

# Variations of Repeat Accumulate Codes

## Literature Survey

Wei Wang

University of California, Berkeley  
Department of Electrical Engineering  
wangwei@eecs.berkeley.edu

May 24, 2002

### Abstract

In this paper I present a class of parallel and serially concatenated linear block codes, focusing specifically on the case of repeat accumulate codes and the more general irregular repeat accumulate codes. The analytical tractability of repeat accumulate codes has enabled the first derivations of coding theorems for any class of turbo-like codes. Experimentally, repeat accumulate codes allow linear-time encoding and decoding with performance comparable to turbo and LDPC codes.

## 1 Introduction

Turbo codes were introduced in 1993 [1] as a practical code construction whose performance approaches Shannon's theoretical limit for the capacity of noisy channels, which was first introduced in 1948 [2]. It has led to the subsequent rediscovery of low-density parity-check codes [3], and the discovery of the connection between iterative decoding and belief propagation [4][5]. There has been a recent explosion of interest in such codes defined on sparse random graphs. [6][7]

However, analysis of turbo codes has mostly been experimental, far outpacing theoretical results. Also, the following fundamental problem has still

not been solved: for a given channel, find a code that has linear-time encoding and decoding, and has rate arbitrarily close to channel capacity. Turbo codes satisfy the linear-time encoding criteria while LDPC codes have encoding times quadratic in the block length. But there is a gap between channel capacity and iterative decoding thresholds for both turbo codes and LDPC codes on binary symmetric and AWGN channels. [8] Although irregular LDPC codes satisfy the capacity achieving criteria on binary erasure channels.

Repeat accumulate (RA) codes were first studied for their analytical tractability as a simple case of a general class of parallel and serially concatenated codes. Divsalar, Jin, and McEliece developed the first rigorous proofs of coding theorems for any class of turbo-like codes, by proving that the maximum likelihood word error probability for RA codes over a memoryless binary-input channel approaches zero as block length  $k \rightarrow \infty$ . [9] Iterative decoding of RA codes is suboptimal but linear-time and gave good experimental performance. RA codes were then generalized to irregular repeat accumulate (IRA) codes. [8] Coding theorems were proven for IRA codes on the binary erasure channel, and less rigorously on the AWGN channel.

In Section 2, I will define a general class of concatenated codes and the cases of regular and irregular repeat accumulate codes. I will illustrate the structure of RA codes using Tanner graphs, and compare them to the structure of LDPC and turbo codes. In Section 3, I will outline proofs of coding theorems bounding maximum-likelihood decoding error probabilities for repeat accumulate codes. In Section 4, I will briefly discuss linear-time iterative decoding algorithms for RA and IRA codes and proofs of capacity achievability for IRA codes. Finally in Section 5, I will mention analysis that has been done on the class of concatenated codes beyond repeat accumulate codes.

## 2 Structures of Concatenated codes

### 2.1 General class of concatenated codes

Consider the class of concatenated codes depicted in Figure 1 with  $q$  encoders (circles) and  $q - 1$  interleavers (boxes). The  $i$ th code  $\mathcal{C}_i$  is an  $(n_i, k_i)$  linear block code, and is preceded by an interleaver  $\mathcal{P}_i$  of size  $k_i$ , except  $\mathcal{C}_1$  which is connected directly to a length  $k$  block of input information bits. The overall structure must be a graph theoretic tree (i.e. have no loops). Define  $s_q = \{1, 2, \dots, q\}$  and its subsets  $s_O = \{i \in s_q : \mathcal{C}_i \text{ connected to output}\}$  and  $\bar{s}_O$ , the complement of  $s_O$ . The overall system in Fig. 1 is an encoder for an  $(n, k)$  block code, where  $n = \sum_{i \in s_O} n_i$ .

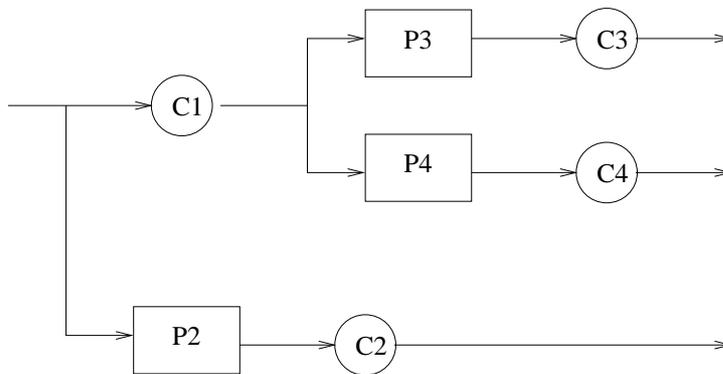


Figure 1: Class of concatenated “turbo-like” codes

### 2.2 Repeat accumulate codes

Special cases of this general class of “turbo-like” codes include parallel concatenated convolutional codes (e.g. classical turbo codes), and serial concatenated codes. In particular, consider the serially concatenated code depicted in Figure 2, where the outer code is a rate  $1/q$  repetition code and the inner code is a truncated rate 1 convolutional code with transfer function  $\frac{1}{1+D}$ . This is a repeat accumulate code (or RA code) of rate  $1/q$ . Put more simply, an information block of length  $k$  is repeated  $q$  times, scrambled by

a fixed random permutation, and put through a block code whose input  $[x_1, \dots, x_{qk}]$  and output  $[y_1, \dots, y_{qk}]$  are related by

$$\begin{aligned} y_1 &= x_1 \\ y_2 &= x_1 + x_2 \\ y_3 &= x_1 + x_2 + x_3 \\ &\vdots \\ y_n &= x_1 + x_2 + x_3 + \dots + x_{qk}. \end{aligned}$$



Figure 2: Repeat accumulate codes

Here I will illustrate the structure of repeat accumulate codes using Tanner graph representations. A Tanner graph  $G = (V, E)$  is a bipartite graph whose vertices are partitioned into variable nodes  $V_m$  and check nodes  $V_c$  with edges  $E \subseteq V_m \times V_c$ . Check nodes represent local constraints on subsets of variable nodes; an edge indicates that a variable participates in a constraint.

For an  $(qk, k)$  RA code, denote the  $k$  information bits by  $U = \{u_1, \dots, u_k\}$ , the  $qk$  code bits by  $X = \{x_1, \dots, x_{qk}\}$ , and the  $qk$  intermediate bits (inputs to the inner code) by  $\{v_1, \dots, v_{qk}\}$ . Represent the  $qk$  equations describing the accumulator inner code with check nodes  $C = \{c_1, \dots, c_{qk}\}$

$$x_i = \begin{cases} v_1 & \text{if } i = 1 \\ v_i + x_{i-1} & \text{otherwise.} \end{cases}$$

Represent both information bits  $U$  and code bits  $X$  by variable nodes. The Tanner graph representation of RA codes, with  $V_m = U \cup X$  (open circles) and  $V_c = C$  (filled circles), is illustrated in Figure 3.

Information Node      Check Node      Code Node

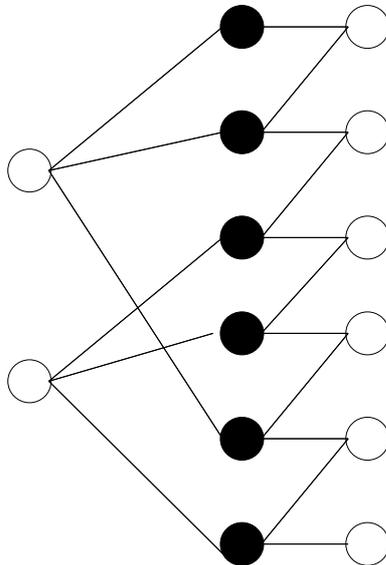


Figure 3: Tanner graph representation of repeat accumulate codes

### 2.3 Irregular repeat accumulate codes

Irregular repeat accumulate codes (or IRA codes) are a generalization of repeat accumulate codes. The Tanner graph representation of IRA codes with parameters  $(f_1, \dots, f_J; a)$ , where  $f_i \geq 0$ ,  $\sum_i f_i = 1$ , and  $a \in \mathcal{Z}^+$ , is illustrated in Figure 4. There are  $k$  variable nodes on the left representing the information bits  $\{u_1, \dots, u_k\}$ , and each such information node is connected to  $i$  check nodes, where the fraction of information nodes connected to exactly  $i$  check nodes is  $f_i$ . There are  $r = (k \sum_i i f_i) / a$  intermediate check nodes, with each check node connected to exactly  $a$  information nodes. The  $ra$  edges joining information nodes and check nodes are configured by a fixed random permutation. Finally the check nodes are connected to  $r$  variable nodes on the right representing the parity bits  $\{x_1, \dots, x_r\}$ .

The value of the parity bits are determined uniquely by the condition that the mod-2 sum of the values of the variable nodes connected to each

check node is zero. Denoting the values on the  $ra$  edges between the information nodes and the check nodes by  $\{v_1, \dots, v_{ra}\}$ , the recursive accumulator formula comes out

$$x_j = x_{j-1} + \sum_{i=1}^a v_{(j-1)a+i}, \quad j = 1, 2, \dots, r$$

where  $x_0 = 0$ . Observe that for fixed  $a$  and as  $n \rightarrow \infty$ , the encoding complexity of IRA codes is  $O(n)$ .

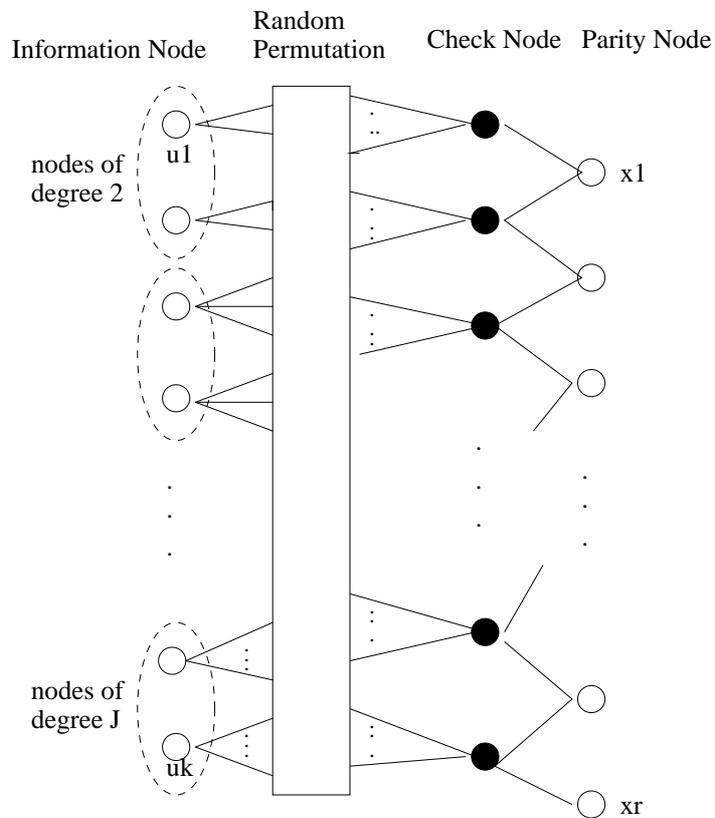


Figure 4: Tanner graph representation of irregular repeat accumulate codes

The IRA code depicted in Fig. 4 can be nonsystematic or systematic. The nonsystematic  $(r, k)$  IRA code encodes the information bits  $\{u_1, \dots, u_k\}$

as  $\{x_1, \dots, x_r\}$ , and has rate

$$R = \frac{a}{\sum_i i f_i}.$$

The systematic  $(k+r, k)$  IRA code encodes  $U$  as  $\{u_1, \dots, u_k, x_1, \dots, x_r\}$ , and has rate

$$R = \frac{a}{a + \sum_i i f_i}.$$

Observe that the regular RA codes are nonsystematic IRA codes with  $a = 1$  and  $f_i = \text{indicator}(i = q)$ .

### 3 Proving Coding Theorems Using ML Decoding

Coding theorems for the family of repeat accumulate codes on memoryless binary-input channels were proven by upper bounding the average maximum-likelihood decoding word error probability with a union bound, and applying the uniform interleaver technique to derive weight enumerators of code ensembles. [9]

Consider a memoryless binary input channel with input alphabet  $\{0, 1\}$ , output alphabet  $\Omega$ , and transition probabilities  $p(y|0), p(y|1)$ . The Bhattacharya parameter  $\gamma$  is defined as

$$\gamma = \begin{cases} \sum_{y \in \Omega} \sqrt{p(y|0)p(y|1)} & \text{if } \Omega \text{ is finite} \\ \int_{\Omega} \sqrt{p(y|0)p(y|1)} dy & \text{if } \Omega = \mathcal{R}^r. \end{cases}$$

$\gamma^h$  is an upper bound on the ML decoder error probability for a binary code with two codewords separated by Hamming distance  $h$ . Denote the weight enumerator of an  $(n, k)$  linear code by  $A_0, A_1, \dots, A_n$ , where  $A_i$  is the number of codewords of weight  $i$  in  $\mathcal{C}$ . Thus for an  $(n, k)$  binary linear code with known weight enumerator, the union bound on the ML decoder word error probability is

$$P_W \leq \sum_{h=1}^n A_h \gamma^h.$$

Consider a system of concatenated truncated convolutional codes as de-

picted in Fig. 1. Define a code ensemble  $\mathcal{C}_{n_1}, \mathcal{C}_{n_2}, \dots$ , where  $\mathcal{C}_{n_i}$  is a set of  $(n_i, k_i)$  codes with common rate  $R_i = \frac{k_i}{n_i}$ . A uniform interleaver is defined as a probabilistic mapping of a given input word of weight  $w$  into all distinct  $\binom{k_i}{w}$  permutations of it with equal probability  $p = 1/\binom{k_i}{w}$ . A code ensemble is thus generated by a uniform interleaver.

Denote the input-output weight enumerator (IOWE) for code  $\mathcal{C}$  by  $\{A_{w,h}\}_{0 \leq w \leq k, 0 \leq h \leq n}$ , where  $A_{w,h}$  is the number of encoder input-output pairs with input weight  $w$  and output weight  $h$ . Given the IOWE  $A_{w_i, h_i}^{(i)}$ 's for the codes  $\mathcal{C}_i$ , the ensemble IOWE  $\bar{A}_{w,h}$  is

$$\bar{A}_{w,h} = \sum_{h_i: i \in S_O, \sum h_i = h} \sum_{h_i: i \in S_I} A_{w_1, h_1}^{(1)} \prod_{i=2}^q \frac{A_{w_i, h_i}^{(i)}}{\binom{k_i}{w}}.$$

For an input block of length  $k$  the union bound on the ML decoder word error probability is

$$P_W^{UB} = \sum_{h=1}^n \left( \sum_{w=1}^k \bar{A}_{w,h} \right) \gamma^h.$$

Define

$$\alpha(w, h) = \limsup_{k \rightarrow \infty} \log_k \bar{A}_{w,h}$$

$$\beta_M = \max_{h \geq 1} \max_{w \geq 1} \alpha(w, h)$$

The parameter  $\beta_M$  is called the interleaving gain exponent.

### 3.1 Interleaving gain exponent conjecture

The IGE conjecture states that there exists a  $\gamma_0$  such that for any  $\gamma < \gamma_0$ , as the block length  $k$  becomes large

$$P_W^{UB} = O(k^{\beta_M}).$$

The IGE conjecture implies that if  $\beta_M < 0$ , then for a given  $\gamma < \gamma_0$  the word error probability of the concatenated code decreases to zero as the

input block size is increased.

### 3.2 RA codes

D. Divsalar, H. Jin, and R.J. McEliece [9] derived the ensemble IOWE for a  $(qk, k)$  RA code

$$A_{w,h} = \sum_{h_1=0}^{qk} \frac{A_{w,h_1}^{(1)} A_{h_1,h}^{(2)}}{\binom{qk}{qw}} \quad (1)$$

$$= \frac{\binom{k}{w} \binom{qk-h}{\lfloor qw/2 \rfloor} \binom{h-1}{\lceil qw/2 \rceil - 1}}{\binom{qk}{qw}} \quad (2)$$

and used it to prove the IGE conjecture for RA codes. They found that

$$\beta_M = -\lceil \frac{(q-2)}{2} \rceil.$$

Thus an RA code can have word error probability gain only if  $q \geq 3$ .

#### 3.2.1 RA codes achieve channel capacity

H. Jin and R.J. McEliece prove that RA codes have the potential to achieve channel capacity. [10] Specifically they show that as the rate of the RA code approaches zero, the average required bit  $E_b/N_0$  for arbitrarily small error probability with ML decoding approaches  $\log 2$ , which is the Shannon limit.

## 4 Iterative Decoding

### 4.1 RA codes

The complexity of ML decoding of RA codes is prohibitively large. Using the Tanner graph representations (see Section 2), message passing algorithms can be applied to decode RA codes. [11] The belief propagation algorithm

is an instance of GDL [5] and, as simulations show [9][12], gives good performance decoding in linear time despite being a suboptimal decoder.

## 4.2 IRA codes

Similarly, an iterative sum-product message-passing decoding algorithm is defined for IRA codes. [8] Furthermore, since the sum-product algorithm simplifies considerably on the binary erasure channel (BEC), H. Jin, A. Khandekar, and R.J. McEliece prove that irregular repeat accumulate codes can be encoded and decoded in linear time at rates arbitrarily close to channel capacity for the BEC. Their proof uses fixed point analysis of iterative decoding and density evolution. They do not have a rigorous proof for the AWGN channel, but demonstrate the performance of IRA codes with numerical results for both channels. In their simulations, IRA codes perform slightly better than turbo codes of comparable complexity, and as good as the best irregular LDPC codes. They show that IRA codes combines the favorable attributes of turbo and LDPC codes in that IRA codes have linear time encoding (like turbo codes) and are amenable to Richardson-Urbanke style analysis (like LDPC codes).

## 5 Beyond Repeat Accumulate Codes

H. Jin, in his thesis [11], applies similar methods used to prove coding theorems for RA codes to variants of RA codes, specifically repeat-delay-delay codes, convolution-accumulate codes, and repeat-accumulate-accumulate codes. He further applies the same method to prove the IGE conjecture for all parallel and serial turbo codes on any memoryless binary-input channel.

[13] demonstrated both analytically and experimentally the performance of serially concatenated trellis coded modulation with one or more inner accumulate codes.

## 6 Conclusion

We have seen that the discovery of repeat accumulate codes has led to the development of a string of provable coding theorems for the class of concatenated truncated convolutional codes. The experimental performance of these simple codes is comparable to turbo codes and LDPC codes, but unlike those other codes, repeat accumulate codes can be encoded and decoded in linear time. It still remains to be proven that repeat accumulate codes or its variations can achieve channel capacity on any memoryless binary-input channel.

## References

- [1] S. Benedetto and G. Montorsi, “Unveiling turbo codes: Some results on parallel concatenated coding schemes,” *IEEE Trans. on Info. Theory*, vol. 42, no. 2, pp. 409–428, March 1996.
- [2] C.E. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, 1948.
- [3] R. Gallager, *Low-Density Parity-Check Codes*, The M.I.T. Press, 1963.
- [4] R.J. McEliece, D.J.C. MacKay, and J.F. Cheng, “Turbo decoding as an instance of pearl’s ‘belief propagation’ algorithm,” *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 140–152, February 1998.
- [5] S. Aji and R.J. McEliece, “The generalized distributive law,” *IEEE Trans. on Info. Theory*, vol. 46, no. 2, pp. 325–343, March 2000.
- [6] D.J.C. MacKay, “Gallager codes - recent results,” .
- [7] D.J.C. MacKay, “Sparse graph codes,” .
- [8] H. Jin, A. Khandekar, and R.J. McEliece, “Irregular repeat-accumulate codes,” in *Proc. 2nd International Symposium on Turbo Codes*, Brest, France, September 2000, pp. 1–8.

- [9] D. Divsalar, H. Jin, and R.J. McEliece, "Coding theorems for 'turbo-like' codes," in *Proc. 36th Allerton Conf. on Communication, Control and Computing*, September 1998, pp. 201–210.
- [10] H. Jin and R.J. McEliece, "Ra codes achieve awgn channel capacity," in *13th Symp. on Applied Algebra, Algebraic Algorithms and Error Correcting Codes*, Hawaii, November 1999, pp. 10–18.
- [11] H. Jin, *Analysis and Design of Turbo-like Codes*, Ph.D. thesis, California Institute of Technology, 2001.
- [12] D.J.C. MacKay, "Decoding times of repeat-accumulate codes," October 1998.
- [13] H.M. Tullberg and P.H. Siegel, "Serial concatenated trellis coded modulation with inner rate-1 accumulate code," .