, where varying q has the greatest effect on the slope. Yet the reduction in slope caused by increasing the repetition factor is only marginal and the curves almost lie on top of each other. Finally, we note that, because of the increasing minimum distance growth rate of the code ensemble with increasing repetition factor q, the transition from the quasi-linear section of the asymptotic TSE to the more steeply increasing section takes place at higher values of α for larger q.

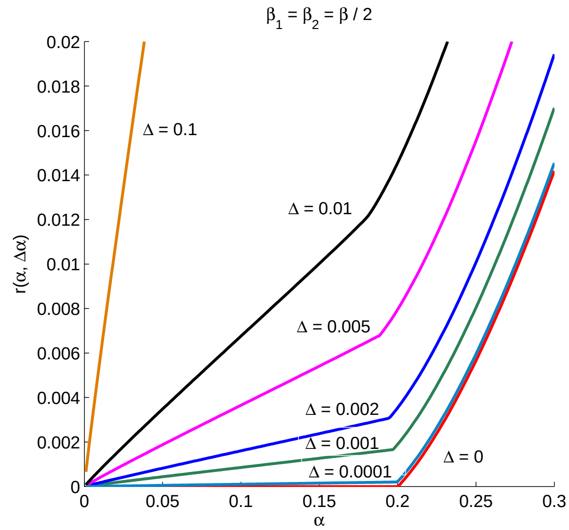

Fig. 4. Asymptotic TSE for different values of Δ and $\beta_1 = \beta_2 = \beta/2$.

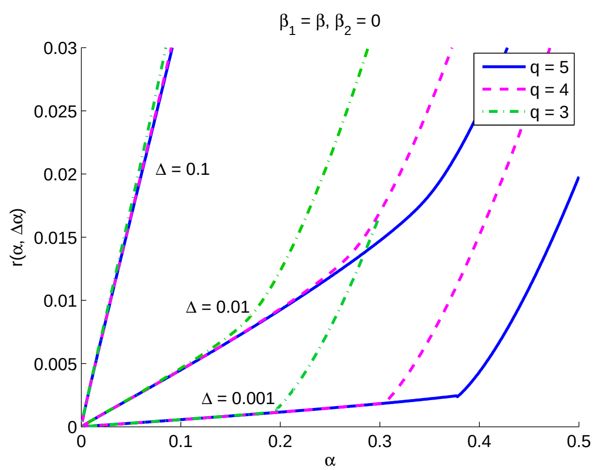

Fig. 5. Asymptotic TSE of the RAA code ensemble for different repetition factors q.

V. Conclusion

In this paper we have described two methods to determine trapping sets for RAA code ensembles. First, we have considered a variant of the Gallager-Zyablov-Pinsker bit flipping decoding algorithm on the factor graph of the code. An estimation of the minimum trapping distance suggests that for all rates smaller than or equal to $R = 1/3$, the minimum trapping distance grows linearly with block length, although the growth rate coefficients are extremely small.

Second, we have presented a simple closed form method to enumerate general (a, b) trapping sets for RAA code ensembles according to the definition in [8]. We have observed that, when unsatisfied check nodes are present in the factor graph, the asymptotic TSE lies strictly above the asymptotic TSE for the case when no unsatisfied check nodes exist in the graph. Further, in contrast to our results for the bit flipping decoder, there exists no regime where the minimum trapping set size grows linearly with block length. Thus, the influence that (a, b) trapping sets have on the error floor must still be evaluated separately.

REFERENCES

[1] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proc. 36th Annual Allerton Conf. Commun., Control, Computing*, Monticello, IL, Sept. 1998, pp. 201–210.

[2] L. Bazzi, M. Mahdian, and D. A. Spielman, "The minimum distance of turbo-like codes," Submitted to *IEEE Trans. Inf. Theory*, May 2003.

[3] N. Kahale and R. Urbanke, "On the minimum distance of parallel and serially concatenated codes," in *Proc. IEEE Int. Symposium on Inform. Theory*, Cambridge, MA, Aug. 1998, p. 31.

[4] A. Perotti and S. Benedetto, "An upper bound on the minimum distance of serially concatenated convolutional codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5501–5509, Dec. 2006.

[5] J. Kliewer, K. S. Zigangirov, and D. J. Costello, Jr., "New results on the minimum distance of repeat-multiple-accumulate codes," in *Proc. 45th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2007, pp. 1097–1102.

[6] H. D. Pfister and P. H. Siegel, "The serial concatenation of rate-1 codes through uniform random interleavers," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1425–1438, June 2003.

[7] D. J. C. MacKay and M. Postol, "Weaknesses of Margulis and Ramanujan-Margulis low-density parity-check codes," *Electron. Notes Theor. Comput. Scie.*, vol. 74, 2003.

[8] T. J. Richardson, "Error floors of LDPC codes," in *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2003, pp. 1426–1435.

[9] J. Kliewer, K. S. Zigangirov, and D. J. Costello, Jr., "On the minimum trapping distance of repeat accumulate accumulate codes," in *Proc. 46th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2008.

[10] R. G. Gallager, *Low-density parity-check codes*, MIT Press, Cambridge, MA, 1963.

[11] V. V. Zyablov and M. S. Pinsker, "Estimates of the error-correction complexity of Gallager's low-density codes," *Problems of Information Transmission*, vol. 11, no. 1, pp. 18–28, January-March 1976.

[12] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 909–926, May 1998.

[13] C. Koller, A. Graell i Amat, J. Kliewer, and D. J. Costello, Jr., "Trapping set enumerators for repeat multiple accumulate code ensembles," Submitted to *IEEE Symposium on Information Theory 2009*, January 2009.

[14] S. Abu-Surra, W. E. Ryan, and D. Divsalar, "Ensemble trapping set enumerators for protograph-based LDPC codes," in *Proc. 45th Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, Sept. 2007, pp. 201–210.